



学校代码: 10286  
分 类 号: TP391.4  
密 级: 公开  
U D C: 004.8  
学 号: 208182



心於至善

医学图像分割中的数据高效学习算法研究

# 东南大学 博士学位论文

## 医学图像分割中的数据高效学习算法 研究

研究生姓名: 何宇霆

导师姓名: 杨冠羽 教授

何宇霆

东南大学

申请学位类别 工学博士 学位授予单位 东南大学

一级学科名称 计算机科学与技术 论文答辩日期 2023年11月22日

二级学科名称 \_\_\_\_\_ 学位授予日期 2023年12月20日

答辩委员会主席 \_\_\_\_\_ 评 阅 人 盲审  
\_\_\_\_\_



2023 年 12 月 27 日

学校代码: 10286

分 类 号: TP391.4

密 级: 公开

U D C: 004.8

学 号: 208182



东南大学

SOUTHEAST UNIVERSITY

博士学位论文

医学图像分割中的数据高效学习算法

研究

研究生姓名: 何宇霆

导师姓名: 杨冠羽 教授

申请学位类别 工学博士 学位授予单位 东南大学

一级学科名称 计算机科学与技术 论文答辩日期 2023年11月22日

二级学科名称 \_\_\_\_\_ 学位授予日期 2023年12月20日

答辩委员会主席 \_\_\_\_\_ 评 阅 人 \_\_\_\_\_ 盲审 \_\_\_\_\_

2023 年 12 月 27 日



東南大學  
博士学位论文

医学图像分割中的数据高效学习算法  
研究

专业名称: 计算机科学与技术

研究生姓名: 何宇霆

导师姓名: 杨冠羽 教授



# DATA EFFICIENT LEARNING ALGORITHM IN MEDICAL IMAGE SEGMENTATION

A Dissertation submitted to  
Southeast University  
For the Academic Degree of Doctor of Engineering

BY  
HE Yuting

Supervised by:

Prof. YANG Guanyu

School of Computer Science and Engineering  
Southeast University

June 2023



## 摘要

随着我国社会的发展，人们对医疗资源的需求正在快速增长，然而医疗工作者的数量和质量、高水平医疗设备的数量却增长缓慢，使得医疗需求和医疗资源之间的矛盾越发尖锐。为缓解这一医疗矛盾，迫切需要利用科技手段释放有限的医疗资源，提高医疗资源分布的均衡性和使用的效率。医学图像分割技术能够自动化提取和可视化医学图像中的目标结构，提高医生的诊断效率和质量，从而更好服务于更广泛的患者群体，是医疗健康技术的重要研究内容之一。尤其是近几年来，以深度学习为代表的人工智能技术以其强大的学习能力，已经在医学图像分割领域展现了巨大的应用潜力。然而，深度学习强大的非线性学习能力使其对图像的表征依赖于对大量数据的学习，而医学图像采集昂贵、图像质量较低、标注者专业要求高且标注过程非常耗时的特点使得难以获得高质量、大规模的医学图像分割数据集。最终导致了医学图像分割任务面临严重的数据效率问题，对该技术在医学图像领域的大规模发展和应用构成了严峻挑战。这些挑战可以概括为以下两个方面：1) 图像表征挑战：医学图像具有巨大的视觉差异，同时图像内的解剖结构不明显，这使得深度神经网络在有限的数据下难以有效表征任务相关特征，并且容易受到任务不相关特征的干扰，从而影响模型的泛化能力。2) 标签不足挑战：由于医学图像标注的高成本以及深度神经网络对大量监督信息的依赖，数据集中有限的标签数量难以支撑可靠的医学图像分割学习，最终限制模型的分割性能。为了解决上述挑战，

本文聚焦于医学图像分割，着重研究如何提升深度学习模型在学习医学图像分割任务时的数据使用效率，旨在提高学习算法对医学图像的表征能力，同时减少对标签数量的依赖，实现医学图像分割中的数据高效学习。本文的研究重点包括以下四个方面：

1) 对于标签数量有限的情况下精细肾动脉分割任务，我们设计了一种基于半监督学习、密集偏置神经网络结构和难区域自适应损失的分割算法。该算法嵌入大量来自无标签图像的知识，并构建对肾动脉具有较强表征能力的神经网络，最终通过一种有效的损失函数，实现了对肾动脉的精细分割学习，为肾癌手术的术前规划提供了技术支持。

2) 对于图像间分布变化大，图像内目标结构特征不显著的肾癌多结构分割任务，我们构建了一种自适应不同目标分割结构分布和图像分布的元学习算法。该算法通过自适应搜索任务分布最适合的图像分布窗口，并采用图像级的动态特征融合元学习方法，实现了高效准确的肾癌多结构分割学习，展现出其在肾癌多结构分割中的有效性，进一步展现出其在肾部分切除术中的潜在临床价值。

3) 为了进一步降低标签需求，构建更通用的医学图像分割算法框架，我们设计了

---

一种基于配准的伪标签数据生成和鲁棒分割学习算法。该算法面向通用任务场景，同时降低了医学图像分割学习对标签数量的依赖。在心脏结构分割和脑组织分割这两个医学图像分割任务中，仅使用少量标签，就取得了非常优秀的效果，有效地降低对医学图像中常规结构分割的成本，展现出其在手术规划、导航等临床任务上的高效性和潜在应用价值。

4) 为了进一步研究无标签情况下的表征学习，提高任意潜在的医学图像分割任务的学习能力，我们设计了一种基于自监督学习的医学图像表征预训练算法。该算法利用医学图像之间的拓扑结构一致性，推动模型学习图像之间的特征一致性，实现了无标签情况下的医学图像表征预训练。在心脏结构、脑结构、冠状动脉分割任务中均取得了更快的收敛性和更高的准确性，展现出了该方法在快速适应目标任务方面的能力，具有提高医学图像分割应用开发效率的潜力。

本文的研究旨在解决医学图像分割学习中的数据效率问题，提出了一系列应用于医学图像分割的数据高效学习方法。这些研究从具体任务出发，逐步提升了算法的数据效率和通用性，为基于深度学习的医学图像分割技术的大规模发展和广泛应用提供了一系列有效的解决方案。

**关键词：** 数据高效学习，深度学习，标签不足，图像表征，医学图像分割

## Abstract

With the development of our society, people's demand for medical resources is growing rapidly, but the amount and quality of medical workers and the amount of high-level medical equipment are growing slowly. This makes the contradiction between medical demand and medical resources more acute. In order to alleviate this medical contradiction, it is urgent to use technology to release the limited medical resources, improve the balance of the distribution, and improve the efficiency of medical resources. Medical image segmentation technology can automatically extract and visualize the target structure in medical images which improves the efficiency and quality of doctors' diagnoses, thus bringing better service for a wider range of patients. Therefore, it has been one of the important research directions in medical health technology. Especially in recent years, artificial intelligence technology (like deep learning) has shown great application potential in the field of medical image segmentation owing to its strong learning ability. However, due to the powerful nonlinear learning ability of deep learning, its representability depends on the learning of a large number of data. It is challenging to obtain high-quality and large-scale medical image segmentation datasets because of the expensive medical image acquisition, low image quality, high professional requirements of annotators, and very time-consuming labeling process in medical images. Finally, the medical image segmentation tasks are facing serious data efficiency problems, which poses a severe challenge to the large-scale development and application of this technology in medical images. These challenges can be summarized as the following two aspects: 1) Representation for Images: Medical images exhibit vast visual variations, and anatomical structures inner the images are insignificant. This will make the deep neural networks have weak representation capabilities for task-dependent features and make them be interfered with by task-independent features with very limited data amount, finally affecting model generalization. 2) Limited Label Amount: The high cost of medical image annotation and the heavy reliance of deep neural networks on supervision make the limited segmentation annotations unable to train a reliable medical image segmentation model, thus extremely limiting the segmentation performance of the trained models.

This thesis focuses on medical image segmentation which works on how to improve the data utilization efficiency of deep learning models in the learning medical image segmentation tasks. It aims at improving the representation ability of learning algorithms for medical images, reducing the dependence on the number of labels, and realizing efficient data learning in medical image segmentation. The research focus of this paper includes the following four

---

aspects:

1) For the task of fine-grained renal artery segmentation with limited labeled data, we propose a segmentation algorithm based on semi-supervised learning, dense bias neural network architecture, and hard region adaptation loss. This algorithm incorporates knowledge from numerous unlabeled images into the learning process, constructs neural networks with strong representation capabilities for renal arteries, and achieves fine-grained renal artery segmentation through an effective loss function. Ultimately, it provides technical support for preoperative planning of renal cancer surgery.

2) For the task of multi-renal structure segmentation with large inter-image distribution variations and insignificant target structures inner images, we develop an adaptive meta-learning algorithm. This algorithm dynamically selects the most suitable image distribution window based on the adaptive search for the task, and it employs image-level dynamic feature fusion for efficient and accurate multi-structure segmentation of renal tumors. Finally, it shows its effectiveness in multi-renal structure segmentation and its potential clinical value in LPN.

3) To further reduce the label requirements and construct a more universal medical image segmentation algorithm framework, we design a registration-based pseudo-label data generation and robust segmentation learning algorithm. This algorithm is tailored for general task scenarios and reduces the dependence of medical image segmentation learning on label quantity. In the heart structure segmentation and brain tissue segmentation tasks, it only used one and five labels and achieved excellent results, which effectively reduces the cost of conventional structure segmentation in medical images. This shows its high efficiency and potential application value in clinical tasks such as surgery planning and navigation.

4) To delve into representation learning without labels and improve the learning capabilities for potential downstream medical image segmentation tasks, we propose a novel medical image representation pre-training algorithm based on self-supervised learning. This algorithm leverages the topological consistency of medical images and drives the model to learn feature consistency among images in the absence of label. It achieved faster convergence and higher accuracy in the heart structure, brain structure and coronary artery segmentation tasks, which demonstrates the ability of this method to adapt to the target task quickly and has the potential to improve the efficiency of medical image segmentation application development.

This thesis aims to address the data-efficiency problem in medical image segmentation and presents a series of effective solutions for efficient learning in medical image segmentation. The studies start with specific tasks and progressively enhance algorithm efficiency and generality, paving the way for the large-scale development and widespread application of medical image segmentation learning.

**Keywords:** Data efficient learning, Deep learning, Limited label amount, Representation of

images, Medical image segmentation



## 目录

<b>摘要</b> . . . . .	<b>I</b>
<b>Abstract</b> . . . . .	<b>III</b>
<b>术语与数学符号约定</b> . . . . .	<b>XI</b>
<b>第一章 绪论</b> . . . . .	<b>1</b>
1.1 研究背景及意义 . . . . .	1
1.2 深度学习简介与研究挑战 . . . . .	3
1.3 本论文主要工作及内容安排 . . . . .	4
1.3.1 主要研究工作 . . . . .	4
1.3.2 结构安排 . . . . .	7
<b>第二章 医学图像分割中的数据高效学习研究现状及进展</b> . . . . .	<b>9</b>
2.1 医学图像分割学习算法及其问题 . . . . .	9
2.1.1 数据量不足 . . . . .	10
2.1.2 数据不均衡 . . . . .	10
2.1.3 特征不显著 . . . . .	11
2.1.4 标签量不足 . . . . .	11
2.2 优化深度学习数据效率的相关设计 . . . . .	12
2.2.1 数据表征 . . . . .	13
2.2.2 标签学习 . . . . .	16
<b>第三章 基于半监督学习和密集偏置网络的肾动脉精细分割算法</b> . . . . .	<b>21</b>
3.1 研究背景和动机 . . . . .	21
3.2 相关工作 . . . . .	23
3.3 方法 . . . . .	25
3.3.1 深度先验解剖 (DPA) 学习 . . . . .	26
3.3.2 密集偏置网络 (DenseBiasNet) . . . . .	27
3.3.3 难区域自适应性 (HRA) 损失 . . . . .	29
3.4 实验设置与结果分析 . . . . .	30
3.4.1 实验方案 . . . . .	30
3.4.2 对比分析 . . . . .	32

3.4.3 消融实验和模型分析 . . . . .	32
3.5 本章小结 . . . . .	36
<b>第四章 基于元灰度自适应网络的肾脏多结构分割算法 . . . . .</b>	<b>37</b>
4.1 研究背景和动机 . . . . .	37
4.2 相关工作 . . . . .	41
4.3 方法 . . . . .	42
4.3.1 灰度兴趣搜索 (GIS) . . . . .	42
4.3.2 元灰度自适应 (MGA) 学习 . . . . .	44
4.3.3 网络细节 . . . . .	47
4.4 实验设置与结果分析 . . . . .	47
4.4.1 实验方案 . . . . .	47
4.4.2 对比分析 . . . . .	48
4.4.3 消融实验和模型分析 . . . . .	50
4.5 本章小结 . . . . .	52
<b>第五章 基于分割配准互补性的少标签通用多结构医学图像分割算法 . . . . .</b>	<b>55</b>
5.1 研究背景和动机 . . . . .	55
5.2 相关工作 . . . . .	58
5.3 方法 . . . . .	60
5.3.1 知识一致性约束策略 (KCC) . . . . .	60
5.3.2 空间-风格采样过程 (S3P) . . . . .	62
5.3.3 混合不对齐正则化 (MMR) . . . . .	64
5.4 实验设置与结果分析 . . . . .	66
5.4.1 实验方案 . . . . .	66
5.4.2 对比分析 . . . . .	67
5.4.3 消融实验和模型分析 . . . . .	70
5.5 本章小结 . . . . .	75
<b>第六章 基于自监督学习的三维医学图像表征预训练算法 . . . . .</b>	<b>77</b>
6.1 研究背景和动机 . . . . .	77
6.2 相关工作 . . . . .	80
6.3 方法 . . . . .	81
6.3.1 几何视觉相似性学习 (GVSL) . . . . .	82
6.3.2 Z形匹配模块 . . . . .	84
6.3.3 基础代理任务 . . . . .	86
6.4 实验设置与结果分析 . . . . .	86

6.4.1 实验方案 . . . . .	86
6.4.2 对比分析 . . . . .	89
6.4.3 消融实验和模型分析 . . . . .	90
6.5 本章小结 . . . . .	93
<b>第七章 总结与展望 . . . . .</b>	<b>95</b>
7.1 全文总结 . . . . .	95
7.2 研究展望：小与大 . . . . .	96
<b>致谢 . . . . .</b>	<b>99</b>
<b>参考文献 . . . . .</b>	<b>101</b>
<b>附录：本文使用的数据集介绍 . . . . .</b>	<b>125</b>
<b>作者简介 . . . . .</b>	<b>127</b>
<b>毕业/学位论文答辩委员会名单 . . . . .</b>	<b>131</b>



## 术语与数学符号约定

$\mathbb{D}$	数据
$\mathcal{N}$	深度神经网络结构、模块、层
$\theta$ 、 $\xi$ 、 $\iota$	深度神经网络内部参数
$\mathcal{L}$	深度神经网络学习目标、损失函数
$f$	特征
$\psi$	仿射变换矩阵、位移向量场
CNN	convolutional neural network, 卷积神经网络
DWT	discrete wavelet transform, 离散小波变换
IDWT	inverse discrete wavelet transform, 离散小波逆变换
DVF	displacement vector field, 位移向量场
ROI	regions of interest, 感兴趣区域
CT	computed tomography, 计算机断层扫描
CTA	CT angiography, CT血管造影
CCTA	Coronary CT angiography, 冠状动脉CT血管造影
MRI	magnetic resonance imaging, 核磁共振成像
LPN	laparoscopic partial nephrectomy, 肾部分切除术
DSC	Dice Similarity Coefficient, Dice系数
ASD	Average Surface Distance, 平均表面像素距离
ACD	Average Centerline Distance, 平均中心线像素距离



# 第一章 绪论

## 1.1 研究背景及意义

党的十九大报告提出“健康中国”战略，强调“加快转变健康领域发展方式，全方位、全周期维护和保障人民健康，大幅提高健康水平，显著改善健康公平，为实现‘两个一百年’奋斗目标和中华民族伟大复兴的中国梦提供坚实健康基础”。然而，十九大报告同时指出“我国社会主要矛盾已经转化为人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾”，在医疗健康领域，人民群众不断增长的医疗健康需求和有限的医疗健康资源之间的矛盾正在日益凸显。随着社会的发展，人们对医疗资源的需求正在快速增长，然而医疗工作者的数量和质量、高水平医疗设备的数量却增长缓慢，这使得医疗需求和医疗资源之间的矛盾越发尖锐。2022年，国务院办公厅印发的《深化医药卫生体制改革2022年重点工作任务的通知》中进一步强调，要“加快构建有序的就医和诊疗新格局，持续推进解决看病难、看病贵问题”。因此，为缓解这一医疗矛盾，迫切需要利用科技手段提高医疗资源分布的均衡性和使用的效率。

随着医学成像系统<sup>[1]</sup>与现代医疗技术的快速发展，医学图像<sup>[2]</sup>对医疗健康产生了革命性的影响。医生和专家能够观测医学图像，以获取人体内部的解剖和功能信息，进而结合疾病的特点产生对临床有益的指导性结论，辅助精确的疾病诊断和治疗。然而，医学图像由于其特殊的成像方式，往往对比度较低，图像中的病灶和重要解剖结构多变且显著性较弱，这使得对图像中信息的解析和理解存在着较大的专业门槛，需要经验丰富的医生花费大量时间进行观察和勾画标注，限制了医学图像的广泛应用和临床潜力。因此，如何有效地从医学图像中精准自动地提取临床感兴趣的区域成为了亟需解决的重要问题。医学图像分割<sup>[3:4]</sup>（如图1-1所示）作为一种能够帮助医生自动化地完成医学图像中目标结构的提取工作的计算机图像处理技术，成为了医疗健康技术的重要研究内容之一。尤其是近几年，以深度学习<sup>[5]</sup>为代表的人工智能技术<sup>[5]</sup>利用其强大的学习能力，从大量的医学图像数据中学习获得了医生的影像分析能力，在医学图像分割方面展现出了巨大的潜力。已有大量面向不同临床场景的医学图像分割学习研究成果被报道，包括肾癌手术辅助<sup>[6:7]</sup>，冠心病诊断和治疗<sup>[8]</sup>，心肌梗死诊断<sup>[9]</sup>，脑功能分析<sup>[10]</sup>等。一旦这些基于深度学习的智能医学图像分割技术成果被广泛应用于各医疗中心，它们将与医生相互配合，提供更高效和精准的治疗手段，从而进一步释放有限的医疗资源，应用于更广泛的患者群体，推进医疗资源的使用效率<sup>[11]</sup>。

虽然深度学习应用于医学图像分割方面已有了一定的研究<sup>[3:4]</sup>，但是该技术的数据<sup>[1]</sup>效率问题使其依赖于大量带有目标分割区域标签的数据，收集和标注成本大，对智

<sup>1</sup>注：本文的“数据”指医学图像及其对应标签。

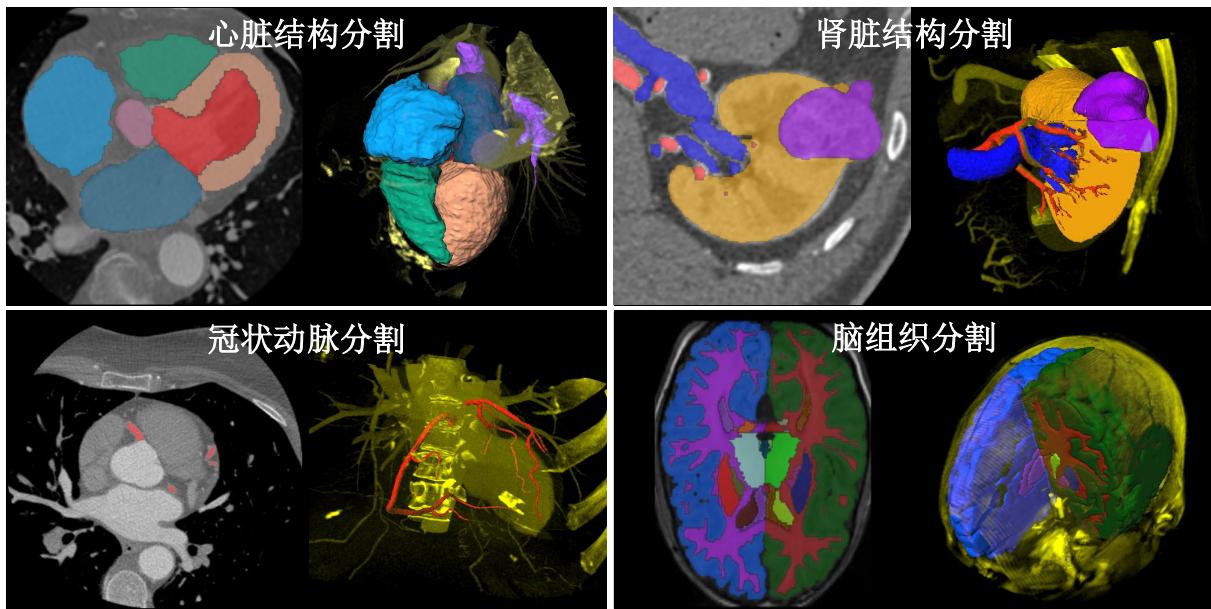


图 1-1: 二维和三维视角下的医学图像及其分割。图中展示了本文中所涉及的四种不同医学图像分割任务，包括，心脏结构分割、肾脏结构分割、冠状动脉分割和脑组织分割。

能医学图像分割技术在医疗健康领域的大规模发展和落地应用<sup>[12]</sup>造成了较大考验。深度学习强大的非线性学习能力使其对图像的表征依赖于对大量数据的学习，而医学图像采集昂贵（千万级专用设备），图像质量较低（非开放式的成像方式），标注者专业要求高（影像专家标注）且标注过程非常耗时（在高维度图像上密集标注）。这些因素导致获取大量高质量数据需要投入巨大的成本<sup>[12]</sup>，最终需要付出非常大的代价才能够构建一个大型标注数据集用于模型训练。此外，医学图像的扫描设备、扫描协议等的多样性，又导致了图像间变化较大，进一步加剧了模型对大型训练数据集的需求。如果数据集无法涵盖大多数的风格变化，那么所训练的医学图像分割模型的泛化能力将受到限制，难以被广泛应用。因此，用于医学图像分割的人工智能技术陷入了不断迭代的图像收集、标注和训练的高成本循环中，大大制约其大规模发展，同时使得医疗数据资源的潜力无法被释放。

2017年初，“人工智能”被首次写入政府工作报告中，科技部将“人工智能2.0”列为科技创新2030重大项目之一。国务院发布的《健康服务产业发展指导意见》也提出了以科技创新和智能化推动医疗健康技术与服务的发展。人工智能是新一轮科技革命和产业变革的重要驱动力量，加快发展新一代人工智能是事关我国能否抓住新一轮科技革命和产业变革机遇的战略问题。人工智能技术的不断进步和应用，作为“头雁”<sup>2</sup>已在多个领域（如物理、化学和生命科学等）催生了大量的新兴交叉学科应用，并逐渐扩展到医疗健康领域，人工智能与生物医学的交叉融合已成为新一轮全球竞争的技术制高点。因此，本论文所研究的面向医学图像分割中的数据高效学习问题隶属于医疗人工智能领域的核心内容，属于面向国际学术发展前沿和国家重大战略需求的前瞻

<sup>2</sup>参考于“中共中央政治局就人工智能发展现状和趋势举行第九次集体学习”。

性应用基础课题，不仅具有理论和应用价值，而且符合“健康中国”和“人工智能2.0”等国家战略需求。

## 1.2 深度学习简介与研究挑战

深度学习应用于医学图像分割方面的数据效率问题对智能医学图像分割技术的广泛应用带来了非常大的挑战。而面对该问题，进一步分析深度学习技术在受到医学图像场景限制情况下的学习问题有着较大的意义。

在获得一个端到端的深度学习模型<sup>[5]</sup>的整个过程中，一个简单的深度神经网络可以被描述为图像 $x$ 被一系列可微的函数 $\mathcal{N}_1, \dots, \mathcal{N}_T$ 嵌套式表征的过程，如公式1.1所示：

$$\mathcal{N}(x; \theta) = (\mathcal{N}_T \circ \mathcal{N}_{T-1} \circ \dots \mathcal{N}_1)(x; \theta) = \mathcal{N}_T(\mathcal{N}_{T-1}(\dots \mathcal{N}_1(x; \theta_1); \dots; \theta_{T-1}); \theta_T), \quad (1.1)$$

其中， $\theta = (\theta_1, \dots, \theta_T)$ 表示这些可微函数中可训练的参数。因此，这使得模型在目标任务中可以通过一些对可微函数的优化方式，如随机梯度下降法，去调节这些参数，以适应当前任务所需要的输出目标。为了达到学习的目的，一个可学习的机制是必不可少的，通常可以描述为一种期望的形式：

$$\begin{aligned} & \min_{\theta} \mathbb{E}_{[x,y \sim \mathbb{D}_{train}]} \mathcal{L}(\mathcal{N}(x; \theta), y) \\ &= \min_{\theta} \mathbb{E}_{[x,y \sim \mathbb{D}_{train}]} \underbrace{(\lambda^s \mathcal{L}_{\theta}^s(\mathcal{N}(x; \theta), y))}_{\text{监督损失}} + \underbrace{\lambda^u \mathcal{L}_{\theta}^u(\mathcal{N}(x; \theta))}_{\text{无监督损失}} + \underbrace{\lambda^r \mathcal{L}_{\theta}^r(\theta)}_{\text{参数正则化}}, \end{aligned} \quad (1.2)$$

其中 $\mathbb{D}_{train} = \{\mathbf{X}, \mathbf{Y}\}$ 是训练所使用的数据集， $\mathbf{X}, \mathbf{Y}$ 是数据集中的图像和对应标签， $\mathcal{L}_{\theta}^s$ 、 $\mathcal{L}_{\theta}^u$ 、 $\mathcal{L}_{\theta}^r$ 分别代表监督损失、无监督损失和模型参数正则化的优化目标，即损失函数，而 $\lambda^s, \lambda^u, \lambda^r$ 则是对应损失在整个学习过程中所占的比重。整个学习函数旨在使被训练的模型能够在训练数据集中所有图像上都预测出优化目标所期待的结果，即分割目标的准确掩膜。最终在应用过程中，被训练好的神经网络将在测试的数据集 $\mathbb{D}_{test}$ 上进行预测，获得输出结果 $\hat{y}$ ：

$$\hat{y} = \mathcal{N}(x; \hat{\theta}), x \sim \mathbb{D}_{test}, \quad (1.3)$$

其中 $\hat{\theta}$ 表示被训练过的模型参数。

由此可见，深度学习整个过程涉及数据 $\mathbb{D}$ ，网络结构 $\mathcal{N}$ ，网络参数 $\theta$ ，和学习目标 $\mathcal{L}$ 四类元素，而深度学习在医学图像分割学习上的数据效率问题也与这四个元素息息相关。本文试图优化这四个元素在医学图像分割学习过程中的表现，以提高深度学习模型性能，降低其学习所需成本，促进医学图像分割学习任务中的高效性。

高效学习（Efficient learning）<sup>[13]</sup>即提升人工智能系统的学习效率以达到降低构建系统成本的目的。由于医学图像数据采集的昂贵，数据质量较低，分割标注专业且耗时，研究人员在医学图像分割中往往更关注数据使用的效率问题，已有一些相关研究被报道<sup>[14;15]</sup>。因此，本文关注如何通过优化上述深度学习中的四类元素，提升图像和

标签的使用效率以降低构建智能医学图像分割系统的成本。本文致力于解决不同医学图像分割任务设置下的以下两个数据高效学习的挑战：

**1) 图像表征的挑战<sup>[16]</sup>**: 由于医学图像的成像设备、协议等的多样性，医学图像间的视觉表现（如图像风格）往往有着较大的变化，而医学图像特殊的成像方式又导致图像中解剖结构在视觉上不显著。因此，在医学图像分割模型的学习过程中，深度神经网络在有限的数据下对任务关注但不显著的特征表征能力较弱，与此同时多变的外在因素也将对图像表征造成干扰，进一步影响分割模型的泛化能力。

**2) 标签不足的挑战<sup>[15]</sup>**: 医学图像分割标注需要专业的影像科医生在高维度图像上进行像素级的标注，这将消耗专业技术人员大量的宝贵时间，导致标注成本高<sup>[12]</sup>。在现实情况下，研究者们难以构建一个足够大的分割标注数据集用于模型训练，而深度学习强大的非线性学习能力使其依赖于对大量数据中监督信息的学习，又使得数据集中有限的标签信息难以驱动可靠的医学图像分割学习。

总的来说，深度学习强大的学习能力为获得一个有效的分割模型提出了大规模精细标注数据集的需求，而有限、低质量、标签不足的医学图像数据又为深度学习提出了数据高效学习的需求。两者之间的矛盾似乎让这项技术陷入一种“死锁”<sup>[17]</sup>状态，既无法提供大量高质量精标注数据进行大规模训练，又无法对有限数据进行高效鲁棒的学习。近年来，在一些自然图像和自然语言领域的研究试图通过构建大规模数据的方式驱动大型深度学习模型走向产品化和应用化，如ChatGPT<sup>[18]</sup>背后是数千TB的数据和人工数据处理量<sup>3</sup>。然而这些成功的道路在医学图像分割任务上是难以复制的，因为医疗数据本身的昂贵性和道德伦理问题<sup>[19]</sup>，无法基于互联网环境进行大范围收集和众包形式的低成本整理标注<sup>[20]</sup>。因此如何优化深度学习算法以提高医学图像分割模型对数据和标签的学习效率，成为了在本课题研究的突破口。

## 1.3 本论文主要工作及内容安排

### 1.3.1 主要研究工作

本文研究医学图像分割中的数据高效学习技术相关问题，重点设计新颖有效的高效深度学习网络架构、数据策略、学习策略和网络参数，包括：1) 基于半监督学习的精细肾动脉分割算法；2) 基于元灰度自适应网络的肾脏多结构分割算法；3) 基于分割配准互补性的少标签通用多结构医学图像分割算法；4) 基于自监督学习的三维医学图像表征预训练算法。如图1-2所示，这项研究旨在逐步提高算法的学习效率，通过提高医学图像分割学习算法对医学图像的表征能力，并且降低对标签数量的依赖，从而致力于解决“基于深度学习的医学图像分割中的数据效率问题”，降低模型学习中的数据成本。研究思路从具体临床相关场景的医学图像分割问题出发，逐步提高所研究算法的适用范围，构建更加通用的医学图像分割学习算法。

<sup>3</sup><https://LifeArchitect.ai/whats-in-my-ai>

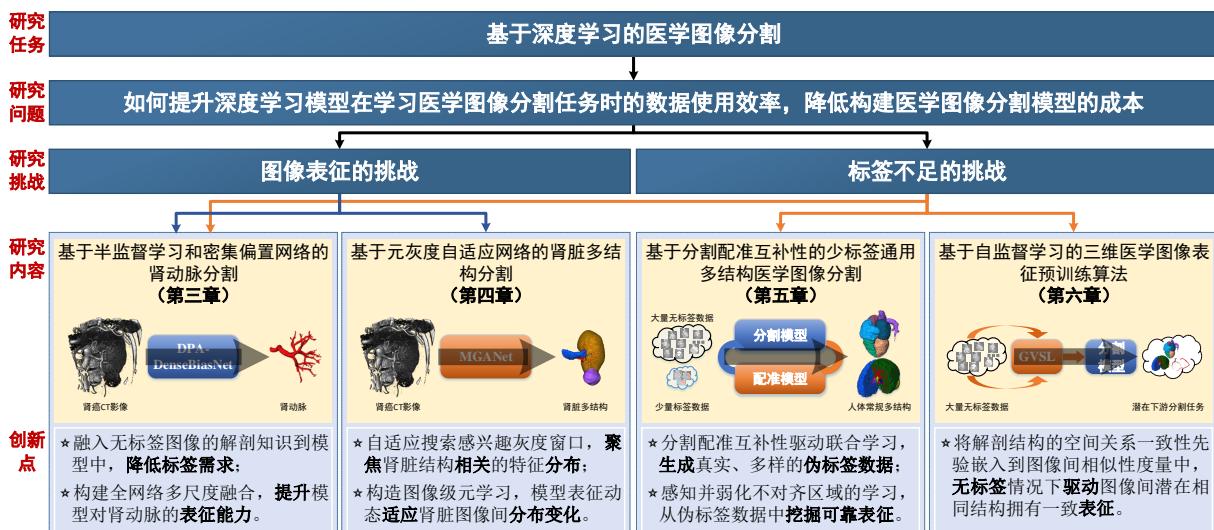


图 1-2: 本文的研究思路和总体技术方案路线图

本文的工作的具体内容包括:

1) 研究了基于半监督学习和密集偏置网络的肾动脉精细分割算法。腹部CT血管造影图像上肾动脉的精细分割是肾癌手术术前计划的重要任务之一,该任务能够为临床医生提供完整的肾动脉树三维模型来辅助其定位肾动脉分支的供血区域,从而制定手术计划。然而,由于肾动脉树自身尺度变化大、不同病人间解剖结构变化大、局部结构细小、图像中体积比小、标签难获得的特点,使其成为了一项具有挑战的任务。在本文中,对于标签数量有限的情况下精细肾动脉分割任务,我们从网络结构( $\mathcal{N}$ )和学习策略( $\mathcal{L}$ )的角度设计了一种基于半监督学习、密集偏置神经网络结构和难区域自适应损失的精细肾动脉分割算法。将更多的无标签图像的知识融入到模型中,构建对肾动脉不同尺度表征能力强的神经网络,最后通过设计一种对肾动脉特征学习更加有效的损失函数实现了肾动脉精细分割学习。我们使用了52个带标签图像和236个无标签图像进行模型训练,获得了88.6%的DSC,为肾癌手术的术前规划提供了技术支持。

2) 研究了基于元学习和灰度自适应搜索的肾脏多结构分割算法。三维集成肾脏多结构分割旨在一次性从腹部CT血管造影图像上提取肾脏、肾肿瘤、动脉和静脉,能够为肾癌手术的精确术前计划和术中导航提供肾癌相关解剖结构的三维视觉模型,是计算机辅助肾癌手术的重要任务之一。然而,由于该任务目标多结构相邻,部分结构间低对比度较低难以区分,而不同图像间又由于肿瘤和人体对造影剂代谢能力相互影响,图像间分布变化较大,使得深度学习模型难以学习到对目标多结构有效表征,同时易受到图像间分布变化的干扰,使得该任务成为一项具有挑战的任务。在本文中,我们从数据分布( $\mathbb{D}$ )和学习策略( $\mathcal{L}$ )的角度构建了一种自适应不同结构分布和图像分布的元学习算法。通过自适应搜索对目标结构分布最有效的多个图像分布窗口,自适应地将分割网络聚焦于任务相关的灰度分布,从而学习用于精细分割的有鉴别能力的表示,并进一步设计了一种图像级元学习方法,根据图像的分布动态地融合来自多个窗

口的特征，以适应图像间的分布变化，从而实现高效准确的肾癌多结构分割学习。我们使用123个肾脏图像并获得了87.9%的DSC，展现出其在肾癌多结构分割中的有效性，同时展现了其在肾部分切除术中的潜在临床价值。

3) 研究了基于分割配准互补性的少标签通用多结构分割算法。人体中拥有多种常规结构，如心脏的多个腔室，脑内的不同脑区，腹部的多个器官结构，在医学影像上对这些结构的自动分割能够为临床医生和研究者提供这些结构的三维视觉模型，从而辅助手术规划、导航手术过程、开展科学的研究，具有非常高的研究价值。然而，这类结构在图像中数量大，标注难度高且耗时，大规模构造此类数据集往往成本较大，使得对多人体常规结构的分割学习模型训练难以实施。我们观察到这类常规结构往往具有空间关系的一致性，可以通过配准算法将不同患者间的此类结构对齐，展现出来配准和分割任务之间潜在的互补性。我们可以利用配准将非常少量的标注图像映射到无标注图像上，以生成大量伪标签图像，进而驱动分割模型的学习，最终实现仅使用非常少量的标签即可获得性能良好的分割模型。在这项研究中，我们从伪标签训练数据生成( $\mathbb{D}$ )和学习策略( $\mathcal{L}$ )的角度，构建了一种可靠的基于配准的伪标签数据生成和鲁棒的分割学习算法。该算法在分割配准互补性的范围内，研究了数据生成的真实性、多样性，以及学习生成数据时的鲁棒性问题，面向更加通用的任务场景同时进一步减少了医学图像分割学习的标签数量需求，仅在1个标签和5个标签的两种设置下，就在心脏结构分割和脑组织分割两个医学图像分割任务上取得了非常优秀的效果。这项工作有效地降低对医学图像中常规结构分割的成本，展现出其在手术规划、导航等临床任务上的高效性和潜在应用价值。

4) 研究了基于图像间几何视觉相似性的自监督表征预训练算法，在无标签情况下提高任意潜在的下游医学图像分割任务的学习能力。在实际场景中，分割标签的标注通常存在滞后性，往往在医学图像数据收集完之后，需要花费大量时间来完成分割标注工作，因此在没有任何标注信息的情况下如何进行预学习以快速服务于潜在的下游分割任务，是一个值得研究的问题。这项研究中，我们进一步利用人体间常规结构存在空间关系的一致性这一先验信息，从优化网络初始化参数( $\theta$ )的角度，构建了一种自监督学习图像间的相似性的模型表征预训练算法。它将解剖结构的空间关系一致性先验嵌入到图像间相似性的度量中，利用配准学习驱动图像间可能为相同结构的区域拥有近似的表征。我们同时考虑了图像间的全局一致性和局部一致性，将其分为相互耦合的两种学习任务，从而驱动模型同时学习到对全局和局部不同语义粒度特征的表征能力，为不同尺度分割对象的下游分割任务都提供良好的迁移性能。我们在心脏结构、脑结构、冠状动脉分割任务中均取得了更快的收敛性和更高的准确性，展现出了该方法在快速适应目标任务方面的能力，具有提高医学图像分割应用开发效率的潜力。

数据效率问题是限制基于深度学习的医学图像分割模型在实际应用中巨大阻碍，数据高效学习算法取得的进展将推动医学图像分割模型在更广泛的医疗场景中发展，尤其是随着算力逐渐充足，计算机基础设施逐渐健全，进行数据高效学习算法研究能

够进一步解放医疗资源，提高医疗水平，推进生命健康进步。

### 1.3.2 结构安排

本文关注医学图像分割的数据高效学习算法，具体从图1-2中的四个任务进行研究。其中第三章，我们设计了一种基于半监督学习的肾动脉分割算法，在标签数量受限的情况下获得更加有效的肾动脉精细分割；在第四章，我们提出了一种基于元学习的肾脏多结构分割算法，面对训练图像间分布变化大，图像内目标结构特征不显著的情况下获得更加精准的肾脏多结构分割结果；在第五章，我们提出了通用的少标签医学图像分割学习算法，进一步降低训练分割模型所需的标签数量，在多个场景中仅使用几个标签即可获得性能良好分割模型；在第六章，我们提出了一种医学图像自监督表征学习算法，为仅有图像没有标签的情况下构建了一种模型预训练的方法，作为模型的初始化参数促进任意潜在下游医学图像分割学习。本文结构安排如下：

第一章对本文选题背景、研究内容和研究挑战进行系统性地分析。

第二章对医学图像分割相关研究现状进行回顾和总结，最后介绍了高效学习的相关研究。

第三章研究了一种基于半监督学习的精细肾动脉分割算法，结合了半监督学习、高效特征融合网络结构和难特征挖掘损失函数的设计，实现了高效的精细肾动脉分割。

第四章研究了基于元灰度自适应网络的肾脏多结构分割算法，结合了元学习和数据分布优化的设计，实现了高效的肾脏多结构分割。

第五章研究了基于分割配准互补性的少标签通用多结构医学图像分割算法，通过优化生成和学习伪标签数据的过程，提升分割模型在缺乏大量标签情况下的学习效率。

第六章研究了基于图像间几何视觉相似性的三维医学图像表征预训练算法，通过自监督学习图像间区域的对齐驱动潜在相同结构区域拥有近似的表征，从而为下游医学图像分割任务提供拥有强大表征能力的初始化参数。

第七章总结了本文的研究内容，讨论了本文的贡献，并就高效医学图像分割学习未来可能的研究方向做了简要的介绍。



## 第二章 医学图像分割中的数据高效学习研究现状及进展

### 2.1 医学图像分割学习算法及其问题

图像分割<sup>[21:22]</sup>是指将一幅图像划分为若干个互不重叠的子区域，每个子区域内包含具有相似属性或特征的像素，是图像处理领域中的重要研究方向之一，是计算机视觉应用的基础之一。在许多实际应用领域中都有着重要的应用，如医学图像分析<sup>[3]</sup>、自动驾驶<sup>[23]</sup>、机器人视觉<sup>[24]</sup>等。根据分割方法的不同，图像分割可以分为传统方法<sup>[21]</sup>和深度学习方法<sup>[22]</sup>两类。传统方法主要包括基于阈值<sup>[25]</sup>、区域生长<sup>[26]</sup>、边缘检测<sup>[27]</sup>等。这些方法通常基于图像的局部信息和底层特征进行分割，处理速度较快，但对噪声和图像质量较敏感，分割效果也受到算法和参数的限制。随着数据量的不断累积，计算机硬件水平的提升，深度学习方法<sup>[5]</sup>以其强大的学习能力和任务适应能力，在图像分割领域中的应用也日益广泛。该方法主要基于神经网络，如FCN<sup>[28]</sup>、U-Net<sup>[4]</sup>等，利用深度学习模型对图像强大的层次化表征学习能力，从而对数据和标签进行学习以适应在不同场景下的图像分割能力，对噪声和图像质量具有较强的鲁棒性。

医学图像分割<sup>[3:4]</sup>是指从医学图像中提取临床医生感兴趣的目标结构，是医学图像分析和诊断中的重要前处理步骤。该任务的目的是提取出具有特定生物学意义的区域，例如肿瘤、器官和病变等<sup>[7]</sup>，从而为医生提供更准确的诊断和治疗指导，也为医疗影像的自动化分析提供图像处理基础。医学图像分割是图像分割技术在医学领域中的应用，其方法通常也包括传统方法和深度学习方法。传统方法上，针对一些特殊的结构，通常会采用特定的更加有效的方法设计，如针对血管这类细小结构<sup>[29]</sup>，通常会采用一些基于中心线<sup>[30]</sup>的或者区域增长的<sup>[31]</sup>分割方法以适应血管细长的结构。近年来，由于深度学习方法在图像分割领域获得的巨大成功，也在医学图像分割上被广泛地研究<sup>[12]</sup>。如图2-2所示，基于深度学习的医学图像分割算法将图像输入深度神经网络中，学习对每个像素所在中心感受野下特征的层次化表征，进而预测像素级特征所属类别，实现对图像的分割<sup>[4:32]</sup>。基于深度学习的医学图像分割研究往往关注解决深度学习方法在医

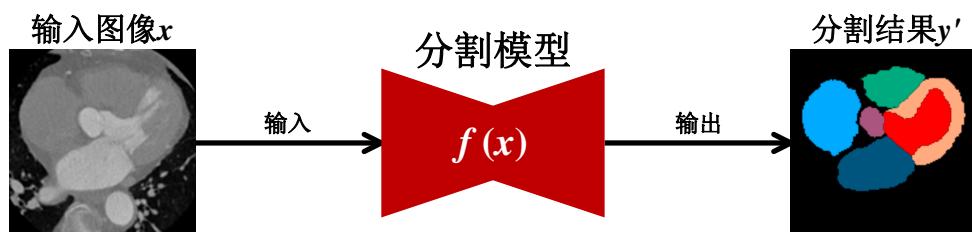


图 2-1: 基于深度学习的医学图像分割示意图。输入一张医学图像，分割模型输出对目标区域的分割结果，图中以CT图像上的心脏结构分割为例。

学图像上的独有挑战，本文关注其中的数据相关的挑战，主要包括以下几个方面：

### 2.1.1 数据量不足

医学图像由于需要专业昂贵的扫描仪器在病人的配合下进行采集，往往非常稀有和昂贵，而医疗数据的私密性也使得这些数据难以在开放平台上获取，因此无法像自然图像一样被大规模地收集。这使得医学图像数据集往往较小，其分布无法代表真实世界的数据分布，最终导致模型过拟合到缺乏代表性的数据分布上，在真实世界医疗场景的应用中表现较差。许多方法采用数据增强<sup>[33]</sup>的方式，如旋转、缩放、弹性形变、灰度变换等，来增加数据的多样性，从而增加训练数据分布的覆盖范围，促进模型在测试时的表现，然而这也引入了人工构造增强操作时的与真实分布不符的噪声。更进一步的，一些方法利用生成对抗网络（GAN）<sup>[34]</sup>来生产高质量的数据以实现数据集的扩充，是一种新型的数据增强手段<sup>[35]</sup>。然而GAN本身的训练不稳定性使生成的数据存在严重的失真问题，尤其是在对细节要求更高的医学图像上，这使得生成数据分布与真实数据分布不符，最终大大限制模型的性能。除此以外，一些方法采用迁移学习<sup>[36]</sup>的策略，将原本就有着较好表征能力的模型迁移到目标任务和目标数据分布上，使其无需很多数据训练即可获得较好的泛化能力。但是获得一个可供迁移的预训练模型，是高成本的。还有些研究试图在模型训练过程中加入精心设计的正则化操作<sup>[37]</sup>，从而是模型在数据量较小情况下不易发生过拟合问题，然而尽管这使得模型更容易表征一些易学习的显著特征，在一些脆弱敏感的分割任务中，如细小血管分割<sup>[38]</sup>，模型将更难以学习这类特征。

### 2.1.2 数据不均衡

医学图像分割中数据本身的分布由于任务目标和人群等原因导致了非常大的不均衡性，这会使得分割模型倾向于部分优势的数据特征和类别，导致在少数类别或罕见特征对象上分割结果较差。具体的，该问题分为两个方面：**a.** 类别级不均衡<sup>[39]</sup>。医生往往关注临床相关的一些细节结构，如细小的病灶，血管等，这些细小区域与较大的脏器和大范围的背景区域相比有着非常大的体积数量差异，这形成了不同类别对象的标签数量的不均衡，导致模型偏向于将分割结果预测为多数类别。一些方法试图通过在损失函数上加权或者设计对来别不均衡鲁棒的损失函数<sup>[40]</sup>来削弱类别级的不均衡问题，而另一些方法通过数据的过采样或者欠采样来实现类别的均衡化<sup>[41]</sup>，尤其是采用数据生成的方式调和类别的数量<sup>[42]</sup>，使得类别不均衡问题进一步得到缓解。然而，这些方法依旧难以解决少数类数量非常少的情况下的模型学习，即长尾问题<sup>[43]</sup>。当某类别数量非常少时，其特征分布无法覆盖真实世界该类别的特征分布，使得模型无法泛化到这些少数类上。**b.** 特征级不均衡。不同的器官和病变具有不同的形状和大小，并且在不同的病例中可能存在大量的变异性，如有些特殊的癌症，变异的解剖结构等，这些罕见的人体特征反应在医学图像上是图像特征的不均衡，这使得模型无法在学习过程

中对这些特征进行有效拟合，最终导致其在测试时对这些特征对象性能不佳，并且与类别级不均衡相比，特征级不均衡未知地潜在存在于数据集中，无法通过人工的方式去除。一些被用于解决类别不均衡的方法也常应用于特征级不均衡，如一些自动化调节损失中对各类别权重策略<sup>[44]</sup>，以及采样难学习的对象给予更多的学习机会<sup>[6]</sup>，这些本质上是给予了那些难学习的少数特征更高的权重和数据量，从而让模型多学习这些少数特征。另一些方法通过设计更加有效的前处理策略，在将数据输入模型前就将这些不均衡特征统一，使得模型能够平等地对待训练数据，但这也潜在地丢失了一些有效的信息<sup>[45]</sup>。

### 2.1.3 特征不显著

医学图像由于其成像特点往往使得，一方面图像内包含噪声和伪影等噪声特征干扰分割模型的学习，另一方面任务感兴趣的解剖结构对比度低（如CT图像上的软组织）和分辨率低（如厚层MRI图像中的细小结构），限制分割模型表征这类对象的特征。一些方法采用提升图像质量的策略，通过超分辨率<sup>[46]</sup>、去噪成像<sup>[47]</sup>、高对比度图像合成<sup>[48]</sup>等策略，将原本在图像中不显著，受干扰的特征增强，使得深度学习模型能够更容易感知到这些不显著的特征，从而增强其表征能力。但是这些图像优化和生成的方法一方面需要额外的训练，增加了计算成本，另一方面也带来了错误累积问题，容易将图像生成过程中引起的误差引入后续的分割学习中，干扰分割性能。另一些方法试图引入一些人工设计的先验知识，通过改善模型框架<sup>[38]</sup>和学习策略<sup>[7]</sup>来增强模型对目标不显著特征的感知能力，从而增强其表征能力。但是一方面人工设计的先验知识往往将额外引入人为的与真实情况不符的噪声，干扰模型学习，另一方面也增加了模型设计的难度与通用性。

### 2.1.4 标签量不足

医学图像分割需要拥有丰富影像学专业知识和经验的专家在高维图像（如3D的CT图像，MRI图像等）上花费大量时间进行逐层缓慢地标注，这使得单位图像的标注成本大，难以获得大型标注数据集。半监督学习是常用于解决该问题的方法<sup>[15]</sup>。这类方法仅为整个数据集提供部分标注信息，试图利用无标签的数据信息引导模型学习到数据的后验分布<sup>[49]</sup>，从而减少训练所需的标签数量。近些年，一些基于一致性<sup>[50]</sup>、不确定<sup>[51]</sup>、多任务学习<sup>[52]</sup>等的半监督学习算法被设计以实现对标签信息更高效的利用，从而降低标注的成本。一些研究关注主动学习<sup>[53]</sup>和交互式分割<sup>[54]</sup>，将人与分割学习耦合，通过分割模型与标注者的互动来降低标注所需要花费的工作量。其中主动学习关注如何从数据集中筛选出难分割的需要标注者参与的案例，而交互式分割关注流程设计，试图让标注过程成为半自动的形式，从而进一步降低标注者的工作量。除此以外，一些研究试图通过弱监督学习<sup>[55]</sup>，将图像级标签<sup>[56]</sup>、带噪声标签<sup>[57]</sup>等低质量不可靠的标注信息引入分割学习过程中，再通过一些人工设计的策略提升模型对这些标签中噪

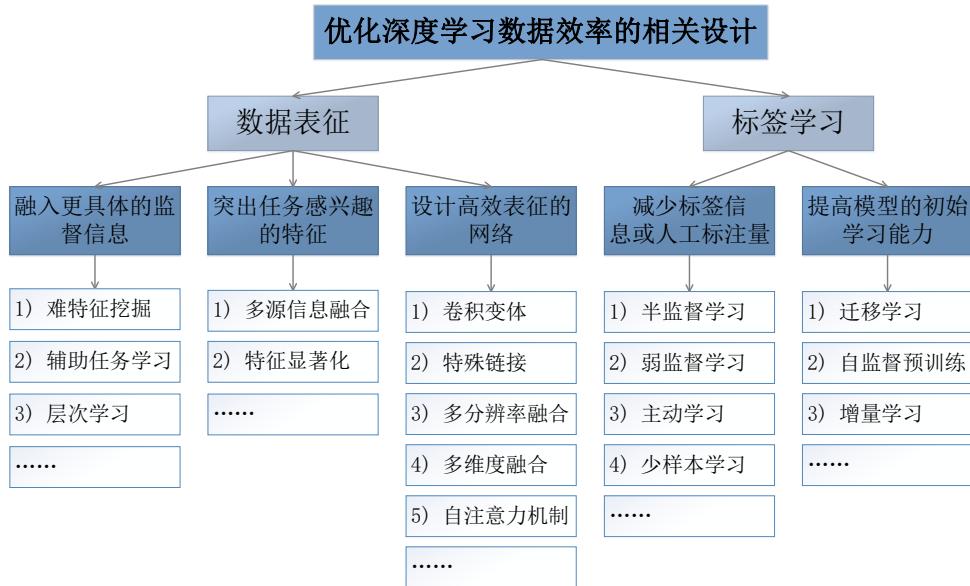


图 2-2: 优化深度学习数据效率的相关设计总览图。

声信息鲁棒性，从而在精细标签不足的情况下也能够获得良好的分割模型。当然以上的三种方法在机器学习层面通常可以相互转换，其本质是引入人为设计的策略和知识，让模型更高效地利用已有的标注信息，学习得到更好的泛化能力。

**讨论** 总的来说，基于深度学习的医学图像分割在当下仍旧是具有挑战的任务。当前的困境在于获得足够大的精细标注数据集成本高昂，无法通过大数据训练的方式让深度神经网络通过反复的训练拟合来获得对数据特征的有效表征。因此，正如前文所述，在标签、数据不足的情况下，如何进行具有较强泛化能力的分割学习是当下大量研究所关注的开放问题，也是本论文试图解决的重要问题。本文的第三和第四章关注在数据质量较低，标签较少的情况下如何提升模型对医学图像的表征能力，重点解决肾癌影像分割中的数据不均衡、特征不显著和标签量不足的问题。本文的第五和第六进一步专注于标签量不足的问题，研究少标签学习和自监督学习，设计通用方法致力在一类分割任务中都取得有效的性能。

## 2.2 优化深度学习数据效率的相关设计

医学图像分割学习的数据效率问题来自于深度学习强大的非线性拟合能力对大量数据的需求与医学图像数据收集和标注的高成本导致的精标注数据量有限之间的矛盾。基于本文第一章的介绍，通过对深度学习方法的优化设计能够为该问题提供一种有效的解决途径，因此本章将从数据表征和标签学习两个角度介绍医学图像分割学习中的优化深度学习效率的相关设计。

## 2.2.1 数据表征

提高数据的表征效率的目的是让深度学习模型更容易地提取到图像上的与任务相关的特征，降低模型过拟合到任务无关特征的风险。相关方法在传统的数据挖掘和机器学习中被广泛研究<sup>[58]</sup>，通常被称之为“特征选择”，相关研究通常采用相似性<sup>[59]</sup>、信息论<sup>[60]</sup>或统计<sup>[58]</sup>的方法筛选获得更加有利于学习的数据特征，从而让模型更容易地学习可靠的性能。在本课题相关研究中，由于深度学习<sup>[61]</sup>强大的非线性拟合能力，只需要通过大量的数据就能够学习到对任务相关特征的提取和表征能力，因此现有的深度学习研究往往不关注数据特征的选择问题。然而当数据量不足时，少量的数据容易使得模型过拟合到与任务无关的特征上，由此限制了模型的泛化能力。尤其是在医学图像上<sup>[62]</sup>，由于医学图像受限于采集过程，通常图像对比度较低，与任务相关的特征不显著<sup>[3]</sup>，再加之标注的专业性与采集的昂贵性问题，使得数据规模较小，最终导致模型非常容易发生上述的过拟合问题。

一种提升数据表征效率的思路是融入更具体的监督信息来引导模型学习如何表征更加与任务相关的特征，根据融入的监督信息与主任务的关系，在本文中我们将其分为难特征挖掘法、辅助任务学习和层次学习：

**1) 难特征挖掘法<sup>[63]</sup>**：该方法旨在动态地对不同样本或特征预测结果的损失进行加权，使网络更加关注数据集中难以学习的特征，从而使模型更关注与任务相关的特征，降低对那些易学习但与任务相关性较低特征的过拟合风险。常见的难负挖掘<sup>[63:64]</sup>、Focal损失<sup>[44]</sup>等对数据或特征学习损失经行采样或加权的策略使得模型能够更加高效地学习难学习但与任务相关的信息。一些均衡数据类别的做法，如DSC损失<sup>[65]</sup>，样本重采样<sup>[66]</sup>等也是在对数据集中属于少数类样本的学习进行人为地增强，从而提高模型对少数样本特征的关注度。但是这些方法缺乏对任务相关特征的直接监督，性能的提升往往是因为缓解了特征的不均衡问题，模型本身依旧容易过拟合到无关特征上。

**2) 辅助任务学习<sup>[67]</sup>**：该方法在学习主任务的同时学习一个对任务感兴趣特征更加具体的辅助监督任务，以多任务学习<sup>[67:68]</sup>的形式引导模型去感知捕获这些信息，从而获得更加高效的数据特征学习。他们利用任务之间的潜在关联性，通过共享参数的方式利用辅助任务引导模型去学习主任务的同时，更好地感知主任务可能感兴趣的特征。在医学图像分割学习中，一些研究同时学习分类任务与分割任务<sup>[69]</sup>，希望模型能够在全局上学习对图像中整体信息的分类，在局部上学习对目标对象的分割，从而引导分割任务更好的获得对全局信息的表征。但是这类方法往往需要考虑辅助任务的可靠性，一旦辅助任务产生的梯度过大，很可能产生“喧宾夺主”的问题，使模型更关注辅助任务的学习而限制主任务的性能。

**3) 层次学习<sup>[70]</sup>**：该方法将学习目标任务分解为多个层级任务的学习过程，从而在学习目标任务前逐步地学习与该任务相关的更底层任务，逐步引导模型学习对目标任务的有效表征。神经网络能够在层次化地学习过程中先学习不容易发生过拟合的底层简单特征，再慢慢拟合更加高级但难学习的任务相关特征，从而更容易学习对任务相

关特征的表征，提高学习表征的效率。由于层次学习更加贴合人们认知事物的由浅入深，由局部特征到全局语义的特点，在强化学习<sup>[71;72]</sup>，分类任务<sup>[73;70]</sup>等难度较大的任务中被深入研究。在医学图像上，已有一些研究通过构建疾病类型的层次化关系，使模型更容易捕获类别特征，取得了更加鲁棒和高效的分类性能<sup>[74;75]</sup>。

另一种思路类似于传统的特征选择，通过对输入数据进行额外的预处理，使得输入数据自身的拥有更显著的任务相关特征和更少的任务无关特征，从而让模型更容易拟合到重要的特征上，同时也从源头避免无关特征的干扰。

**1) 多源信息融合<sup>[76;77]</sup>**：这是一种通过增加更多额外信息的方式来让输入数据中与任务相关的特征更加丰富和显著，由此促进模型对这些信息的感知。该方法通常将来自不同数据源<sup>[78]</sup>、不同模态<sup>[76]</sup>或不同视角<sup>[79]</sup>的信息进行整合和融合，利用深度学习模型进行联合学习或协同训练来进行学习。因此，这类方法可以充分挖掘多源数据之间的相关性和互补性，从而获得更显著的任务相关特征以提高模型的精度和泛化性能。但是，由于多源数据中也包含了更多的任务无关的冗余信息，这也潜在地增加了无关特征干扰的风险。

**2) 特征显著化**：这是一种通过减少任务无关信息来凸显出输入数据中任务相关特征的方法，由此帮助模型更容易表征这些相关特征。在医学图像中最简单的方法是通过改变窗宽窗位<sup>[80]</sup>来使图像灰度分布处于感兴趣对象显著的范围内，从而使得在视觉上这些对象和区域拥有更高的显著性<sup>[7]</sup>。当医学图像中感兴趣特征本身显著性较弱时，一些方法<sup>[81]</sup>先通过某种方式优化图像质量，来提升图像上感兴趣特征的显著性，然后进行后续任务的学习。另外一些研究<sup>[82]</sup>采用目标检测算法提前将任务感兴趣区域裁剪出来，从而去除大量与任务无关的区域，使得图像中的主要信息为任务感兴趣信息，进而帮助模型更加高效地学习表征。

除了对图像的优化和对学习过程的精心设计外，另一种思路是通过针对性地设计网络结构，从而获得对任务相关特征更加高效的表征过程，从而促进深度学习模型对这些特征的表征能力：

**1) 卷积变体**：针对任务特点，一些特殊的卷积结构被设计以获得对模型更有效的表征。空洞卷积<sup>[83]</sup>将卷积核在图像网格上以一定距离向外扩张，使卷积核内部存在多个空洞，从而达到扩大卷积感受野的同时保持参数量与计算量不变，从而提升模型对全局特征的表征能力。形变卷积<sup>[84]</sup>在学习卷积核参数的同时学习对卷积核的形变，从而打破原始卷积从图像网状拓扑上的表征模式，通过可学习的形变来获得适应图像中不同结构拓扑形态的表征模式。图卷积<sup>[85;48]</sup>通过某种方式构造一种图结构的拓扑数据，然后在该拓扑上使用卷积进行学习，从而对图数据进行处理。卷积结构的设计与优化相关研究主要集中在卷积拓扑的设计上，无论是空洞卷积、形变卷积还是图卷积，本质上都是改变原始卷积为了适应图像网格所设计的 $k^N$ 的矩形结构（ $k$ 是卷积核边长， $N$ 是卷积维度），使卷积核本身的拓扑拥有更加适应任务的能力，从而在一系列任务上获得更好的性能。

**2) 特殊连接:** 在网络内部采用一些特殊的连接模式, 能够打破单一的前向路径, 构造更加多元化的特征表达路径和梯度回传路径, 从而获得更强大的表征能力。残差连接最初于ResNet网络<sup>[86]</sup>中被提出, 它采用跳跃连接将特征图传递到后续卷积层的输出位置, 并将输出的特征图于传递而来的特征图相加, 构建了一种残差的连接模式。因此梯度可以从残差连接中直接传递到网络浅层, 从而有效地避免梯度消失问题, 为构建更深的网络带来了巨大潜力。密集链接<sup>[87]</sup>进一步将网络中每个卷积层的特征图通过跳跃连接传递到其后的每一层, 从而密集地融合来自不同层次的不同粒度的特征, 进一步增强了网络的表征能力。同时密集的连接也使得浅层与深层之间拥有直接的梯度连接, 最终使得在训练的过程中梯度能够快速地传递到浅层的每一层, 构成了一种隐式的深度监督<sup>[87]</sup>, 从而提高准确性并加速训练。

**3) 多分辨率融合结构:** 多分辨率融合结构学习图像在不同分辨率和不同感受野下的表征, 并进行跨分辨率和感受野的融合, 从而将多感受野和多分辨率信息在整个网络中融合, 从而将图像中多尺度信息融合的同时保持对细节信息的敏感性。典型的HRNet<sup>[88]</sup>采用多分辨率融合的策略, 将网络划分为多个不同分辨率路径, 并在路径间进行密集的特征融合, 从而协调了多尺度特征, 同时在分类、分割、目标检测等多个任务上取得了成功。特征金字塔网络<sup>[89]</sup>也将多分辨率特征融合的思路应用于网络设计, 构建了能够适应目标在不同尺度上变换的目标检测网络架构, 成为了目标检测任务中的基础网络结构。在医学图像上, U-Net++<sup>[32]</sup>也将这种设计思路融入U-Net<sup>[4;90]</sup>网络中, 从而构建了多分辨率的密集融合, 也在多个医学图像分割任务上取得了良好表现。

**4) 多维度融合结构:** 面对医学图像中各向异性分辨率问题<sup>[3]</sup>, 如三维CT图像中x、y轴与z轴方向间分辨率的差异可达到十倍以上, 多维度融合结构将不同维度的卷积表征能力融入到一个网络中, 从而适应图像的维度变化。典型的方法将二维网络与三维网络输出结果融合<sup>[91]</sup>, 从而在网络之间形成多维度融合, 但是在单独的网络中学习多维度的表征无法对于图像内部不同的结构自适应地构造对不同维度的表征。另外一些工作将网络的高分辨率层使用二维卷积, 然后在低分辨率层使用三维卷积, 从而构建网络内部的多维度融合<sup>[92]</sup>, 获得了比基于多个网络的多维度融合更好的适应能力, 但是该策略限定了维度所适用的分辨率阶段, 模型的多维度分辨率适应能力仍旧受限。最新的网格网络(MNet)<sup>[93]</sup>在卷积单位内部构造了兼容多维度的卷积形式, 从而实现了对各向异性医学图像的维度自适应。该网络内部融合了多种不同的二维三维卷积融合路径, 使得在学习的过程中融合了前两者优势的同时拥有了更强大的自适应能力, 从而在各向异性数据上获得了良好的性能。

**5) 自注意力机制(Transformer):** 与卷积神经网络不同, 基于自注意力机制<sup>[94]</sup>的网络不直接使用神经元学习对数据特征的预测, 而是通过神经元学习对数据特征的注意力的预测, 利用注意力作为权重对特征相乘从而间接动态地实现对数据的表征。如此一来, 该过程为每个数据都形成一套独有的注意力, 即独有的对特征的权重<sup>[95]</sup>, 因此拥有对每个数据建模的个性化表征能力, 动态适应能力更强; 其次, 注意力本身的

动态性又增大了网络内部的参数空间，从而大大增加了网络的“容量”，这使其对大数据有着较好的可扩展能力，为通用的大型基础模型打下了基础<sup>[96]</sup>；最后，注意力机制拥有良好的可解释性，能够通过绘制热图的方式了解模型所关注的图像特征，从而为理解模型的行为打下基础<sup>[97]</sup>。基于以上的优势，Transformer模型已经在自然语言处理<sup>[94;98]</sup>中取得了成功，大量成功的案例，如GPT系列<sup>[99]</sup>等，展现出Transformer在学习能力上的巨大优越性。视觉 Transformer（ViT）<sup>[100]</sup>进一步将Transformer应用于计算机视觉相关的研究中<sup>[101]</sup>，它将自然图像划分为 $16 \times 16$ 个小块作为视觉词，使用全连接层对划分的小块特征编码后输入Transformer网络进行学习，并结合位置编码以保持图像的空间信息，取得了自然图像上的有效性能。这种将图像划分小块作为视觉词从而使得图像能够如自然语言一样适用于Transformer，启发了一系列计算机视觉的研究<sup>[102;103;104]</sup>，使得计算机视觉的网络结构也能够和自然语言一样拥有大规模训练的能力。

大量的网络结构融合了上述的设计思路，从而相辅相成构造强大的整体网络。如已经介绍的UNet++<sup>[32]</sup>、HRNet<sup>[88]</sup>、DenseUNet<sup>[105]</sup>、MNet<sup>[93]</sup>等融合了多分辨率、特殊连接、多维度等设计方式，解决深度学习模型学习表征时所需应对的切实挑战。更进一步地，在大量具体场景的医学图像分析任务中，研究者根据这些场景任务的特点，构造针对性的设计方案，以网络架构或网络模块的形式嵌入模型中从而显著提升模型在具体任务的性能<sup>[106]</sup>。这些针对性的工作通常将临床知识或人为观察作为基础，嵌入网络结构中，因此能够有效地提升网络在表征任务感兴趣特征时的能力。与基于卷积的网络结构类似，基于transformer的网络结构也得益于卷积网络中被广泛使用的跳跃连接、多分辨率融合等技术，为整个网络构建更好的表征能力<sup>[107;108;109]</sup>。

## 2.2.2 标签学习

提高医学图像分割模型的标签学习效率是指在标签信息不足的情况下提高模型的学习能力，从而获得可靠的分割性能<sup>[15]</sup>。传统的基于监督学习的分割学习通常需要大量的标签数据来训练模型，但是分割标签的获取成本很高，特别是在医学图像领域，由于数据标注的专业性使得大量的分割标签是高成本的，难以获得的。已有的方法往往通过利用半监督学习<sup>[110]</sup>、弱监督学习<sup>[111]</sup>、迁移学习<sup>[36]</sup>、主动学习<sup>[53]</sup>、自监督学习<sup>[112]</sup>、少样本学习<sup>[113]</sup>等来降低模型对标签的依赖，使得在有限的标签数据下能够取得与使用大量标记数据相当的模型性能。我们首先介绍半监督学习、弱监督学习、主动学习和少样本学习，这四类方法都试图在模型学习目标任务的过程中减少对标签信息或人工标注量的需求：

**1) 半监督学习<sup>[15;110]</sup>:** 半监督学习是一种同时利用有标签数据和无标签数据进行训练的学习方法，相较于仅利用有标签数据进行训练的传统监督学习模式，半监督学习可以通过学习来自无标签数据的大量信息来提高模型的泛化性能。因此这类方法减少了模型学习对精细标注数据的强烈依赖，节省了标注数据成本，已经被大量地研

究<sup>[114;115;116;117;118;119]</sup>。然而，在半监督学习的过程中，由于无标签数据的质量难以保证，半监督学习方法的鲁棒性在一些极端数据情况下较差，容易产生大量有悖于事实的监督信息，从而反过来限制模型的学习。由于半监督学习展现出了良好的缓解智能医学图像分割任务中数据效率问题潜力，也在医学图像中被广泛地研究<sup>[120;121;122;123;124]</sup>。一些研究更进一步的考虑如何加入临床或几何相关的先验信息来构建鲁棒的半监督学习策略，从而约束模型即使面对一些极端情况也能获得鲁棒的学习<sup>[125;126;127]</sup>。然而如何更进一步地减少半监督学习中的标签数据数量，增大无标签数据的作用，这仍然是一个半监督学习中的开放问题。

**2) 弱监督学习<sup>[111]</sup>:** 弱监督学习与半监督学习类似，都是同时利用监督信息和无监督信息一起学习，从而降低模型训练所需的监督信息量，提升模型的泛化能力。但是弱监督学习并不区分有标签数据和无标签数据，而是使用比完整标签更弱的标签信息，例如仅使用图像级别的标签（如分类）完成像素级别的任务（如分割）<sup>[128]</sup>，或部分标注的标签<sup>[129]</sup>等。相比传统的监督学习，弱监督学习可以减少标注成本，但也与半监督学习类似，无监督信息中的噪声将在极端情况下限制模型的准确性。在医学图像分割中，虽然获得精细且专业的标注较为困难，但是一些粗糙的非专业标注是较容易获得的，如一些涂鸦<sup>[130]</sup>、包围盒<sup>[131]</sup>等，因此如何利用这些弱的信息来训练医学图像模型受到了广泛的关注。

**3) 主动学习<sup>[53]</sup>:** 主动学习利用深度神经网络的表达能力和自适应性在学习的同时自动筛选需要标注的数据，从而使得人们只需要标注部分数据集而不是整个数据集，从而大大减少标注数据量，降低标注成本。同时模型训练也无需提前准备好大量的标注数据，在边训练边标注的情况下就可以开始进行学习。医学图像分割学习中已有一些主动学习的研究被报道<sup>[132;133]</sup>，在标注成本高的医学图像场景中有着较大的应用价值。然而，人工的参与也意味着人为不确定性的出现，往往这类模型的标注和训练的交替性使得标注质量无法被第三方即时检查，导致人工标注数据的质量和数量影响模型的性能<sup>[53]</sup>，同时数据选择的策略设计也容易发生误判，导致选择的数据标注价值较小<sup>[3]</sup>。

**4) 少样本学习<sup>[113]</sup>:** 少样本学习是指从带有非常少量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<sup>[113;134;135;136]</sup>。近年来，在少样本学习已经在一些任务中取得了成功，如图像分类<sup>[136]</sup>、视频分类<sup>[134]</sup>、神经架构搜索<sup>[135]</sup>等。由于医学数据的标注成本高昂且数据稀缺<sup>[62]</sup>，少样本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<sup>[137;126;127;125;138]</sup>。然而，正如 Wang 等人<sup>[113]</sup>所指出的，少样本学习的核心问题是在学习中产生的不可靠的优化方向（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。

提高模型的初始学习能力<sup>[139]</sup>也是另一种降低标签需求的有效方法。该方法是一种

利用已有的被训练的模型中的参数来初始化新模型，从而提高模型收敛速度和泛化能力，特别是在样本数据较少的情况下，可以大大提高模型的表现，是一种降低标签需求的有效方法。根据训练这些初始参数的方法和迁移这些参数的目标任务的差异，我们在这里介绍迁移学习<sup>[36;15;140]</sup>，自监督预训练<sup>[141;142;143]</sup>和增量学习<sup>[144;145]</sup>三种方法：

1) **迁移学习**<sup>[36;15;140]</sup>: 基于迁移学习的方法的主要思想是将一个在源任务上训练好的模型的参数应用于新的目标域任务，使得目标任务的学习可以在有着一定表征性能的模型基础上进行。由此，模型可以将在源域中获得的高级表征能力直接迁移到目标域中，从而在目标任务中大大减少训练时间和数据量，并缓解过拟合的问题，提升学习效率并达到更高的性能。然而，一方面由于模型已经学习到了大量的源域表征偏好，这些偏好将被带入目标域任务的学习中，如果源域和目标域之间存在较大的差异，将潜在地限制模型在目标域上的学习；另一方面，由于源域中所训练的参数与网络架构绑定，迁移时无法调整网络架构，因此使得在目标域任务中无法“因地制宜”地调整网络的结构参数<sup>[15]</sup>。在过去的研究中，已有大量的工作将在一些大型数据集，如Imagenet<sup>[146]</sup>，上的训练模型迁移到其他任务中，验证了该参数初始化方法的有效性。

2) **自监督预训练**<sup>[141;142;143]</sup>: 与基于迁移学习的方法不同，自监督预训练关注源域任务采用无标签的数据进行自监督学习，从而学习数据自身的特征来提高模型的泛化性能。然后使用预训练好的参数初始化模型，在目标域有标注数据的任务进行迁移，以达到提高模型性能和学习效率的目的。该方法可以利用大量的无标注数据，避免了需要大量标注数据的瓶颈，在标签数据稀缺或不可用的情况下能够大大提升模型的可用性，同时利用从大规模数据中学习到的具有较强泛化性的表征能力大大提升在目标任务中的性能。然而，由于该方法的大量大规模无标签数据需求，以及较低的无标签学习效率，获得一个自监督预训练模型通常计算成本较高，并且预训练导致的“偏好”问题也同样存在于该方法中。现有的一些方法通常采用生成式或对比式的自监督预训练<sup>[147]</sup>。生成式的方法采用如重建<sup>[148]</sup>、拼图<sup>[149]</sup>、旋转<sup>[150]</sup>等接任务使网络通过图像上下文来判断图像中内容的分布情况，从而模型提前学习到对图像内容的表征能力。而对比式的方法采用对比学习<sup>[151;152;153]</sup>，训练模型学习对象间的相似性与相异性，学习对特征的内在聚类。

3) **增量学习**<sup>[144;145]</sup>: 与前两者不同，增量学习关注对目标任务学习的同时保持原有任务的性能。它不断地在网络中加入新的任务和新的数据，将前置任务上学习获得的参数作为后续任务的初始化参数，并且在学习后续任务时不遗忘前置任务的性能，使得模型逐步进化。因此，该方法主要应用于数据量不断增加或者任务不断演化的场景中，从而利用前序任务中学习到的知识降低对后续任务的数据和标签依赖，同时获得不断提升的模型性能和越来越丰富的模型功能。然而，增量学习也与前两者一样，存在前置任务所学习到的表征偏好使模型容易陷入局部最优解，难以发现在后续任务上更好的全局最优解，同时在学习后续任务时，对于前置任务的灾难性遗忘问题<sup>[154]</sup>也是

该任务的巨大挑战。

**讨论** 总的来说，上文所述的数据高效学习设计的多个方面都在试图解决深度学习过于强大的表征能力与有限的精标注数据、特征不显著的医学图像数据的矛盾，本文研究医学图像分割中的数据高效学习，针对数据表征和标签学习两个挑战本文进行深入研究。在第三章中，本文设计了一种半监督学习算法以提升肾动脉分割学习的标签效率，并构建了一种融合密集连接的新型网络结构，提升模型对肾动脉多尺度特征的表征能力，最后利用一种自动化调整难表征区域学习程度的损失函数以挖掘难以学习的特征，最终获得精细的肾动脉分割模型。在第四中，本文设计了一种自动调节输入图像窗宽窗位是算法以凸显出任务感兴趣目标结构的特征，并进一步动态融合来自多个不同窗口的图像特征以适应对不同图像的表征，最终实现了肾癌多结构分割模型。在第五中，本文融合了配准学习和分割学习，构造了一种少样本学习算法，大大减少模型学习的标签需求。在第六中，本文进一步设计了一种基于人体间拓扑一致性的自监督学习算法，提高了模型的初始学习能力，从而促进下游分割任务的学习。通过融合上述多种数据高效学习的潜在基础方法，能够有效地环节深度学习模型对于医学图像分割任务的数据效率问题，有着较大的研究价值。



### 第三章 基于半监督学习和密集偏置网络的肾动脉精细分割算法

本章从“肾动脉精细分割”这一与临床结合紧密的单结构医学图像分割任务出发，针对如何减少肾动脉精细分割学习中所需标签数量，以及如何提高模型对细小的肾动脉结构的表征效率，这两个问题展开研究。考虑到这一任务的实际情况，本章着力解决了四个与标签数量和数据表征密切相关的挑战：血管尺度变化大、解剖结构差异大、肾动脉体积占比小以及标签数量有限。为了应对这些挑战，本章的研究从网络结构 ( $\mathcal{N}$ ) 和学习策略 ( $\mathcal{L}$ ) 两个角度出发，设计了新的神经网络、新的训练策略和损失函数，以提供更高效的算法用于肾动脉精细分割，具体如下：为了应对标签数量有限和解剖结构差异大的挑战，本章的算法采用去噪自编码器从大量无标签图像中学习通用的解剖结构表征能力，构建一种新型的半监督学习算法，从而减少整体学习过程中对标签的需求，以适应多样化的解剖结构特征。为了克服尺度变化大的挑战，本章设计了一种全网络密集连接的新型神经网络结构，通过整合不同网络阶段的特征，获得了强大的多尺度表征能力。对于肾动脉体积占比较小的挑战，本章设计了一种新型的损失函数，能够有效学习肾动脉这种体积较小且难以分割的目标的特征表征能力。

#### 3.1 研究背景和动机

腹部CT血管造影（CT angiography, CTA）图像上的肾动脉精细分割是腹腔镜下肾部分切除术（laparoscopic partial nephrectomy, LPN）术前计划中最重要的任务之一<sup>[155;156;157;158]</sup>。它旨在分割获得图像中能够到达肾叶间动脉末端的三维肾动脉树掩模，临床医生能够通过它定位每个肾叶间动脉对肾脏的供血区域，从而提前确定在肿瘤切除时影响的肾脏范围，从而构建精准术前规划<sup>[159]</sup>。如图3-1 a) 所示，它能够帮助医生提前知晓对肿瘤供血的叶间动脉（红色箭头）位置获得术中需要夹闭的肾段动脉，制定夹闭的策略，从而帮助手术顺利进行<sup>[156;157]</sup>。

然而，自动化地分割精细的肾动脉是一项具有挑战的任务，因为存在以下几个原因：1) 血管尺度变化大。如图3-1a) 所示，同一患者的不同肾动脉区域之间存在着很大的尺度变化。肾叶间动脉的直径小于1.5毫米，仅为最粗部分的五分之一。因此，我们需要使用多尺度特征表示来适应动脉内的大尺度变化。2) 解剖结构差异大。如图3-1b) 所示，不同患者的肾动脉分支结构、开口和进入肾脏的方式各不相同，存在较大的解剖差异。一项在仅有461个肾脏上的研究<sup>[160]</sup>发现了11种不同的肾动脉结构。由于在数据集中难以覆盖一些独特的解剖结构，因此在有标签数据集较少的情况下模型的泛化能力较差。3) 体积占比小。肾动脉仅占整个肾脏感兴趣区域的0.27%，导致了前景与背景的严重类别不平衡。因此，在训练过程中，图像中的背景区域将产生更高损

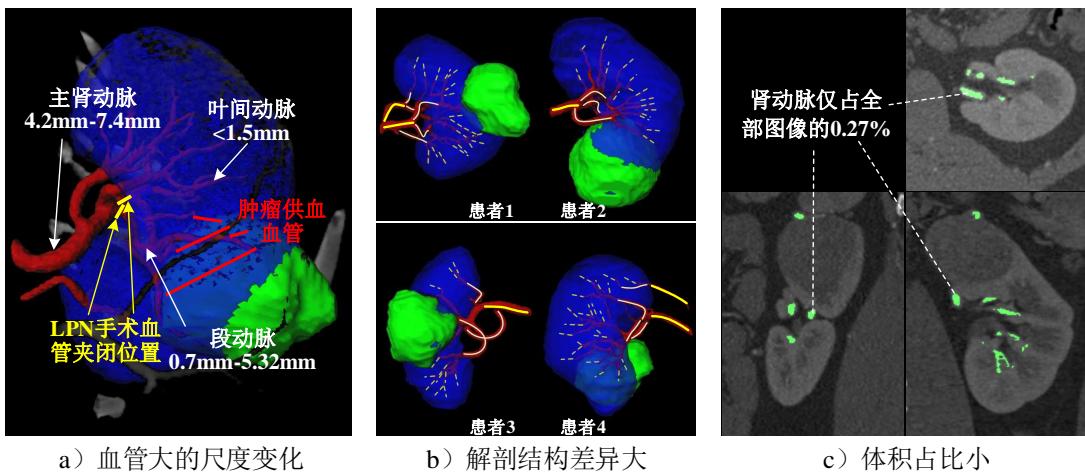


图 3-1: 精细肾动脉分割任务的挑战。a) 同一肾动脉不同区域之间存在大的尺度变化, 肾动脉末端的直径只有主肾动脉直径的四分之一。b) 不同患者之间的动脉分支结构、开口和进入肾脏的方式等不同, 存在较大的解剖结构变化。c) 肾动脉区域仅占肾脏感兴趣区域的0.27%, 前景背景类别数量差异巨大, 导致严重的类别不均衡问题。

失从而使网络更偏向于将像素分类为背景区域, 忽略数量占少数血管区域。4) 标签数量有限。肾动脉呈现出细长的管状结构, 其最细的部分直径可小于1.5毫米, 这种细小的结构有着非常精细的边缘、表面、末端等, 使得手动勾画肾动脉标签非常困难。勾画过程将花费比一般结构更高的标注成本, 这使得精细的肾动脉标签数量有限, 难以训练分割模型取得对结构多变的肾动脉有效泛化能力。

半监督学习可以提高模型的泛化能力<sup>[16]</sup>, 适用于解决解剖结构多变和标签数量较少的挑战, 但是以往的基于半监督学习的分割方法<sup>[121;122]</sup>往往难以适应这项研究中细小的肾动脉结构。这些已有方法通常使用无标签数据生成伪标签标签以进一步学习泛化能力, 这种方法很容易丢失像血管这样的细小结构, 并将这些错误信息作为标签来训练分割模型, 使得模型拟合到错误的信息上, 而肾动脉作为细小又脆弱的分割对象容易受到较大干扰, 导致性能下降。为此本章设计了一种深度先验解剖学习策略, 该策略使用大量无标签数据通过无监督学习训练一个自编码器网络用于从大量无标签数据中学习对多变解剖结构的通用表征能力, 进而将这种表征能力嵌入到肾动脉分割的监督学习模型中, 从而引导模型获得对多标的肾动脉解剖结构的适应能力。因此, 整个框架将第一阶段的无监督学习与第二阶段的监督学习相结合, 构成了一个半监督学习框架<sup>[110]</sup>。

密集连接<sup>[87]</sup>将卷积神经网络中每个卷积层的特征图通过跳跃连接传递到其后的每一层, 从而融合多个不同粒度的特征, 使得整体网络对不同尺度的对象都具有较好的表征能力。然而, 由于大量的特征复用对内存的巨大占用, 基于密集连接的网络的深度和宽度受到了很大的限制。本章设计了一种密集偏置连接结构, 它在特征传输之前先压缩特征图, 以提纯对任务有效的特征信息的同时大大降低特征对内存的占用, 从而保持密集连接的优势。基于此, 本研究构建了整个网络的密集偏置连接架构, 在网

络中形成有效的多分布特征融合，从而适应图像中不同尺度的特征以应对血管结构尺度变化问题。此外，该设计采用的全网络密集连接模式保持了网络中的密集信息流，构建了一种隐式的深度监督<sup>[162]</sup>，提高模型精度的同时加速收敛。

通过过采样少数类别数据和欠采样多数类别数据的方法来重新划分数据，是应对类别不均衡问题的常见方法<sup>[66]</sup>。但在本章的分割任务中，由于类别不均衡来自于图像内部，因此这种方法受到了很大的限制。本研究设计了一种难区域自适应损失函数（“难区域”指图像中难以表征和学习的图像区域，如边缘、细小分支等），根据图像内部的分割质量动态采样损失计算区域，从而使网络聚焦到难以分割的区域（如边缘、表面、末端等），最终自适应地增强对这些区域的分割质量。

本章提出了一种三维半监督医学图像分割框架（DPA-DenseBiasNet）来克服肾动脉精细分割中的四大挑战，并有效地获得肾动脉精细分割掩膜。该框架基于本文所研究的深度先验解剖学习（deep prior anatomy, DPA）策略，密集偏置网络（dense biased network, DenseBiasNet）和难区自适应（hard region adaptation, HRA）损失函数，以引导对模型标签不足情况下学习图像间的解剖结构差异，表征图像内部的尺度变化，聚焦于难分割的区域以动态保持类别均衡。我们工作的详细贡献如下：

- 我们设计了一种新型的三维半监督框架肾动脉精细分割框架，该框架将为LPN手术规划提供三维肾动脉结构的可视化，帮助医生顺利完成术前规划以辅助手术的顺利进行，对肾癌治疗有着巨大的临床价值。
- 我们提出了一种对细小结构表征稳定的半监督学习策略，深度先验解剖学习，以引导模型学习适应更丰富解剖结构特征，从而降低缺少标签的限制。它通过预训练的自编码器提取对解剖结构特征的表征能力，并将其嵌入到分割网络中，从而提升网络对更广泛特征的表征能力。
- 我们构建了一种密集偏置连接方法，该方法通过密集融合不同感受野和分辨率的特征，以获得对多尺度特征的优秀表征能力，从而适应图像中对象尺度的变化。基于这种密集偏置连接，我们进一步设计了一个全网络的密集链接架构——密集偏置网络，以实现这种有效的学习。
- 我们设计了一种难区域自适应损失函数以在训练过程中保持类别均衡，学习难以表征的肾动脉区域特征。该损失函数能够根据分割质量自动地采样用于训练的图像内部区域，使网络动态地关注到难分割的区域，以保持类别均衡并提高难分割区域的分割质量。

## 3.2 相关工作

**血管分割** 血管分割对于疾病诊断和术前规划非常重要，一些研究已经在该任务上取得了成果。一些传统方法<sup>[29]</sup>，如一些区域生长方法<sup>[163;164]</sup>，主动轮廓方法<sup>[165;166]</sup>等，已

经在许多血管分割场景中取得了成功。这些方法需要根据任务场景进行大量的手动设计，一旦一些特殊的数据出现，这些方法将很容易发生失误。随着深度学习的发展，基于深度学习的血管分割方法<sup>[167;168;169;170]</sup>因其速度快、准确度高和泛化能力强而被广泛研究。然而，这些基于深度学习的血管分割方法需要依赖于大量的精细标注数据集，当标签量较小时，它们的性能会受到较大限制。

精细肾动脉分割的研究仍然处于初步阶段，在本章研究工作<sup>[6]</sup>发布前还没有成功的案例被报道，除了一些肾脏相关的多对象分割任务的研究包含了粗糙的肾动脉分割对象结果<sup>[106;171]</sup>。Li等人<sup>[106]</sup>使用400张动脉期CT图像训练一个卷积神经网络来实现包括粗糙的肾动脉在内的多肾结构的分割。而Taha等人<sup>[171]</sup>则提出了一种Kid-Net网络以实现肾脏外部的肾动脉和静脉分割。这些方法对于肾动脉分割的精细程度较低，并且需要依赖较大的训练数据集，无法直接应用于本章的任务。

**半监督学习** 半监督学习利用无标签数据学习额外的特征表征能力<sup>[161;110]</sup>对提高神经网络的泛化能力具着巨大潜力，已经在图像处理领域取得了大量的成功<sup>[114;115]</sup>。在医学图像分割中，一些半监督方法<sup>[121;122]</sup>已经被用来处理有限的标记数据问题。Bai等人<sup>[122]</sup>将条件随机场与全卷积神经网络相结合，使用了240个无标签数据来提高模型的准确性，实现了心脏MRI图像分割。Nie等人<sup>[121]</sup>利用生成对抗网络来获得一种置信度图，进而使用置信图来将无标签数据的分割结果中高置信度的区域作为网络训练的标签以获得有效的半监督分割。然而，在面对脆弱而细小的肾动脉结构式，这些策略所带来的不准确标签将大大影响对血管这种脆弱结构的表征，一旦出现错误监督，血管的学习将受到非常大的干扰，反而削弱了模型性能。本章所设计的DPA策略直接利用特征嵌入方法将来自大量无标签数据的知识融合到监督学习模型中，避免了不准确信息对学习的干扰，稳定地提高了模型的准确性。

**密集连接** 密集连接<sup>[87]</sup>通过优化信息和梯度在网络中的传递来提升模型的训练，但是由于其大量的特征复用对内存的巨大占用，使其难以应用于大型三维医学图像。混合尺度密集网络<sup>[172]</sup>采用了密集连接方法并实现了生物医学图像分割。同样，Li等人<sup>[105]</sup>也设计了一种混合密集连接的U-Net（DenseUnet）实现了肝肿瘤分割。而DenseASPP<sup>[173]</sup>将密集连接与空洞卷积相结合，进一步融合多尺度信息并提高语义分割的准确性。由于这些网络的密集链接结构在重复使用所有特征图时会消耗大量内存，因此输入数据大小受到了较大的限制。特别是在3D医学图像分割任务中，往往需要对数据进行切片或下采样来降低训练时的内存占用，并且网络不能太大。本章所设计的密集偏置连接在特征传输前先压缩特征图，降低特征数量的同时降低特征中的冗余，从而实现在保持密集连接性能优势的同时节省内存占用。该方法为构建全网络密集连接模式成为了可能，从而能够将每个隐藏层都将通过跳跃连接以较短的距离通往损失函数，构成了一种隐式的深度监督<sup>[162;87]</sup>，从而提高准确性并加速训练。

**自编码器网络** 自编码器<sup>[174]</sup>是一种无监督学习的神经网络架构，通过编码压缩特征和解码恢复原图两个步骤，使网络无监督地学习到对解剖结构特征的表征能力，因此可以用于嵌入监督学习以构建半监督学习模式，从而高模型性能。应用自编码器到半监督学习的方法主要有三种。1) 梯度嵌入。该方法将预训练好的编码器网络的特征用作具体监督学习网络的辅助优化目标，从而额外的梯度被嵌入到训练过程中使模型学习对预训练得到的特征表征能力<sup>[120]</sup>。2) 辅助任务。该方法将无监督学习的自编码器网络与监督学习的目标任务网络共享部分参数，并同时训练自编码和目标任务，通过辅助任务的形式促进目标任务的学习<sup>[118]</sup>。3) 特征嵌入。该方法使用预训练的自编码器提取特征，并将提取的特征嵌入到监督学习的神经网络中<sup>[119;175;176]</sup>。不同于前两种方法，特征嵌入方法直接将来自大量未标记数据的知识融合，避免对训练过程产生影响，对本研究中脆弱细小的肾动脉非常友好。因此，本研究使用特征嵌入的方法来构建有效的半监督学习架构。

**类别不均衡** 类别不均衡是指数据集中不同类别之间的数量差异巨大，严重的类别不均衡会导致模型偏向于预测为多数类别而不利于少数类别的性能<sup>[44]</sup>，在医学图像分析中，一些研究已经针对该问题进行了探索。在冠状动脉钙化评分任务中，Wolterink等人<sup>[64]</sup>采用了一种两阶段的难负挖掘方法，先定位到难分类少数类别，然后训练模型倾向于这种难的类别，从而削弱了类别不均衡的影响。DSC损失<sup>[65]</sup>通过先计算不同类别的DSC值后进行平均作为损失值，避免了类别间的相互影响。加权交叉熵给予少数类别更高的权重，使网络更关注这些类别从而使得损失保持均衡，但权重是该损失额外的超参数，难以确定。Focal损失<sup>[44]</sup>在加权交叉熵的基础上增加了一个调制因子，根据网络的预测结果，自动化地使网络更关注难预测区域，从而自动化均衡类别。在本章研究中，我们设计了一个难区域自适应损失，将大部分已经分割好的区域剔除不参与训练过程，使得网络倾向于学习难学习的区域，动态地保持类别的均衡。

### 3.3 方法

如图3-2所示，DPA-DenseBiasNet由三个部分组成：1) 预训练的自编码器用于提取和嵌入解剖结构特征以引导模型适应解剖结构变化，形成了一种稳定的半监督学习过程；2) 基于密集偏置连接构建的密集偏置网络，DenseBiasNet，密集融合多感受野和多分辨率特征图以获得强大的多尺度表征能力以适应图像内目标对象的尺度变化；3) 难区域自适应损失函数（HRA）根据分割质量动态采样损失区域，从而使网络专注于难分割的区域并保持类别均衡。整体的训练过程包括两个阶段：a) 使用大量无标签的数据集（无监督）训练自编码器<sup>[174]</sup>以获得对大量解剖特征的表征能力；b) 来自自编码器的特征嵌入到DenseBiasNet中作为先验信息，以适应解剖结构变化。

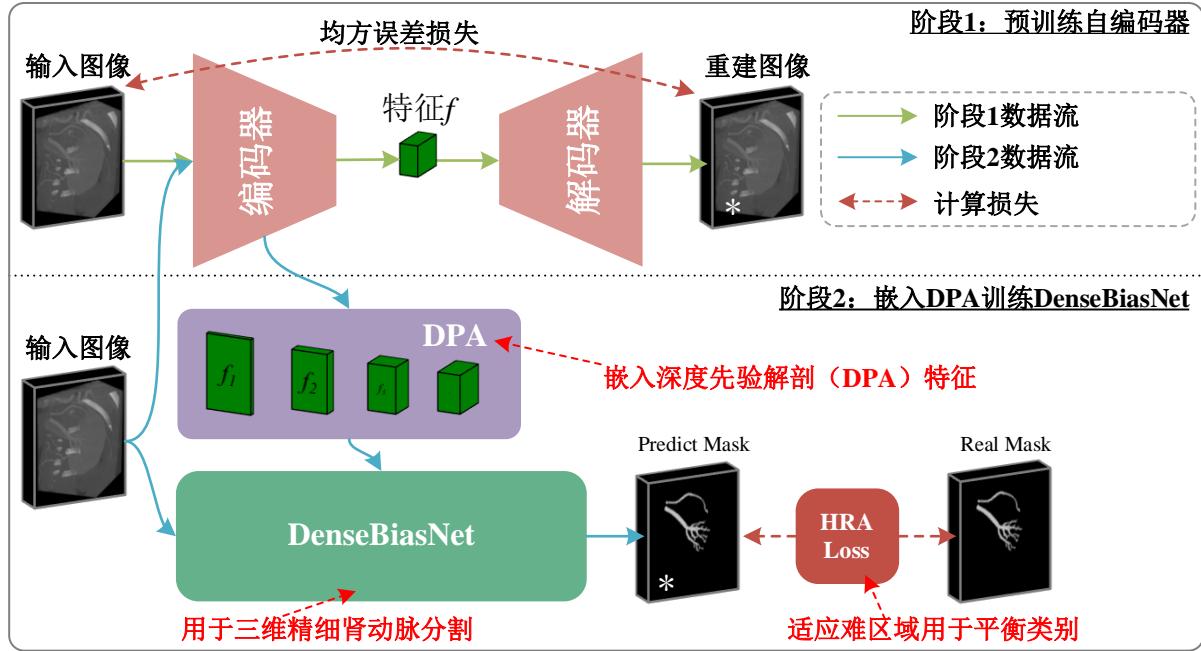


图 3-2: DPA-DenseBiasNet的整体架构。在第一个阶段通过自编码预训练对解剖特征的表征能力，从而提取DPA特征。进而在第二阶段将DPA特征嵌入到并训练DenseBiasNet以实现对肾动脉的有效分割。

### 3.3.1 深度先验解剖 (DPA) 学习

如图3-2所示，DPA的学习过程通过预训练自编码器，然后从编码器的不同深度提取反映解剖信息的特征，并将这些特征融合到分割网络的不同深度以嵌入对解剖特征的表征能力。

**去噪自编码器预训练** DPA学习采用自编码器学习大量无标签数据以获得对解剖特征的表征能力。然后在框架中使用预训练的编码器部分来提取DPA特征。如图3-3所示，DPA学习采用去噪自编码器结构<sup>[177]</sup>通过学习利用图像像素的上文信息修复和重建被破坏的内容以获得对特征的表达能力。整个网络包含一个编码器 $\mathcal{N}_l^{enc}$ 和一个解码器 $\mathcal{N}_\xi^{dec}$ ，编码器通过连续的下采样将图像编码为低分辨率的解剖特征，然后这些特征通过解码器进行上采样以重构和修复原始图像，其中 $l$ 和 $\xi$ 是网络中的参数。具体地，无标签的图像 $x$ 经过一个Dropout层<sup>[178]</sup>，以获得被噪声破坏后的图像 $x'$ ，然后将该图像输入网络。编码器提取并压缩 $x'$ 的特征以获取对图像中重要信息的编码特征 $f = \mathcal{N}_l^{enc}(x')$ ，然后通过解码器重建图像 $\hat{x} = \mathcal{N}_\xi^{dec}(f)$ 。最后，计算重建图像 $\hat{x}$ 和原始图像 $x$ 之间的均方误差 (MSE) 损失值<sup>[174]</sup>以进行网络优化。大量无标签数据将被用于训练该网络以学习对解剖特征的提取能力。

**DPA特征嵌入** 如图3-2所示，DPA策略通过预训练的编码器从输入图像中提取DPA特征，并将其嵌入DenseBiasNet中以适应解剖结构变化。具体地，在模型训练的第二阶

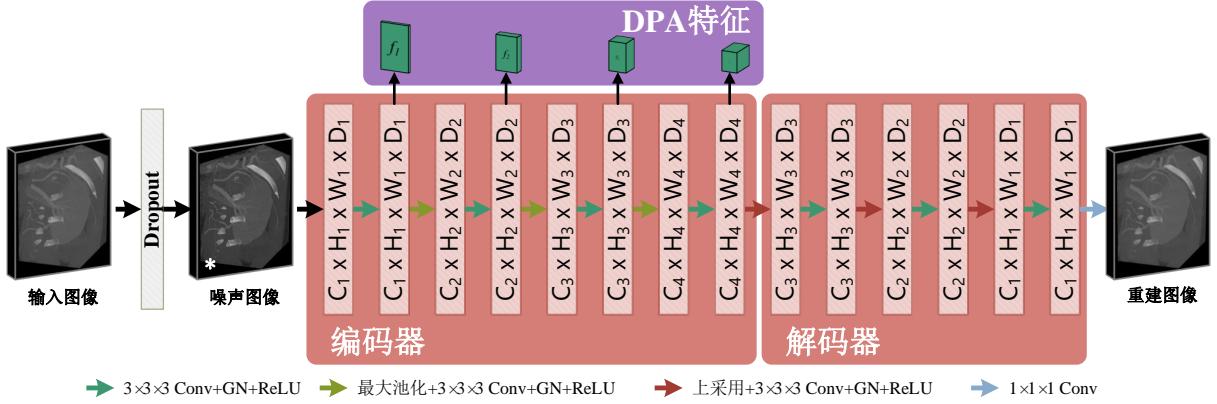


图 3-3: 用于学习DPA表征的去噪自编码器的网络结构。去噪自编码器通过大量无标签数据训练, 以获得对解剖特征的表示能力, 最后从编码器的不同深度输出四组具有不同分辨率的DPA特征用于嵌入监督学习模型。

段, 它将图像 $x$ 输入到在第一阶段被预训练的编码器网络 $\mathcal{N}_t^{enc}$ 中, 提取不同语义级别的DPA特征 $\{f_1, f_2, f_3, f_4\} = \mathcal{N}_t^{enc}(x)$ , 其中1,2,3,4表示编码器的不同分辨率阶段。如图3-5所示, 这些特征被拼接到DenseBiasNet编码路径对应分辨率阶段的特征图中以嵌入对解剖特征的表征, 最终整个DenseBiasNet利用原始图像 $x$ 和这些DPA特征来获得分割结果 $\hat{y} = \mathcal{N}_{\theta}^{DBN}(x, f_1, f_2, f_3, f_4)$ , 其中 $\theta$ 是密集偏置网络的参数。预测结果 $\hat{y}$ 和标签 $y$ 一起计算损失值用于优化训练DenseBiasNet  $\mathcal{L}_{\theta}(\hat{y}, y)$ 以学习对肾动脉的精细分割。

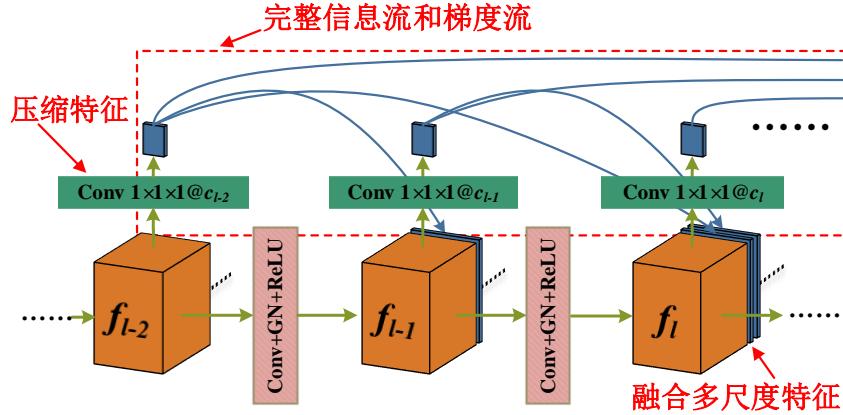
**讨论: DPA学习的属性** 1) 稳定且适用于细小结构。与其他半监督分割方法<sup>[121;122]</sup>采用不精确的优化目标不同, DPA学习直接将来自无标注数据的知识融合到监督模型中, 在目标任务上采用单一准确的优化目标以避免不准确信息的干扰, 从而稳定地提升模型的学习能力。2) 提高泛化能力。DPA将从无标签图像中学习到的特征提取能力嵌入到分割网络中, 以引导模型适应解剖结构特征变化, 有效地提高模型的泛化能力。

### 3.3.2 密集偏置网络 (DenseBiasNet)

DenseBiasNet通过构建一种密集偏置连接, 将每个层中的特征压缩并传输到每个前向层, 从而融合多感受野和多分辨率特征以获得强大的多尺度表征能力, 适应图像内对象的尺度变化。这种密集连接模式的信息流和梯度流在网络中能够高效传输, 从而能够加速训练过程, 提高模型准确性。

**密集偏置连接** 如图3-4所示, 密集偏置连接将每一层中的特征图压缩并传递到其前向的每一层中, 从而融合来自不同尺度不同分辨率的特征信息, 构建对多尺度特征的有效表征。公式3.1展示了密集偏置连接构建网络中第 $l$ 层的计算过程。对于不同分辨率的特征图, 我们使用最大池化或上采样来获得统一尺度的特征图。

$$f_l = \mathcal{N}_l^H(f_{l-1} \circ \mathcal{N}^{C1^3}(f_{l-2}; c_{l-2}) \circ \dots \circ \mathcal{N}^{C1^3}(f_0; c_0)) \quad (3.1)$$



**密集偏置连接:**将每一层输出的特征图压缩并传输到每一个前向层

图 3-4: 密集偏置连接的内部细节。密集偏置连接将每一层的特征图压缩传输到每一个前向层，减少特征冗余，保持信息流和梯度流的完整性，融合多尺度特征。

其中， $f_l$ 是第 $l$ 层输出的特征图， $\{f_0, \dots, f_{l-2}\}$ 是前面各层的特征图。 $N^{C1^3}$ 是一个 $1 \times 1 \times 1$ 的卷积层，将特征图压缩成少量的特征图， $c_0, \dots, c_{l-2}$ 是卷积层输出的特征图数量，即偏置量。 $N_l^H$ 是一个组归一化<sup>[179]</sup>、ReLU、池化、上采样或卷积（Conv）等构成的组合函数用于对这些特征进行进一步运算，我们在实验中将其设置为一个卷积层、一个组归一化层和一个ReLU激活层。“ $\circ$ ”是将特征图拼接起来的操作。

**DenseBiasNet设计** 基于密集偏置连接，DenseBiasNet采用标准的3D U-Net作为基本结构<sup>[90]</sup>构建全网络的密集链接架构。如图3-5所示，整体的编码部分使用连续的卷积和Maxpooling下采样来在提取特征的同时，有效地节省了GPU存储空间，而解码器部分将低分辨率的特征图上采样恢复到具有原始分辨率的特征映射，从而实现了具有相同分辨率的端到端输出。密集偏置连接将不同分辨率和不同感受野的特征图融合在一起，使网络具有不同语义层次的特征表征能力。具体地，每个分辨率阶段都拥有两个 $3 \times 3 \times 3$ 卷积层-组归一化层-ReLU层构成的卷积组相连来提取特征。在网络的最后，使用一个 $1 \times 1 \times 1$ 卷积层和一个Softmax激活层将特征通道数压缩到分割的类别数以实现分割。整体网络的密集连接使得所有层后都直接与损失函数构建联系，形成了一个隐式的深度监督，从而提升学习效率的同时又大大增强了性能。

**讨论：密集偏置连接的属性** 1) 适应尺度变化。该结构将具有不同感受野和分辨率的特征图融合起来，使网络具有强大的多尺度表示能力，从而适应不同尺度的血管。2) 训练加速。来自损失函数的梯度直接沿着密集的跳跃连接直接优化每一层构成了一种隐式的深度监督，加速训练过程。3) 简化的网络冗余。相比于直接的密集连接，密集偏置连接压缩特征图并去除前向传递过程中的大量信息冗余，降低存储空间消耗，更适用于大尺寸的医学图像。

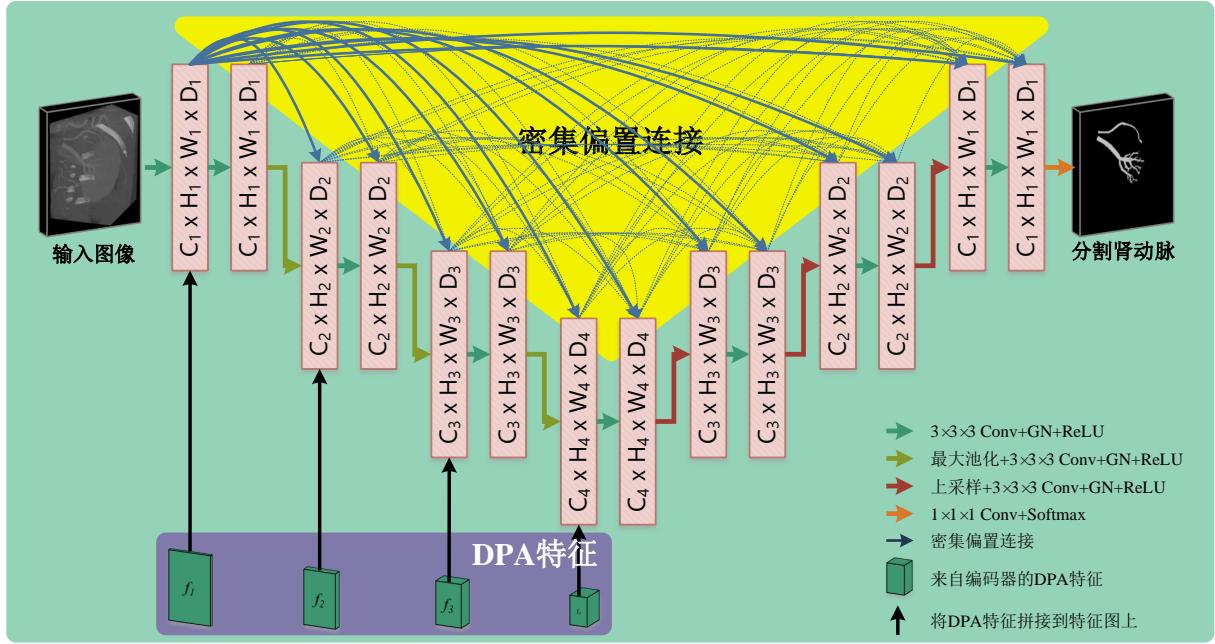


图 3-5: DenseBiasNet 的网络细节。DenseBiasNet 采用密集偏置连接并融合 DPA 特征，取得了有效的肾动脉分割性能。

### 3.3.3 难区域自适应性 (HRA) 损失

难区域自适应损失动态评估并采样图像中难分割的区域，使得网络关注这些难区域，从而动态地保持训练过程中的类别均衡，提高网络的准确性。

**HRA定义** HRA损失根据分割结果的分割质量动态地找到难以分割的区域，即“难区域”，并通过采样针对性地优化这些区域以保持类别均衡。如图3-6所示，红色区域是用于计算损失的像素位置。当使用标准的交叉熵损失训练分割网络时，损失的计算范围覆盖所有的区域，对难易区域都给予相同的关注度。由于血管非常小的类别占比，因此存在严重的类别不均衡问题。当使用HRA损失时，先分割预测结果和真实值之间的 $L_1$ 距离，并通过设定一个阈值以发现难分割的区域如血管边界、细小末端。然后，根据公式3.2所示的方式在这些区域上计算交叉熵损失：

$$\mathcal{L}_{\theta}^{HRA-CE} = -I(y, \hat{y})y \log \hat{y} \quad (3.2)$$

其中， $y$ 是标签值， $\hat{y}$ 是模型预测值。 $I(y, \hat{y})$ 是计算难区域并对这些区域采样的策略，本文采用 $L_1$ 距离，即，如果 $|y - \hat{y}| > T$ ，则 $I(y, \hat{y}) = 1$ ，否则 $I(y, \hat{y}) = 0$ 。超参数 $T$ 用作控制发现难区域的阈值，当 $T = 0$ 时，损失值将在整个图像上计算，即退化为原始的交叉熵损失。

在训练过程中，易分割的区域，如大块的背景区域，会很快被神经网络所拟合，因此当使用HRA损失时，大量的容易拟合的背景区域将被丢弃，从而使得难以学习的血管区域被学习，有助于取得类别均衡。随着模型的训练，如果网络过于关注难区域并

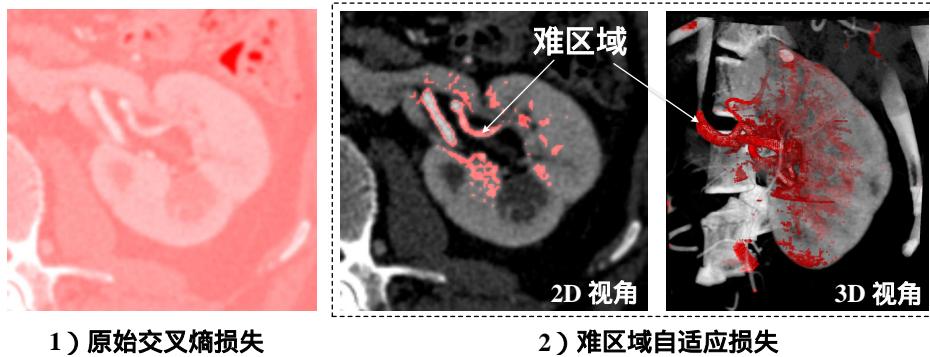


图 3-6: 难区域自适应损失与普通交叉熵损失的差异。1) 标准的交叉熵损失覆盖所有区域, 在本章的肾动脉分割任务中深受类别不均衡问题影响。2) 难区域自适应损失关注难以分割的区域, 从而动态地均衡类别。

开始降低易区域的准确性时, 易区域的 $L_1$ 距离将会增加, 这样网络就会重新关注这些区域。因此, 易区域和难区域的分割质量都将得到动态提升。在本节中, 我们以交叉熵为例, 介绍了HRA损失的定义, 但其形式不限于此。基于图像内采样策略的这种损失函数设计方法也可以用于改进其他损失函数, 如DSC损失和均方误差损失以促进类别均衡。

**讨论: HRA损失的属性** 1) 训练过程自适应。HRA根据分割质量动态选择难的区域计算损失函数。因此, 随着模型收敛, HRA将逐渐排除易于分割的区域, 聚焦于适应于训练实际情况的困难区域。2) 基于特征的类别均衡。与那些直接人工设置加权和采样的方法不同<sup>[66]</sup>, HRA通过动态选择难区域来均衡数据, 使得在均衡数据的同时, 将模型关注的特征表达放在难学习的地方。

## 3.4 实验设置与结果分析

### 3.4.1 实验方案

**数据集** 本研究从江苏省人民医院放射科中回顾性地收集196名肾癌患者的腹部CTA图像, 这些患者的一侧肾脏带有肾肿瘤, 而另一侧肾脏健康。肾脏大小在85.76ml到262.78ml之间, 肿瘤大小在1.75ml到144.82ml之间。肾脏肿瘤类型包括透明肾细胞癌、乳头状肾细胞癌、嫌色肾细胞癌、血管平滑肌脂肪瘤和嗜酸性腺瘤。所有图像都在西门子双源64层CT扫描仪上采集, 并在采集过程中注射了造影剂。图像的体素大小在0.59mm/体素到0.74mm/体素之间。 $z$ 方向上的切片厚度和间距固定为0.75mm/像素和0.5mm/像素。预处理过程中, 首先通过多模板匹配法<sup>[180]</sup>自动地剪裁出大小为 $152 \times 152 \times Z$ 的肾脏感兴趣区域, 共得到392张肾脏感兴趣区域图像。受到专业训练的专家对其中的156张图

表 3.1: DPA-DenseBiasNet与5种方法的对比分析。DenseBiasNet的性能优于三种监督学习方法(V-Net、3D U-Net和Kid-Net)，而DPA策略提高了泛化能力取得了远好于其他两种半监督方法(SemiFCN、ASDNet和CPS)的性能。HRA损失均衡了类别，进一步提高了模型性能。

方法	类型	DSC <sub>%<math>\pm std</math></sub> ↑	ACD <sub>voxel<math>\pm std</math></sub> ↓	ASD <sub>voxel<math>\pm std</math></sub> ↓
V-Net <sup>[65]</sup>		71.5 <sub>±16.7</sub>	3.93 <sub>±2.84</sub>	2.47 <sub>±2.60</sub>
3D U-Net <sup>[90]</sup>		80.1 <sub>±11.4</sub>	3.28 <sub>±2.55</sub>	2.60 <sub>±3.03</sub>
Kid-Net <sup>[171]</sup>	全监督	77.1 <sub>±13.1</sub>	4.36 <sub>±3.56</sub>	3.07 <sub>±3.67</sub>
<b>DenseBiasNet (CE)</b>		86.9 <sub>±8.2</sub>	2.02 <sub>±1.61</sub>	1.45 <sub>±1.87</sub>
<b>DenseBiasNet (CE+HRA)</b>		87.2 <sub>±9.5</sub>	1.87 <sub>±1.49</sub>	1.16 <sub>±1.54</sub>
SemiFCN <sup>[181]</sup>		68.5 <sub>±28.1</sub>	4.06 <sub>±4.59</sub>	3.36 <sub>±5.35</sub>
ASDNet <sup>[121]</sup>		61.7 <sub>±14.9</sub>	6.70 <sub>±4.70</sub>	5.20 <sub>±4.89</sub>
CPS <sup>[182]</sup>	半监督	85.0 <sub>±13.2</sub>	2.09 <sub>±1.70</sub>	<b>1.04</b> <sub>±1.47</sub>
<b>DPA-DenseBiasNet (CE+HRA)</b>		<b>88.4</b> <sub>±7.6</sub>	<b>1.86</b> <sub>±1.42</sub>	1.09 <sub>±1.49</sub>

像进行了半自动标注，确保分割的肾动脉能够叶间动脉末端。本实验使用其中的52张用作为监督训练集，104张作为测试集，其余的236张无标签图像作为无监督训练集。在训练过程中，本实验使用 $152 \times 152 \times 64$ 大小的随即裁剪，x、y和z轴方向的随机翻转和镜像以及 $[-10^\circ, 10^\circ]$ 的随机旋转来增强图像。

**对比设置** 为了展示DPA-DenseBiasNet框架在精细肾动脉分割上的优势，本实验与三种通用的监督学习方法(3D U-Net<sup>[90]</sup>, V-Net<sup>[65]</sup>, Kid-Net<sup>[171]</sup>)和三种半监督学习方法(Semi-FCN<sup>[181]</sup>, ASDNet<sup>[121]</sup>, CPS<sup>[182]</sup>)进行了比较。为了比较的公平性，所有方法采用相同的训练策略和设置在相同的数据集上进行训练。

**实施细节和评价指标** 所有的模型都采用Adam优化器<sup>[183]</sup>进行优化，其学习率、 $\beta_1$ 、 $\beta_2$ 和衰减率分别为 $10^{-4}$ 、0.9、0.999和 $10^{-5}$ 。由于内存的限制，训练时批大小为1，并且在对应的数据集上迭代了200个epoch。本实验使用基于Tensorflow后端的Keras库进行实现，并在一张8 GB内存的 GeForce GTX 1080 GPUs上实施训练。本实验使用平均Dice系数(Dice Similarity Coefficient, DSC)、平均中心线像素距离(Average Centerline Distance, ACD)和平均表面像素距离(Average Surface Distance, ASD)来评估本章所提出的方法。DSC通过计算分割掩膜与分割标签之间的重合程度以评估分割的准确性<sup>[184]</sup>；ACD先根据文献<sup>[185]</sup>中的方法从肾动脉掩膜中抽取中心线，然后计算分割掩膜和标签中心线间的平均Hausdorff距离<sup>[184]</sup>以评估肾动脉分割的位移；ASD先从肾动脉掩膜中提取血管表面，然后计算平均Hausdorff距离以评估肾动脉管腔厚度和表面重合情况。由于计算标准的Hausdorff距离时受到离群点影响较大，对于本研究中肾动脉结构较为敏感，因此我们使用平均Hausdorff距离作为距离度量<sup>[184]</sup>。

表 3.2: 超参数分析。研究密集偏置连接中的偏置量 ( $m$ ) 和HRA中的阈值 ( $T$ )，以分析它们的影响。

$c$	$T$	$DSC_{\% \pm std} \uparrow$	参数量 $M \downarrow$
1	0.1	86.1±9.1	1.30
2	0.1	85.7±8.9	1.41
4	0.1	<b>87.2±9.5</b>	1.61
4	0.2	84.8±10.3	1.61
0.1r	0.1	87.1±7.9	1.63
0.2r	0.1	83.9±11.9	2.10

### 3.4.2 对比分析

**定量对比分析** 如表3.1所示，DPA-DenseBiasNet在三个指标上均取得了最优的性能。DenseBiasNet使用HRA损失时，其DSC、ACD和ASD分别为87.2%、1.87和1.16，远优于其他三种有监督学习方法。而DPA为模型嵌入了更丰富的无标签数据的解剖信息，将DSC提高到了88.4%，而SemiFCN和ASDNet由于对不准确信息的学习，使得细小敏感的肾动脉发生严重的分割失误，削弱了网络性能。他们仅取得了68.5%和61.7%的DSC，甚至比3D U-Net还要低。而CPS采用双网络的形式进行相互的正则化限制过拟合到错误的信息上，取得了85.0%的DSC，但是其性能任然低于我们的DPA-DenseBiasNet。总的来说，DPA学习、DenseBiasNet以及HRA损失阶梯性地逐步提升了对肾动脉分割的性能，合在一起共同解决了CTA图像上的肾动脉分割任务的五大挑战，有效提升了深度学习模型在该任务上的表现。

**定性对比分析** 如图3-7所示，DPA-DenseBiasNet在精细肾动脉分割任务上取得了成功，获得了在不同尺度和奇异结构上良好的分割质量。案例4中的蓝色框区域显示，DenseBiasNet能够精细地分割肾动脉细小的末端区域，而其他两个模型在没有密集连接的情况下对这些细小结构敏感，无法分割这些区域。案例2中的黄色框区域显示，DPA学习策略引入了更加丰富的解剖结构特征的表征能力，使模型在奇异的肾动脉区域也拥有高质量的分割结果。在DPA特征时，3D U-Net、ASDNet和DenseBiasNet均失去了黄色框中的扭曲的奇异结构。而DPA-DenseBiasNet由于来自丰富无标签数据的表征能力，取得了对该区域完整、连续的分割。由于对无标签数据优化目标的不准确，ASDNet在肾动脉这种细小敏感的结构上容易失误，在案例1中将脊柱和其他器官的血管都分割为了肾动脉，导致了严重的误分割。

### 3.4.3 消融实验和模型分析

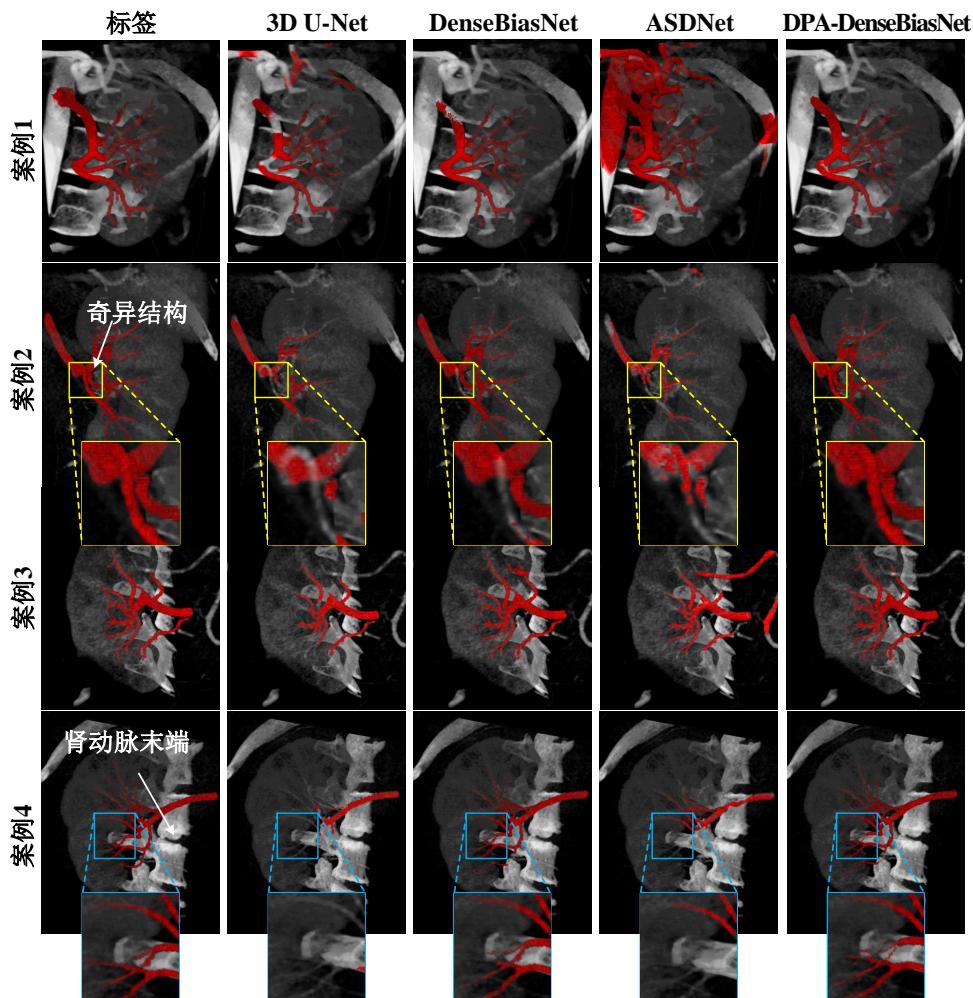


图 3-7: 对肾动脉精细分割结果的可视化。蓝色方框中的区域显示了DenseBiasNet细小的肾动脉末端的分割能力，而黄色方框区域显示了我们的DPA学习对奇异血管结构的性能提升。

**超参数分析** 如表3.2所示，本实验对密集偏置连接中的偏置量( $c$ )和HRA中的阈值( $T$ )进行了消融实验，以分析这些超参数的影响。我们尝试了两种偏置量的设置方法：网络中所有层采用相同数量的偏置量；偏置量是每层输出的特征图数量 $r$ 的比例。本实验可以得到两个有趣的结论。结果表明，当 $c=4$ 且 $T=0.1$ 时，DenseBiasNet(CE+HRA)能够达到最佳性能和相对较小的参数数量。而域置 $T$ 的增大不利于模型关注到难分割的区域，因此 $T$ 增大到0.2时，模型性能反而弱于0.1时的性能。当参数量越大，模型表现出越好的性能，但当过大时性能开始下降。 $c=4$ 和 $c=0.1r$ 时参数量均达到了1.6兆左右，且均取得了最高的87%以上的DSC，但进一步增加参数量到2.1兆时，模型性能下降到了83.5%DSC。这是因为过大的模型容易在标签数量不足的情况下发生过拟合问题，从而降低性能。

**密集偏置连接对训练过程的促进作用分析** 如图3-8和图3-10所示，密集偏置连接加快了收敛速度并且提高了准确性。本实验比较了DenseBiasNet、3D U-Net和没有密集偏

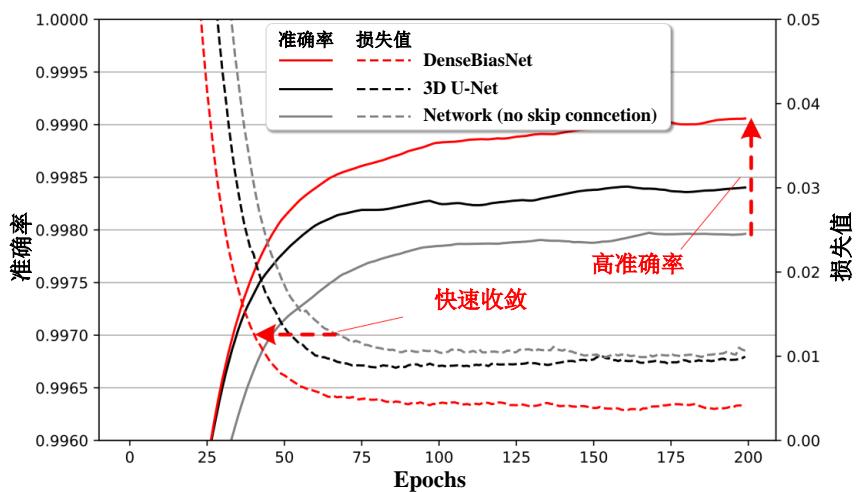


图 3-8: 密集偏置连接能够加速收敛过程并提高了准确性。DenseBiasNet通过全网络的密集连接构成一种隐式的深度监督，从而获得了比其他两个网络更快的收敛速度和更高的准确性。

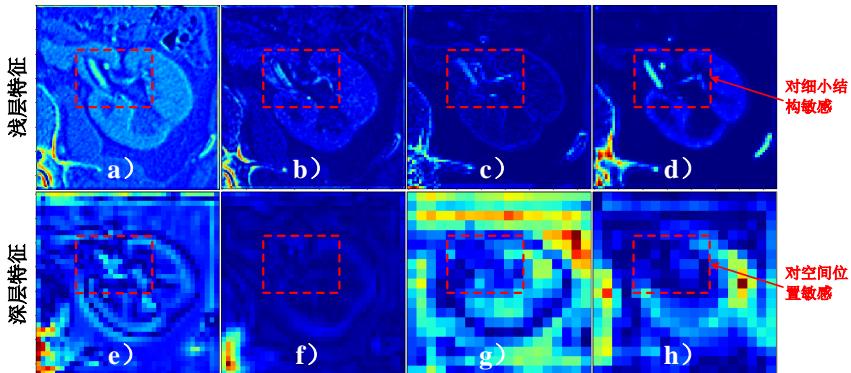


图 3-9: 不同感受野下的特征图可视化。a) -h) 显示了当超参数 $c = 1$ 时，DenseBiasNet的编码部分的特征图。a) -d) 是来自前两个分辨率阶段的浅层特征图；e) -h) 是来自第三和第四阶段的深层特征图。红色方框显示，密集偏向连接中的浅层特征对肾动脉的敏感性比深层特征更高。

置连接的基本编解码网络结构的准确率和损失值。由于DenseBiasNet全网络的密集连接，在训练过程中保持了完整的信息流和梯度流，因此与其他两种方法相比具有更高的准确性和更快的收敛速度。3D U-Net由于其部分跳跃连接而获得比完全没有跳跃连接的模型更快的收敛速度和准确性，但是由于其跳跃连接所带来的梯度流不充分，其收敛性与准确性仍旧弱于DenseBiasNet。根据ROC和PR曲线我们也能够发现，DenseBiasNet的准确性也比其他三个网络（3D U-Net、V-Net、Kid-Net）更高。

**密集偏置网络对细节特征的增强作用分析** 如图3-9所示，红色框中的结果表明，浅层小感受野下的特征对于肾动脉的敏感性更强，而深层大感受野下的特征拥有更好的空间定位能力，因此通过密集偏置连接从每一层向前传递并融合这些特征，反复增强在浅层特征中高响应的细小结构信息，并融合深层的大感受野的定位能力，将增强网络对于细小结构的敏感性与准确性，从而兼顾大区域和细小结构的分割质量。

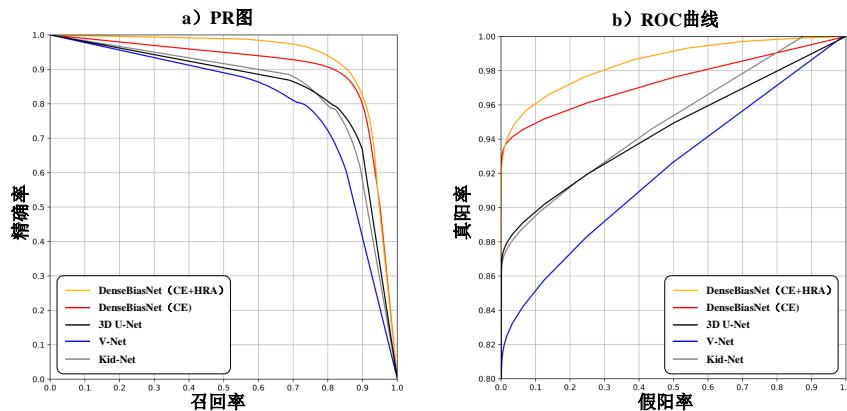


图 3-10: ROC 和 PR 曲线。DenseBiasNet 具有比其他网络更高的分割准确性，而 HRA 损失函数进一步提高了模型的性能。

表 3.3: 在更少的标签数据量下模型的性能对比。即使有标签的数据集规模减小，DPA 依然能够保持网络性能，表格显示了 DenseBiasNet 和 DPA-DenseBiasNet 在不同有标签图像数量下训练的 DSC。

数据量	$DenseBiasNet_{DSC\% \pm std} \uparrow$	$DPA-DenseBiasNet_{DSC\% \pm std} \uparrow$
100%	$87.2 \pm 9.5$	$88.4 \pm 7.6$
50%	$83.6 \pm 11.7$	$85.3 \pm 9.4$
20%	$75.7 \pm 14.6$	$78.5 \pm 12.9$
10%	$63.0 \pm 20.0$	$70.1 \pm 18.3$

**HRA对类别均衡的促进作用分析** HRA可以动态地均衡类别数量，从而降低类别不均衡问题对体积非常小的肾动脉的限制。如图3-11所示，本实验展示了直接使用交叉熵(CE)和加入了HRA后的交叉熵在前景血管与背景间的损失值比率。HRA损失的比率在前50轮迭代中快速下降，这意味着网络正在逐渐关注到难分割的区域。然后在后期的训练中，该比率始终保持在10以下使得类别接近均衡，而标准的交叉熵损失比率始终在100左右波动，说明在训练过程中类别严重不均衡，导致收敛速度缓慢并且容易丢失血管结构。如图 3-10 所示，HRA通过均衡类别数量，有效地提升了准确性。我们可以发现，HRA损失的ROC和PR曲线覆盖了标准的交叉熵，这意味着通过HRA训练的模型有更强的性能。

**DPA在更少的标签量上的有效性分析** 如表3.3所示，在有标签数据集较小时，本章所提出的DPA学习能够有效提高模型性能。本实验随机将训练集减少到原始大小的50%、20%和10%以评估DPA学习在更小的数据集上的有效性。我们发现，随着有标签图像数量的减少，DPA的增强能力逐渐体现出来，当数据量减少到10%时，DPA-DenseBiasNet相对于DenseBiasNet将DSC提高了7.1%。

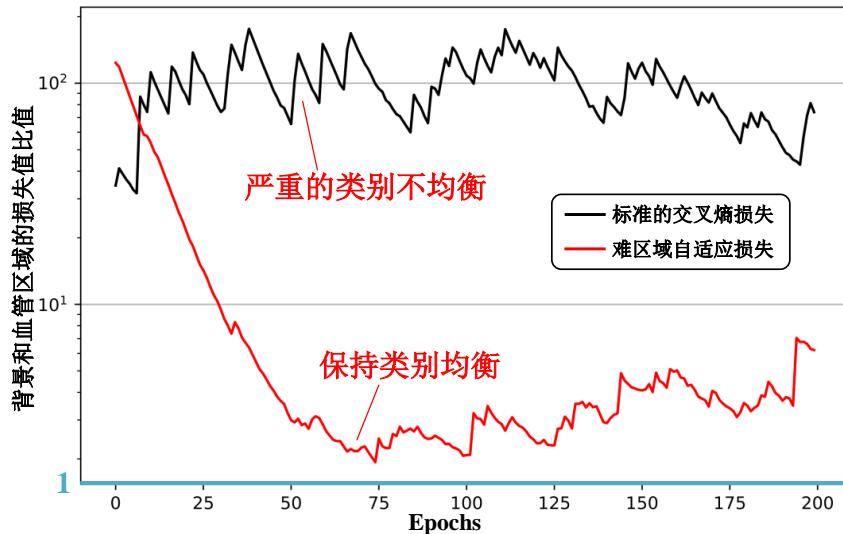


图 3-11：训练过程中背景和血管区域的损失值比折线图。HRA通过均衡不同类别的数量，防止了训练过程中类别不均衡导致的模型退化。在训练过程中，背景和血管类别之间的损失值比值表明，HRA损失可使比值保持在10以下，从而保持类别均衡，而原始的交叉熵损失函数的比值则在100，发生了严重的不均衡问题。

### 3.5 本章小结

本章的研究设计并实现了高效的半监督肾动脉精细分割，DPA-DenseBiasNet。它设计了一个对肾动脉这种尺度变化大的结构表征高效的网络结构（DenseBiasNet），一种半监督学习方法（DPA）和一种对细小结构特征学习更加有效的难特征挖掘的损失函数（HRA），从而在标签不足，分割目标难度大情况下实现了强大的肾动脉精细分割能力。实验结果表明，本章的方法能够有效地提升肾动脉学习效率，降低获得可靠的肾动脉精细分割模型的成本。本章的研究成果被发表在MedIA和MICCAI 2019上，详见作者简介中的文献[A1]和[A4]。

然而本章的研究局限于肾动脉精细分割这一针对单具体分割目标结构的医学图像分割任务上，尚未在多分割目标的任务上进行研究，通用性较差。与此同时，尽管这项研究一定程度上面向标签数量不足的挑战，但是仍旧需要较多的分割标签来驱动分割学习，整体学习过程中所花费的标签标注成本仍旧较高。在后面几章的研究中，我们将进一步从针对多结构的表征效率、标签学习效率和算法整体的通用性三个方面进行深入设计，以求可以面向更多分割目标和更加通用的场景。

## 第四章 基于元灰度自适应网络的肾脏多结构分割算法

上一章所设计的医学图像分割学习算法融合了深度先验解剖学习、密集偏置网络和难区域适应损失，在肾动脉精细分割这一单分割目标任务中取得了良好的性能，并且一定程度上降低了模型训练的标签数量需求，提高了分割学习的数据效率。然而，在面对多个分割目标结构时，如肾脏多结构分割任务，模型需要准确区分图像内多个结构之间的边界，获得对多个目标有区分度的表征。因此如何提高多分割目标的表征效率问题，尚未得到充分研究。此外，在上一章的研究中，为了解决图像间分布（解剖结构）变化大的问题，利用了大量无标签数据进行学习。然而，当数据有限且无法提供大量无标签数据时，图像间的分布变化问题仍然无法得到有效应对。

因此，本章在上一章研究的基础上，进一步从“肾脏多结构分割”任务出发，针对如何在特征不显著情况下学习适应图像间分布变化，以及如何提高模型对肾脏相关的多个结构的表征效率，这两个关键问题展开研究。根据任务的实际情况，本章致力于解决两个数据表征相关的挑战：图像内感兴趣区域分布狭窄导致的分割目标结构间分布区分度低，以及肾肿瘤亚型差异导致的图像间分布变化大。为此，本章的研究从数据分布( $D$ )和学习策略( $L$ )的角度出发，构建了一种自适应不同结构分布优势和图像分布优势的元学习算法，为肾脏多结构分割提供更加高效的算法，具体如下：对于分割目标间区分度低的挑战，本章基于医学图像中常用的窗口化技术，设计了一种新型的灰度分布搜索策略，自动化地调整输入图像的窗宽窗位以增大目标结构间灰度区分度，进一步使用多个窗宽窗位来适应不同结构的表征偏好，从而提高对多结构医学图像分割学习的表征效率。对于图像间分布变化大的挑战，本章算法结合元学习，设计了一种动态网络权重的策略，针对不同的图像分布动态地输出适应该部分网络参数，从而适应数据分布变化情况下的表征。

### 4.1 研究背景和动机

CTA图像上的三维一体化肾脏多结构分割能够同时服务于LPN手术的术前规划和术中导航，对手术的展开非常重要<sup>[156;186;187]</sup>。如图4-1所示，其目标是在一次前向运算推理中实现三维肾脏、肾肿瘤、肾动脉和肾静脉的分割掩膜，医生能够通过它观察分割获得了三维可视化模型，用于精确的术前规划<sup>[158;188]</sup>和术中辅助。在术前，医生能够根据肾动脉预估肾脏灌注模型<sup>[159]</sup>，从而轻松地查找到对肿瘤血供的肾动脉分支并提前确定血管夹闭位置<sup>[157]</sup>。而肾肿瘤和肾脏模型也将直观地向医生展现病变区域，从而提前规划肿瘤切除范围（图4-1b）。在手术中，医生能够把术前规划的标注与腹腔镜手术视频一起显示在屏幕上，以指导LPN手术的顺利进行<sup>[189;190;191]</sup>。而在肾门外的肾脏血

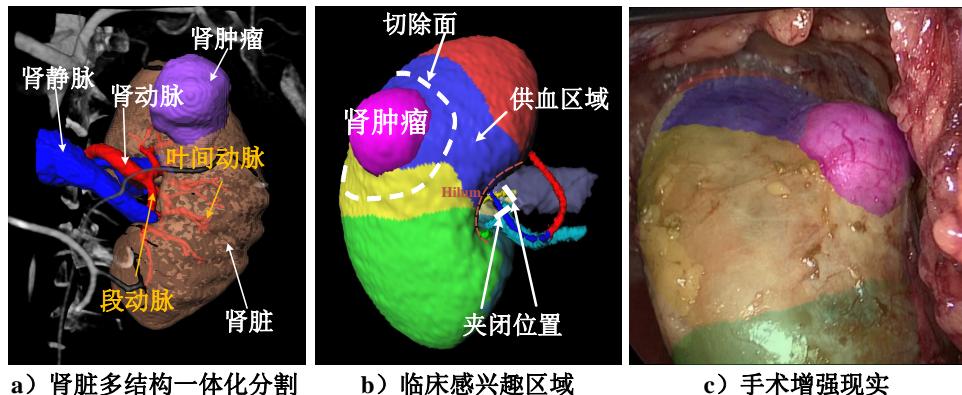


图 4-1: 三为一体化肾脏多结构分割能够为LPN手术提供精准的术前规划和术中指导。

管（肾静脉和肾动脉）也能够通过可视化清晰地显示动脉夹闭的区域，从而帮助医生快速找到夹闭位置，大大节约手术时间。同时，三维模型还是术中混合现实的模型基础，能够实时指导医生在术中做出适当的决策（图4-1c），提高手术品质。

尽管自动化的三为一体化肾脏多结构分割有着很高的临幊上价值，但在本章的研究成果<sup>[7]</sup>被发表前，还没有相关的研究被报道。最近的大部分研究集中在部分肾结构分割上<sup>[192;193;194]</sup>，缺乏分割完整的多个肾癌相关结构的对策。此外，一种直观的方法是对每个结构进行单独分割然后拼合，但是这又会带来重叠问题，特别是错综复杂在肾门区域，有着大量交错的肾静脉和动脉，容易发生不同类别结构重叠的问题。其他一些致力于多肾结构分割的研究<sup>[106;171]</sup>通常对于血管的分割品质较为粗糙，并且缺乏结构类别的完整性，限制了其临幊应用。

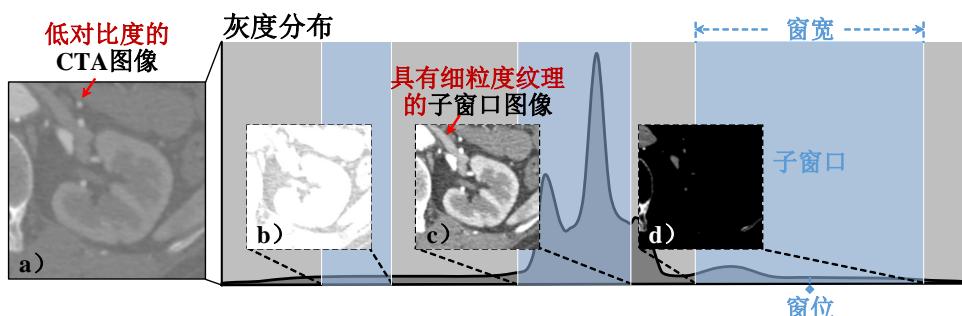


图 4-2: 挑战1——多结构间分布区分度低示意图。狭窄的任务相关区域的灰度分布使得原始CTA图像对比度较低，在灰度分布上使用窗口化操作能够使子窗口图像拥有细粒度的纹理。(a) 原始的CTA图像具有窄分布，使得图像对比度低。(b, d) 当窗口覆盖感兴趣区域无关灰度分布时，肾脏结构将被从图像中丢失。(c) 当窗口覆盖感兴趣的区域相关的灰度分布时，子窗口图像将具有清晰且细粒度的纹理。

究其原因，两个固有的挑战限制了本章所研究任务：**挑战1：**任务关注的多个对象在图像中灰度分布狭窄，对比度低难以分割。CT图像具有远大于自然图像的的灰度分布范围，但是肾脏相关结构仅在各自非常狭窄的范围内较为显著，这使得原始的CTA图像有着不清晰的软组织区域和区域间的边界。如图4-2a) 所示，与任务无关的分布占据

了非常大的灰度范围，将任务感兴趣的多个肾脏结构灰度分布挤压在狭窄的范围内，使得CTA图像从视觉上看模糊且具有较低的对比度。因此，深度神经网络将无法敏锐地区分灰度分布相近的相邻结构，并且鉴别结构之间的边界等细节特征的能力较弱。**挑战2：**图像间分布的变化使得模型难以适应可泛化的表征。肾肿瘤不同的亚型具有不同的体积和对造影剂的代谢能力<sup>[195]</sup>，因此，这使得采集到的CTA图像由于造影剂分布不同具有非常大的灰度分布变化。如图4-3所示，透明肾细胞癌呈现出高亮的外观和较大的体积，而乳头状肾细胞癌则表现出灰暗的外观和较小的体积。五种肿瘤亚型的灰度分布的中心和峰值各不相同，使得网络难以稳定地学习对底层信息（风格、纹理等）的有效表征，从而导致模型的偏好在不同图像间大幅波动，难以获得可泛化的表征能力。尤其是在本章任务中，那些对比度不明显且对变化敏感的静脉结构在这种多变的学习过程中将表现出较差的性能。

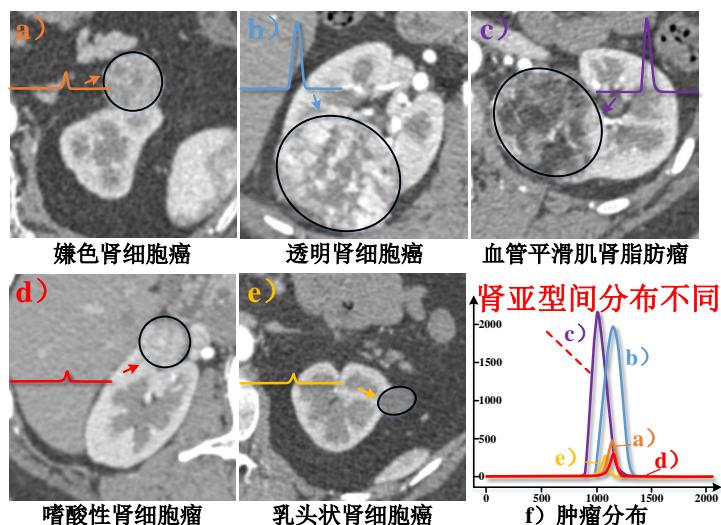


图 4-3: 挑战2——图像间分布差异大的示意图。由于不同肾肿瘤亚型拥有不同的体积和对造影剂的代谢能力，导致图像分布发生变化<sup>[195]</sup>。

**创新点1：**关注任务相关的灰度分布范围。窗口化技术（Widnowing）<sup>[80]</sup>通过扩展任务相关的分布，去除无关的灰度范围，增强感兴趣区域的对比度，使得网络能够聚焦于更广的肾脏结构分布范围以感知更细粒度的纹理。然而，一些常用的手工设置窗口化参数（窗宽，窗位）的方法对于本章的多结构分割任务而言存在较大的限制，这是因为该任务必须依赖于多个分布，并且图像之间具有较大的分布变化，如果仅设置单一的灰度分布将限制对多结构的有效学习。如图4-2所示，在示例图像的灰度分布中，当灰度分布的子窗口覆盖肾结构的分布范围（c）时，子窗口范围内的图像具有细粒度的纹理，肾相关结构区域对比度高。当子窗口超出肾结构的分布范围（b、d）时，肾结构将会变得不显著。因此，本章提出了一种名为灰度兴趣搜索（grayscale interest search, GIS）的算法，利用参数搜索的方式将分割网络聚焦于CTA图像中的任务相关灰度分布上。具体的，该算法通过控制和优化窗口化的两个相关参数来调整窗口宽度和中心来模拟医生在不同子窗口中观察不同结构，从而实现自动搜索任务相关分布。最终，利

用该方法为整个模型提供具有细粒度纹理的显著区域，从而促进网络学习细粒度的表示以进行精细分割。

**创新点2：**适应灰度分布的变化。元学习能够通过任务间的知识快速生成模型参数<sup>[196]</sup>，从而动态地适应任务间的分布变化，是一种任务级的动态适应策略，为本课题所关注的图像级的自适应问题提供了一种潜在的解决方案。在本研究中，我们提出了元灰度自适应（meta grayscale adaptation, MGA）学习（图4-4c），将每个图像作为一个子任务，动态生成适应图像间分布变化的参数，以适应图像的分布变换。该方法同时学习针对该任务预测的固定训练核，以及针对每个图像个性化特征动态输出的元卷积核，从而一起构建动态分割肾脏相关结构的元学习模型。它有三个关键元素：1) 鲁棒地表征图像的分布。从不同子窗口覆盖的多个分布中集成特征表示能力，从而提取丰富多样的特征以减少模型在多变分布下大对表征偏好，获得鲁棒的表征。2) 根据图像的分布动态生成个性化参数。将每个图像作为一个子学习任务，进一步构建一种元感知器来感知其分布特征，从而生成针对每个图像的个性化参数（元卷积核）以适应每个图像的分布变化。3) 个性化特征的融合。设计了一种元残差卷积（ResConv）将用于肾脏一体化分割的训练核和代表每个图像分布个性化变化的元卷积核通过残差融合，以适应图像间分布变化的个性化融合参数，从而提高泛化性。

本研究从灰度分布范围自动调整和灰度变化自适应的角度提出了一种灰度分布自动优化方法，在三维肾脏多结构一体化分割任务中取得了成功。本研究所提出的元灰度自适应网络（Meta Greyscale Adaptive Network, MGANet）可以在一次前向推理中同时分割CTA图像上的肾脏、肾肿瘤、肾动脉和肾静脉，具有良好的效果。本研究的具体贡献总结如下：

- 我们在三维肾脏多结构一体化分割上首次取得了成功，为精准LPN术前规划和术中导航提供了有效的模型基础，具有较高的临床价值。
- 我们提出了一种自适应图像灰度分布的一体化分割网格网络框架，MGANet，构建了一种精细的特征表示和个性化的特征融合，取得了最佳的肾脏多结构分割结果。
- 我们设计了一种新颖的超参数优化算法，GIS，该算法模拟了医生对CT图像上不同结构的观察过程，自动搜索与任务相关的灰度分布范围，从而为模型提供具有细粒度纹理和高对比度区域的输入图像，使网络将学习精细的表征能力。
- 我们设计了一种图像级元学习策略，MGA学习，该策略通过感知图像个性化的分布特征动态生成模型参数以帮助网络适应图像的灰度分布的变化，取得了鲁棒的分割性能。

## 4.2 相关工作

**肾脏结构分割** 肾脏结构分割对于LPN的精准术前规划和术中导航至关重要<sup>[158;188]</sup>，具有非常高的临床价值。本文第三章的研究工作<sup>[194]</sup>设计了一种半监督学习框架以实现肾动脉的精细分割，该技术对于选择肾动脉夹闭位置至关重要。Wang等人<sup>[192]</sup>采用了一种基于图割<sup>[197]</sup>的算法，构建了张量切割框架用于肾动脉分割。Taha等人<sup>[171]</sup>提出了一种基于卷积神经网络的肾血管分割框架，Kid-Net，用于分割肾动脉，肾静脉和输尿管。Li等人<sup>[106]</sup>提出了一个基于残差结构的U-Net<sup>[90]</sup>网络用于分割肾脏内部结构。然而，这些工作在我们的任务中存在两个方面的局限性：1) 其中的一些工作<sup>[194;171;90;192]</sup>仅关注部分肾结构的分割，缺乏一体化分割肾脏多结构的设计，如果使用这些方法单独分割每个肾结构，分割结果的重叠问题将难以解决。2) 其中的另一些工作<sup>[106;171]</sup>精度较低，缺乏被分割结构的细节，例如对于构建肾脏灌注模型<sup>[159]</sup>是至关重要的肾叶间动脉区域，该问题限制了这些方法在临床下游任务的实施。

**图像分割中的集成学习** 集成学习<sup>[198]</sup>通过构建一组学习器，并将它们集成起来以获得更准确的结果。在该方法的学习器训练过程与集成过程中嵌入了特定的设计策略能够解决多种挑战，如偏差纠正、类别不均衡等问题<sup>[199]</sup>。集成分割模型<sup>[91;200;201]</sup>能够将多个分割学习器的优势融合起来，提高整体的分割质量。Xia等人<sup>[91]</sup>设计了一种体积融合网络（VFN），在三维图像的x、y和z轴方向切片上分别训练一个二维分割网络，并使用一个三维网络来融合分割结果。Zheng等人<sup>[201]</sup>进一步在Xia等人的研究中加入了一个三维的子学习器，并将元学习器用于集成过程，取得了进一步的性能提升。虽然这些现有工作在其问题上表现出良好的结果，但这些方法难以应用于本研究任务中。这些方法忽略了任务中肿瘤类型变化所引起的灰度分布变化挑战，使得网络难以稳定地学习统一的浅层知识（风格、模式等），最终导致模型无法动态地适应分布变换，限制其泛化能力。本研究所设计的MGA学习根据各图像分布特征个性化地生成集成参数，使整体架构能够有效地适应不同图像的分布变化。

**超参数优化** 优化神经网络模型中的超参数<sup>[202;203]</sup>能够有效地提升网络的性能。许多先前的研究<sup>[86;204;205]</sup>表明，更深、更宽的网络将捕捉到更丰富和更复杂的特征，从而获得更准确的性能，但也会大大增加网络的非线性能力，提高过拟合的风险。Tan等人<sup>[203]</sup>同时研究了宽度、深度和分辨率的复合缩放对网络准确性的影响，并搜索出使用更少计算资源获得更高准确性的EfficientNet。在医学图像分析中，医学图像的灰度分布范围是另一个重要的超参数，却鲜有研究关注。CT图像的成像过程使得医学图像具有很大的灰度范围，导致原始图像呈现低对比度和模糊的软组织。当医生观察医学图像时，会使用窗口化方法<sup>[80]</sup>来移除无关的灰度范围并扩展感兴趣的分布，从而提高对比度并呈现更清晰的细节。然而，目前对于该超参数的优化研究仍旧缺乏。许多医学图像分析研究<sup>[194;90;106;171]</sup>忽略了对这一超参数的优化，仅在原始灰度范围上训练模型从而限制了

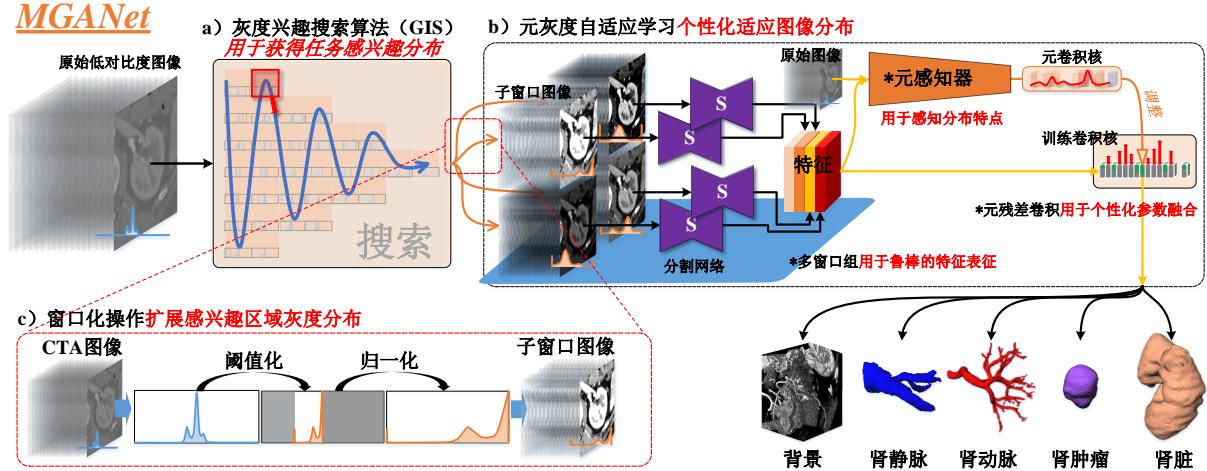


图 4-4: MGANet 通过 GIS 算法和 MGA 学习自适应地优化医学图像的灰度分布, 从而实现精细的肾脏多结构一体化分割

对感兴趣分布的分割质量。而其他一些研究<sup>[92;206]</sup>仅根据研究人员的经验或临床标准手动设置灰度范围, 缺乏任务导向的自动化设置能力, 通常获得的是次优解。本研究所设计的GIS策略, 自适应地搜索任务相关的灰度范围, 使网络能够学习到更细致的特征表示, 以更好地处理感兴趣的分布。

**用于参数生成的元学习** 用于参数生成的元学习能够动态地直接预测模型参数<sup>[196]</sup>避免了反向传播所需的反复迭代优化, 从而为模型生成个性化的参数并提供动态适应数据分布的能力。超网络<sup>[207;208]</sup>通过训练一个网络生成另一个网络的参数以实现快速学习, 在模型压缩和多任务学习中取得了成功。元网络<sup>[209]</sup>将元学习器预测的参数与针对任务训练的权重相结合, 实现了快速迁移泛化。WeightNet<sup>[95]</sup>统一了SENet<sup>[210]</sup>和Cond-Conv<sup>[211]</sup>这两个元学习网络的优势, 在训练网络预测每个卷积核参数, 能够动态地为分类任务生成权重。然而, 这些研究仅关注任务级的参数生成, 缺乏适应图像间灰度分布的变化的相关元学习研究。本研究的MGA学习策略将每个图像视为一个子学习任务, 为每个图像动态生成参数, 从而形成一个图像级元学习框架以适应分布变化。

### 4.3 方法

如图4-4所示, MGANet通过自适应优化医学图像的灰度分布以获得有效的三维一体化肾脏多结构分割。它通过灰度兴趣搜索算法(GIS, 第4.3.1节)和元灰度自适应学习(MGA, 第4.3.2节)策略, 自动化搜索并学习对任务和分割图像灰度分布有效的模型参数, 取得了精细的分割效果。

#### 4.3.1 灰度兴趣搜索 (GIS)

如图4-4a) 所示, GIS使用两个彼此相关的参数来缩放灰度窗口的宽度和中心, 从

而自适应地发现与我们的任务相关的灰度分布，并通过窗口化操作<sup>[80]</sup>来扩展该分布从而在图像上获得肾脏结构细粒度的特征。因此，该算法将使分割网络关注CTA图像上任务相关的灰度分布，从而学习精细的特征表示。

**任务公式化** 如图4-2所示，通过对灰度窗口的宽度 $w$ （窗宽）和中心位置 $c$ （窗位）进行缩放可以实现自动搜索感兴趣的灰度范围。窗宽 $w$ 决定了灰度窗口覆盖的灰度范围大小，窗宽越宽，覆盖的分布信息越丰富，但更宽的窗口会削弱不同结构的分布之间的区分度，从而降低图像的对比度和特征显著性。窗位 $c$ 决定了窗口的位置，窗位距离任务相关分布越近，图像中感兴趣的结构的细节越清晰。我们将训练数据集定义为 $\{X, Y\} = \mathbb{D}_{train}$ ，子窗口中的图像集定义为 $X^{w,c} = \mathcal{W}^{w,c}(X_{[g_{min}, g_{max}]})$ ，其中 $\mathcal{W}$ 是窗口化过程（见第4.3.1节）， $g_{min}$ 和 $g_{max}$ 是定义的原始灰度范围的最小值和最大值。在本研究任务中设置 $g_{min} = 0$ ,  $g_{max} = 2048$ 。在子窗口图像上训练的分割网络为 $\mathcal{N}_{\theta^{w,c}}$ ， $\theta^{w,c}$ 为对应子窗口上的模型训练参数，预测的分割结果集和为 $Y(w, c)$ ，因此这个过程定义为 $Y(w, c) = \mathcal{N}_{\theta^{w,c}}(\mathcal{W}^{w,c}(X_{[g_{min}, g_{max}]}))$ 。窗口边界不能超过原始灰度范围。我们将所有目标肾结构的DSC<sup>[184]</sup>作为指标，从而在搜索的子窗口图像上评估这些分割结果以评估搜索结果的好坏，因此任务相关灰度分布的搜索过程如下：

$$\begin{aligned} & \max_{w,c} DSC(Y(w, c)) \\ & s.t. \quad Y(w, c) = \mathcal{N}_{\theta^{w,c}}(\mathcal{W}^{w,c}(X_{[g_{min}, g_{max}]}) \\ & \quad c - \frac{w}{2} \geq g_{min} \\ & \quad c + \frac{w}{2} \leq g_{max}. \end{aligned} \tag{4.1}$$

本研究采用网格搜索方法<sup>[202]</sup>来优化这两个超参数，并使用两个彼此相关的系数 $\alpha$ 和 $\beta$ 将初始窗口宽度 $w_0 = \phi$ 和中心 $c_0 = \phi$ 按整数倍进行缩放，以减少搜索空间：

$$\begin{aligned} & \text{width: } w_\alpha = \alpha \phi \\ & \text{center: } c_\beta = \beta \phi \\ & s.t. \quad \alpha \in \mathbb{Z}^+ \\ & \quad \beta \in \mathbb{Z}^+ \\ & \quad \alpha + 2\beta \leq \frac{2g_{max}}{\phi} \\ & \quad \alpha + 2\beta \geq \frac{2g_{min}}{\phi}, \end{aligned} \tag{4.2}$$

其中， $\alpha$ 和 $\beta$ 是缩放系数，且必须满足宽度和中心的条件 $\frac{2g_{min}}{\phi} \leq \alpha + 2\beta \leq \frac{2g_{max}}{\phi}$ 。在实验中，我们设置 $\phi = 256$ ，使得灰度窗口网格（图4-5）有在32个子窗口上进行搜索。最终我们发现最佳值为 $\alpha = 3$ 和 $\beta = 5$ 。

**自适应搜索任务感兴趣的灰度范围** 如图4-5所示，GIS通过构造一个窗口网格，然后在该网格上进行窗口化操作来逐步搜索任务相关的灰度分布范围。

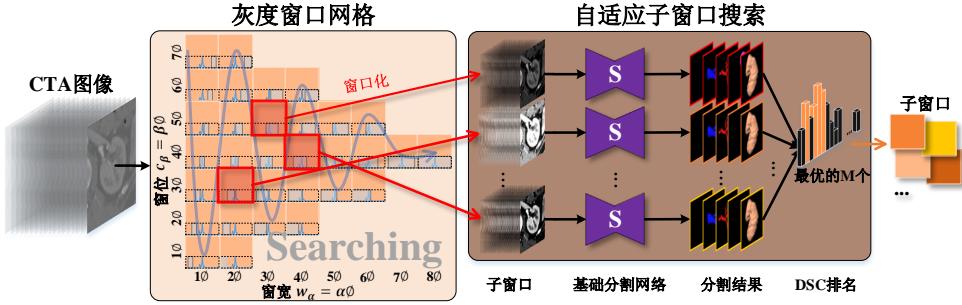


图 4-5: GIS 算法示意图。该算法搜索任务感兴趣的灰度分布，从而使得网络能够获得对细节精细的表征能力。

**窗口化操作：**如图4-4c) 所示，窗口化操作过程 $\mathcal{W}$ 会移除不相关的灰度范围并扩展与感兴趣结构相关的分布范围，从而提高对比度并提取显著的感兴趣区域，使网络能够关注肾脏结构相关的更显著的分布以获得其精细纹理的有效表征。该过程首先从灰度窗口网格中获取窗宽 $w_\alpha$ 和窗位 $c_\beta$ ，然后将原始图像 $x$ 在子窗口的边界 $c_\beta - \frac{w_\alpha}{2}$ 和 $c_\beta + \frac{w_\alpha}{2}$ 处进行阈值化，得到阈值图像 $x_{thr}$ 。最后，将阈值图像归一化为子窗口图像 $x^{w,c}$ ，使肾脏结构更加显著。因此，整个窗口化过程如下：

$$x^{w,c} = \mathcal{W}^{w,c}(x) = \frac{\max(\min(x, c_\beta - \frac{w_\alpha}{2}), c_\beta + \frac{w_\alpha}{2}) - (c_\beta - \frac{w_\alpha}{2})}{w_\alpha}. \quad (4.3)$$

**自适应子窗口搜索：**自适应子窗口搜索通过在灰度窗口网格上迭代地训练多个分割网络，并使用DSC指标<sup>[184]</sup>评估最佳的窗宽 $w$ 和窗位 $c$ ，以获得覆盖任务相关分布的子窗口参数。如图4-5所示，灰度窗口网格上的多个子窗口图像迭代地训练其对应的分割网络，以获得在不同子窗口上的分割结果。因此，这些窗口上的分割结果可用于计算本研究任务中肾脏结构的平均DSC。进一步，对这些平均DSC进行排名，获得对分割对象有优势的子窗口参数。然后在后续的元灰度自适应学习中，本框架选择前 $M$ 个窗口作为子窗口，进行集成与融合。根据模型性能和计算效率的实验评估， $M = 4$ 时取得最佳性能。在GIS算法采用DenseBiasNet<sup>[6]</sup>作为分割网络（在第4.3.3节中介绍）。

**讨论：GIS的属性** 1) GIS算法能够自适应地发现分割任务感兴趣的灰度分布，从而自动为深度学习模型提供更适当的灰度分布范围超参数。2) GIS算法去除了任务无关的灰度范围，并扩展了任务相关的分布，从而减弱了与任务无关分布变化的干扰，增强了感兴趣结构区域的对比度，使分割网络能够更有效关注肾脏结构分布以感知图像中细粒度纹理。

### 4.3.2 元灰度自适应 (MGA) 学习

如图4-4b) 所示，MGA学习通过GIS中获得的多个子窗口从多个分布范围提取多样化的特征，然后通过一个元感知器感知每个图像的分布特点形成一种针对每个图像的

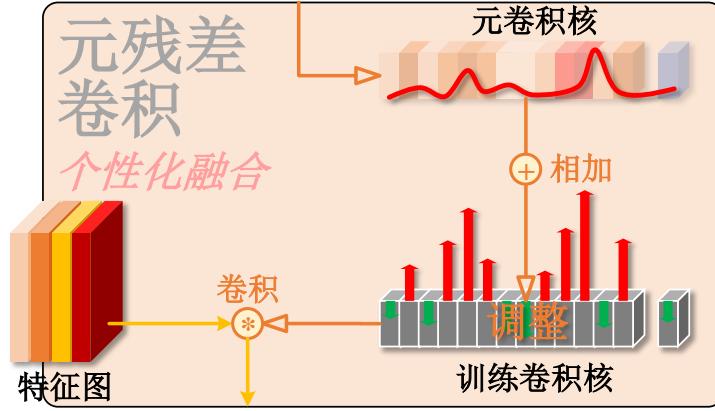


图 4-6: 元残差卷积示意图。同时使用训练卷积核和元卷积核用于个性化地适应每张图的分布特点。

元知识，最后利用一种元残差卷积将知识融入分割过程，从而适应每张图像的分布变化。

**多窗口组用于多样化的特征提取** 多窗口组从多个不同子窗口所覆盖的多个分布中提取特征，从而使特征更加多样化，减少整体模型对单个分布的偏好。它从GIS算法获得的窗口排序中获取前 $M$ 个子窗口图像 $\{x^0, \dots, x^M\}$ 以及在这些子窗口上独立训练的分割网络 $\{\mathcal{N}_{\theta^0}, \dots, \mathcal{N}_{\theta^M}\}$ ，从而获得覆盖这些子窗口的分布的表示。在前向运算过程中，将相应的子窗口图像输入这些网络中，获得最后一层输出卷积层之前的特征图，然后将这些特征被拼接为特征 $\mathcal{F}$ 以混合来自多个分布的多样化的特征：

$$\mathcal{F} = \text{cat}(\mathcal{N}_{\theta^0}(x^0), \dots, \mathcal{N}_{\theta^M}(x^M)), \quad (4.4)$$

其中，'cat'表示拼接过程。通过该过程，来自不同分布的特征将被组合在一起，形成具有更丰富的表征能力的特征组，从而更有效地应对分布变化。

**元感知器用于预测分布偏好** 元感知器 $\mathcal{P}$ 通过感知每个图像的分布特点，动态地融合来自 $M$ 个灰度分布的分割网络的特征。该方法将来自多窗口组的特征组 $\mathcal{F}$ 与原始图像 $x$ 一起输入到元感知器 $\mathcal{N}_{\xi}^P$ 中，生成一个元卷积核 $\hat{\mathbf{W}}_{meta}, \hat{b}_{meta}$ 。该元卷积核将被用于微调由分割网络输出层的权重，从而根据图像的分布动态地为每个图像提供个性化的调整：

$$\hat{\mathbf{W}}_{meta}, \hat{b}_{meta} = \mathcal{N}_{\xi}^P(\mathcal{F}, x). \quad (4.5)$$

**元残差卷积用于个性化特征融合** 元残差卷积将两个卷积核融合形成了一种个性化的动态参数生成过程，包括固定的训练卷积核  $\mathbf{W}_{conv}, b_{conv}$  和来自元感知器的元卷积核  $\hat{\mathbf{W}}_{meta}, \hat{b}_{meta}$ 。训练卷积核表征了用于本研究分割任务的知识，而来自元感知器的元卷积核则表征每张图像的个性化知识。因此，元卷积核通过残差连接<sup>[86]</sup>动态地微调训练卷积核，针对性地为每个图像生成更加合适的特征融合参数，从而提高框架在不同分布的图像

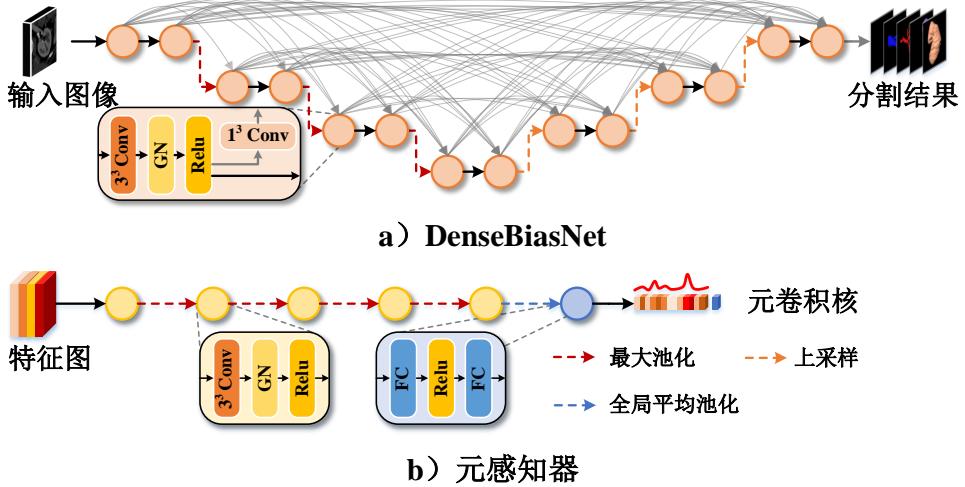


图 4-7: DenseBiasNet 和元感知器的网络结构细节。

上的泛化能力。如图4-6所示，来自GIS的已训练的分割网络的输出卷积层的权重被用来初始化元残差卷积的训练卷积核，而来自元感知器的元卷积核 $\hat{\mathbf{W}}_{meta}, \hat{b}_{meta}$ 被加到训练卷积核 $\mathbf{W}_{conv}, b_{conv}$ 中，形成一个动态的元卷积核 $\hat{\mathbf{W}}, \hat{b}$ ，从而动态调整模型对特征 $\mathcal{F}$ 中不同分量的注意程度。因此，模型将根据每张图像的分布特征对不同特征进行动态融合。该过程的公式为：

$$\begin{aligned}\hat{\mathbf{W}} &= \hat{\mathbf{W}}_{meta} + \mathbf{W}_{conv} \\ \hat{b} &= \hat{b}_{meta} + b_{conv} \\ \hat{y} &= \sigma(\hat{\mathbf{W}}^T \times \mathcal{F}_n + \hat{b}).\end{aligned}\quad (4.6)$$

其中， $\sigma$ 是Softmax激活函数， $n \in N$ 是三维特征图中特征向量的位置， $\times$ 是矩阵乘法， $\hat{\mathbf{W}}^T$ 是矩阵 $\hat{\mathbf{W}}$ 的转置。当元卷积核趋近于零 $\hat{\mathbf{W}}_{meta} \rightarrow 0, \hat{b}_{meta} \rightarrow 0$ 时，元残差卷积将退化为只是用训练卷积核的直接输出，MGA学习将退化为平均集成策略<sup>[199]</sup>。因此，MGA学习本身是一种对平均集成策略的动态增强，进一步考虑了每个图像的分布变化的个性化优化，因此它将带来不同分布间的泛化能力和稳定的性能。

**讨论：元灰度自适应学习的属性** 1) 多窗口组融合了来自多个分布的特征表示能力，提取了更多样化的特征，并减少了整体框架对单一分布的表征偏好。2) 元感知器感知每个图像的分布特点，并利用元残差卷积生成个性化的模型参数以融合不同子窗口的特征，从而适应图像的分布变化。3) 元残差卷积模块使用残差生成基于训练的卷积核的新融合参数，从而大大减小了元卷积核的输出空间，使生成的融合参数不会产生大的波动。

### 4.3.3 网络细节

**DenseBiasNet** 如图4-7a) 所示, 本研究采用DenseBiasNet<sup>[6]</sup>作为基础分割网络, 它融合了多重感受野和多分辨率特征以适应尺度变化, 并已经在肾动脉分割任务中取得了成功。它拥有七个分辨率阶段, 每个阶段有两个[ $3 \times 3 \times 3$ 的卷积层-组归一化层<sup>[179]</sup>-ReLU激活层]组 (Conv-GN-ReLU组)。最大池化和上采样层用于改变分辨率。最终阶段后, 它使用一个 $1 \times 1 \times 1$ 的卷积层和一个Softmax激活层来输出分割结果。网络内部, 每个Conv-GN-ReLU层的特征图都通过 $1 \times 1 \times 1$ 的卷积层压缩并传输到每个向前层, 从而形成了一个全网络的密集连接模式。在实验中, 我们使用交叉熵 (ce) 损失函数 $\mathcal{L}^{ce}$ <sup>[171]</sup>训练DenseBiasNet。

**元感知器** 如图4-7b) 所示, 元感知器采用连续的卷积、池化、激活和全连接层, 用于生成元卷积核, 这些核将对特征映射中的判别组合给出更高的响应。它具有五个分辨率阶段, 每个阶段都有一个Conv-GN-ReLU组。每个阶段最后都使用一个最大池化层以下采样特征图。最后, 通过一个全局平均池化层<sup>[212]</sup>将特征图压缩为一维特征向量, 并将该特征向量输入到两个全连接层中, 以进一步编码预测元卷积核。这些元卷积核通过参加连接添加到训练卷积核上输出预测的分割结果, 然后使用 $\mathcal{L}_{ce}$ 进行优化训练。

## 4.4 实验设置与结果分析

### 4.4.1 实验方案

**数据集** 本研究从江苏省人民医院放射科中回顾性地收集了123名仅有一侧肾脏肿瘤的患者的腹部CTA图像, 每张图像都采集于西门子双源64层CT扫描仪, 并在采集过程中注射造影剂。该数据集包含五种不同的肾脏肿瘤类型, 包括透明肾细胞癌、乳头状肾细胞癌、嫌色肾细胞癌、血管平滑肌脂肪瘤和嗜酸性腺瘤, 因此具有较大的异质性和分布差异。这些图像的体素大小介于0.47mm/体素和0.74mm/体素之间, 切片层厚为0.5mm/体素。预处理过程使用多模板匹配法<sup>[180]</sup>自动地剪裁出大小为 $150 \times 150 \times 200$ 的癌侧肾脏感兴趣区域。标注过程由两名专家参与, 对肾脏、肾肿瘤、肾静脉和肾动脉进行细致地标注, 并进行交叉验证。为了提高模型的泛化能力, 实验中使用 $[-20^\circ, 20^\circ]$ 随机旋转、 $128 \times 128 \times 128$ 大小随机裁剪、x、y和z轴方向随机镜像、 $[0.75, 1.25]$ 倍随机缩放和随机弹性变形的组合进行数据增强, 从而将数据集增加1,000倍。

**对比设置** 为了验证框架的优越性, 本实验将MGANet与V-Net<sup>[65]</sup>、3D U-Net<sup>[90]</sup>、Res-U-Net<sup>[106]</sup>、Kid-Net<sup>[171]</sup>、DenseBiasNet<sup>[6]</sup>和一个集成模型VFN<sup>[91]</sup>在进行比较。为了进一步研究GIS算法的特性, 实验进一步分析了分割网络在不同子窗口中的性能, 以及随着集成学习器数量的变化集成性能的变化情况。

表 4.1: 定量对比分析表。MGANet在三维肾脏多结构一体化分割任务上取得了比对比方法更好的性能。“EN” 表示模型是整体方法。

方法	肾脏			肿瘤		
	DSC <sub>%<math>\pm</math>std</sub> $\uparrow$	HD <sub>mm<math>\pm</math>std</sub> $\downarrow$	ASD <sub>mm<math>\pm</math>std</sub> $\downarrow$	DSC <sub>%<math>\pm</math>std</sub> $\uparrow$	HD <sub>mm<math>\pm</math>std</sub> $\downarrow$	ASD <sub>mm<math>\pm</math>std</sub> $\downarrow$
V-Net <sup>[65]</sup>	94.3 $\pm$ 1.6	14.77 $\pm$ 8.59	0.69 $\pm$ 0.26	81.5 $\pm$ 19.7	23.81 $\pm$ 25.63	3.40 $\pm$ 4.35
3D U-Net <sup>[90]</sup>	94.9 $\pm$ 1.6	19.40 $\pm$ 11.09	0.79 $\pm$ 0.50	82.0 $\pm$ 19.4	31.22 $\pm$ 27.98	2.78 $\pm$ 3.27
Res-U-Net <sup>[106]</sup>	94.4 $\pm$ 1.2	14.52 $\pm$ 9.52	0.62 $\pm$ 0.17	80.3 $\pm$ 14.0	53.87 $\pm$ 30.38	5.69 $\pm$ 5.71
Kid-Net <sup>[171]</sup>	94.3 $\pm$ 1.5	14.33 $\pm$ 8.70	0.66 $\pm$ 0.20	82.7 $\pm$ 19.1	22.04 $\pm$ 25.31	2.42 $\pm$ 3.32
VFN (EN) <sup>[91]</sup>	94.5 $\pm$ 1.5	15.70 $\pm$ 9.17	0.65 $\pm$ 0.36	86.3 $\pm$ 10.3	<b>16.34</b> $\pm$ 23.77	2.45 $\pm$ 3.82
DenseBiasNet <sup>[6]</sup>	94.2 $\pm$ 1.7	20.44 $\pm$ 12.70	0.76 $\pm$ 0.50	82.6 $\pm$ 15.1	35.60 $\pm$ 32.25	4.23 $\pm$ 5.32
<b>MGANet (EN)</b>	<b>95.1</b> $\pm$ 1.5	<b>12.28</b> $\pm$ 7.92	<b>0.53</b> $\pm$ 0.14	<b>86.4</b> $\pm$ 11.8	29.85 $\pm$ 31.17	2.76 $\pm$ 3.12

方法	肾静脉			肾动脉			平均
	DSC <sub>%<math>\pm</math>std</sub> $\uparrow$	HD <sub>mm<math>\pm</math>std</sub> $\downarrow$	ASD <sub>mm<math>\pm</math>std</sub> $\downarrow$	DSC <sub>%<math>\pm</math>std</sub> $\uparrow$	HD <sub>mm<math>\pm</math>std</sub> $\downarrow$	ASD <sub>mm<math>\pm</math>std</sub> $\downarrow$	
V-Net <sup>[65]</sup>	76.4 $\pm$ 9.2	21.99 $\pm$ 18.66	1.20 $\pm$ 0.71	84.3 $\pm$ 4.5	53.51 $\pm$ 14.86	0.76 $\pm$ 0.77	84.1 $\pm$ 6.5
3D U-Net <sup>[90]</sup>	73.7 $\pm$ 13.5	30.82 $\pm$ 21.06	2.12 $\pm$ 2.69	80.2 $\pm$ 8.2	50.82 $\pm$ 14.49	1.47 $\pm$ 1.83	82.7 $\pm$ 7.7
Res-U-Net <sup>[106]</sup>	76.8 $\pm$ 7.9	19.04 $\pm$ 10.84	1.27 $\pm$ 0.93	84.5 $\pm$ 8.6	27.60 $\pm$ 14.36	1.20 $\pm$ 1.89	84.0 $\pm$ 6.6
Kid-Net <sup>[171]</sup>	75.4 $\pm$ 10.4	18.25 $\pm$ 10.61	1.19 $\pm$ 0.96	78.0 $\pm$ 7.1	35.15 $\pm$ 16.76	0.91 $\pm$ 1.02	82.6 $\pm$ 7.3
VFN (EN) <sup>[91]</sup>	76.9 $\pm$ 9.0	<b>18.14</b> $\pm$ 12.45	1.20 $\pm$ 1.40	81.8 $\pm$ 6.1	23.33 $\pm$ 15.55	0.84 $\pm$ 1.16	84.9 $\pm$ 6.5
DenseBiasNet <sup>[6]</sup>	75.3 $\pm$ 11.3	23.47 $\pm$ 15.67	1.61 $\pm$ 2.53	86.1 $\pm$ 8.0	25.68 $\pm$ 14.77	1.27 $\pm$ 2.43	84.6 $\pm$ 6.8
<b>MGANet (EN)</b>	<b>81.0</b> $\pm$ 7.2	18.94 $\pm$ 13.86	<b>0.90</b> $\pm$ 0.70	<b>89.0</b> $\pm$ 5.1	<b>22.67</b> $\pm$ 13.33	<b>0.60</b> $\pm$ 0.70	<b>87.9</b> $\pm$ 5.1

**实施细节和评价指标** MGANet有两个训练阶段，在第一阶段中，初始窗宽 $w^0$ 和中心 $c^0$ 为256和256，逐步在灰度窗口网格上搜索了32个子窗口，在每个子窗口上分别训练相应的分割网络（DenseBiasNet<sup>[6]</sup>）。训练过程中，优化器设置为Adam<sup>[183]</sup>，批大小为1，学习率为 $1 \times 10^{-4}$ ，迭代次数为 $1.6 \times 10^5$ 。在第二阶段中，固定搜索到的多个子窗口上的DenseBiasNet 的权重以提取特征，然后将其输出卷积层的核作为训练核用于元残差卷积。与第一阶段相同，元感知器使用Adam<sup>[183]</sup>作为优化器，批大小为1，学习率为 $1 \times 10^{-4}$ ，迭代次数为 $1.6 \times 10^5$ 。该框架使用PyTorch<sup>[213]</sup>实现，并在NVIDIA TITAN Xp GPU上进行训练和测试。实验中所有模型都采用相同的训练设置，并使用五折交叉验证以进行全面评估。

本研究采用两个方面的指标来评估每种方法的分割性能<sup>[184]</sup>。1) 基于区域的指标：采用DSC来评估分割结果和的重叠指数。2) 基于距离的指标：采用对分割离群点不敏感的ASD来评估表面的重合程度。为了进一步比较异常值的分割质量，我们还采用了对离群点敏感的Hausdorff距离（HD）进一步评估分割结果的聚合性。所有肾脏结构的评价指标的标准差同时被计算与报告（ $\pm$ std）。

#### 4.4.2 对比分析

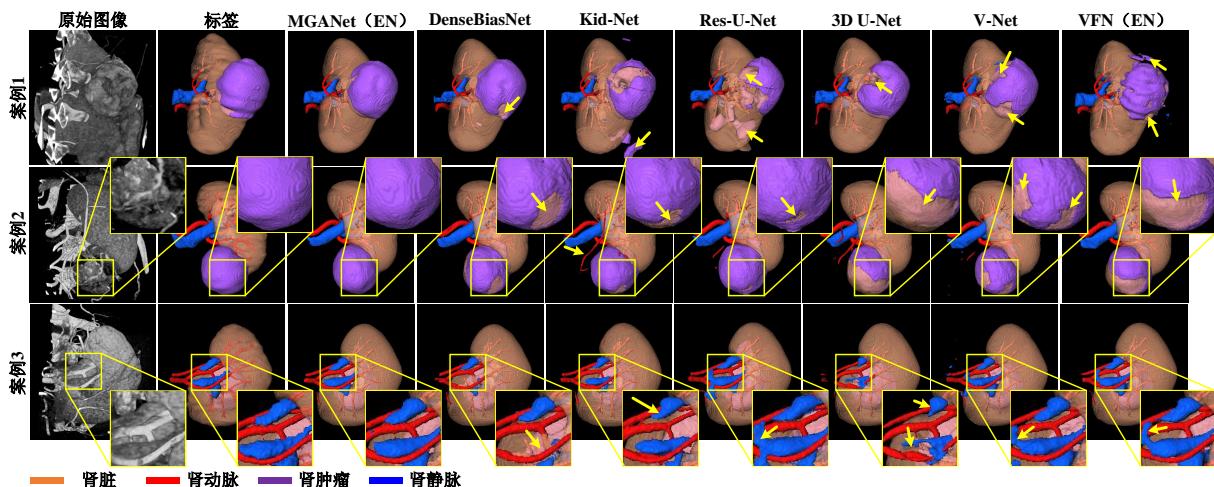


图 4-8: 定性对比分析表。MGANet拥有优质的可视化效果。

**定量对比分析** 如表4.1所示，MGANet相较于其他方法表现出了卓越的性能。MGANet取得了非常精准的肾动脉分割结果（DSC为89.0%，HD为22.67mm，ASD为0.60mm），这将为临幊上供血区域估计和动脉夹闭位置选择提供有效的帮助。MGANet同时取得了95.1%、12.28mm和0.53mm的肾脏DSC、HD和ASD；86.4%、29.85mm和2.76mm的肿瘤DSC、HD和ASD；81.0%、18.94mm和0.90mm静脉DSC、HD和ASD，通过有效的可视化，精细的三维分割模型能够为手术过程提供有效的视觉引导。集成模型VFN将不同视角的二维信息融合，在肿瘤分割上表现良好（86.3%的DSC），并在静脉分割上具有较少的离群点（18.14mm的HD）。然而，由于原始图像对比度较低且二维切片中的动脉特征不显著，因此其在动脉分割上表现不佳（仅为81.8%的DSC）。Kid-Net采用了深度监督，因此在肿瘤上取得了有竞争力的性能，获得了较小的ASD（2.42mm），但其在动脉和静脉上的DSC仅为78.0%和75.4%，对细小血管分割性能较差。直接使用基础的分割网络DenseBiasNet能够在四个结构上都获得良好的表现，并在原始窗口上取得了84.6%的平均DSC。但在进一步采用本研究的GIS和MGA之后，MGANet在每个肾结构上平均提高了3.3%的DSC。

**定性对比分析** 如图4-8所示，MGANet在视觉上展现出了显著的可视化优势，展现出肾脏手术视觉指导的潜力。在案例1中，MGANet在视觉上展现出良好的肾脏多结构分割效果，而 DenseBiasNet、Kid-Net和Res-U-Net在肾脏和肿瘤之间存在严重的错误分割（黄色箭头）。由于肿瘤灰度范围狭窄和灰度分布的变化，V-Net和VFN在肿瘤区域的分割结果较为粗糙。在案例2中，放大的细节显示了肿瘤边界处良好分割质量。由于原始窗口的低对比度和肿瘤分布的变化，所对比的方法都将肾脏外的肿瘤区域分割到肾脏内部（黄色箭头）。在案例3中，MGANet在放大的肾门区域分割品质良好，展现出本研究方法在复杂区域的出色分割质量。由于肾脏、静脉和动脉结构拥挤在狭窄的肾门区域，其他比较方法在这样一个复杂的环境中，静脉区域的低对比度存在严重的误分

表 4.2: 本章算法的消融实验分析表。其中“平均集成-4”采用平均集成策略来融合四个网络。

GIS	平均集成-4	MGA-4	肾脏	肿瘤	肾静脉	肾动脉	平均DSC
			$94.2 \pm 1.7$	$82.6 \pm 15.1$	$75.3 \pm 11.3$	$86.1 \pm 8.0$	$84.6 \pm 6.8$
✓			$94.5 \pm 2.0$	$83.6 \pm 14.1$	$79.2 \pm 6.7$	$88.0 \pm 5.3$	$86.3 \pm 5.7$
✓	✓		$95.0 \pm 1.4$	$85.5 \pm 13.0$	$79.5 \pm 8.2$	$88.1 \pm 7.0$	$87.0 \pm 5.6$
✓		✓	<b><math>95.1 \pm 1.5</math></b>	<b><math>86.4 \pm 11.8</math></b>	<b><math>81.0 \pm 7.2</math></b>	<b><math>89.0 \pm 5.1</math></b>	<b><math>87.9 \pm 5.1</math></b>

割。

#### 4.4.3 消融实验和模型分析

**消融实验** 如表4.2所示，本研究的创新点为整体框架带来了显著的提升。原始窗口下的Dense-BiasNet在肾脏、肿瘤、静脉和动脉上分别取得了94.2%、82.6%、75.3%和86.1%的DSC。当加入GIS算法时，模型在最佳子窗口( $w = 768, c = 1280$ )上增强了图像的对比度，从而使网络关注感兴趣的肾脏结构分布以感知细粒度纹理，最终在静脉和动脉分别取得了3.9%和1.9%的DSC提升。表4.2中的“平均集成-4”是一个采用平均集成策略<sup>[199]</sup>的集成模型，计算在最优的四个子窗口上训练的网络的分割结果的平均值。当采用平均集成策略时，框架在肾脏、肿瘤、静脉和动脉上分别达到95.0%、85.5%、79.5%和88.1%的DSC相比于单独的DenseBiasNet取得了有效提升。当进一步使用MGA学习使网络个性化融合不同分布的偏好时，MGANet取得了进一步的显著提升，尤其是在静脉上，相比平均集成策略取得了显著的1.5%DSC提升，并最终获得了87.9%的平均DSC。

**灰度兴趣搜索分析** 如图4-9所示，GIS算法获得了多个子窗口，这些子灰度窗口下的图像在肾脏结构中的表现优于直接使用原始的CTA图像。**1)**对于子窗口发生变换时，不同肾脏结构的敏感性不同。肾脏有相对较大的体积、有规则的形状和较宽的灰度范围，这使得特征的表征能力对于灰度窗口的变化相对稳定，因此在子窗口改变时，模型的表现相对稳定。而相比之下，由于肾静脉、肾肿瘤和肾动脉体积较小且灰度范围较窄，一些极端的子窗口难以覆盖这些结构，使得它们对窗宽和窗位的变换敏感。**2)**由于肾脏结构具有不同的灰度分布，因此他们具有不同的优势子窗口。肾静脉、肾脏、肾肿瘤和肾动脉的最佳子窗口分别为( $w = 1280, c = 1024$ )、( $w = 1536, c = 1024$ )、( $w = 1280, c = 1280$ )和( $w = 768, c = 1280$ )。考虑到整体性能的优越性，在实验中本研究计算四个目标分割结构的平均DSC并选择排名前四的子窗口(( $w = 768, c = 1280$ )、( $w = 1280, c = 1024$ )、( $w = 512, c = 1280$ )、( $w = 1280, c = 1280$ ))用于MGANet。**3)**在覆盖任务感兴趣分布的子窗口中训练的网络比在原始窗口中训练的网络表现更好，在原始窗口中训练的网络的平均DSC为84.6%，比在最佳子窗口中训练的网络低1.7%。

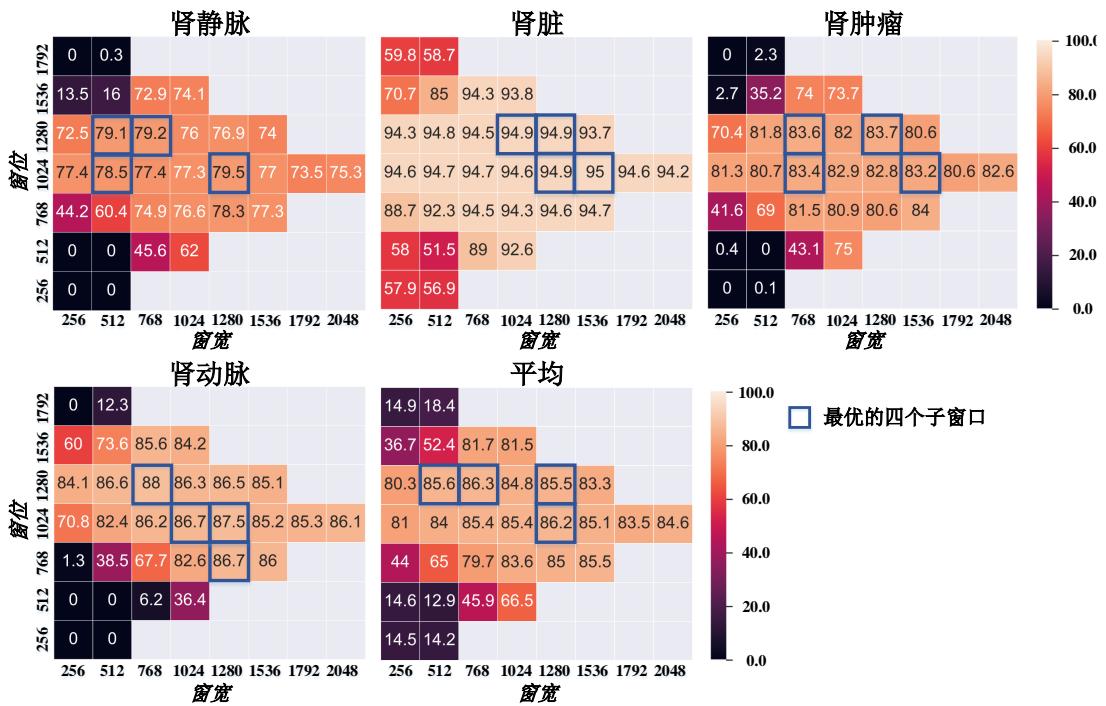


图 4-9: GIS 算法中不同窗口在各肾脏结构上的性能热力图。

**融合的网络数量分析** 如图4-10所示，随着有好到差融合更多的在GIS算法过程中训练的网络，整体框架的性能先增加后减小。在一开始，随着更多网络的融入，模型能够感知到更丰富的分布特征，从而性能逐渐增加。但是一些特定子窗口中某些肾脏结构分割准确性的严重降低形成了噪声网络，随着融合网络数量的增加，这些噪声网络融入整个集成框架，使得框架整体性能下降。肾脏结构对于融合网络数量的增加有不同的敏感度。由于肾脏体积较大，灰度范围较广，因此其DSC波动较小。而肿瘤结构较为敏感，当模型融合超过20个网络时，噪声网络将严重干扰整体准确性并造成性能的严重下降。因此，综合考虑模型性能和计算效率，本研究将融合子窗口的数量设置为4，使得整体框架拥有最高的平均DSC和相对较小的参数量。

**元卷积核分析** 如图4-11所示, MGA学习中的元卷积核调整训练卷积核, 从而动态适应灰度分布的变化。对于不同的肾脏结构, 感知器输出不同的元卷积核值, 从而以不同程度调整训练卷积核。由于肾结构具有不同的分布特点, 如肾脏灰度分布变化较小而肿瘤变化较大, 元感知器对于不同结构输出不同的微调程度(即元卷积核), 以针对性地增强最终结果。针对不同图像, 元感知器也将输出不同的元卷积核来个性化地调整训练卷积核以适应分布的变化。由于MGA学习在静脉方面有显着的提升, 本实验展示了来自两个不同图像的静脉的元卷积核。随着分布在不同的图像之间变化, 元感知器能够感知这种变化, 从而元卷积核也动态发生变化。

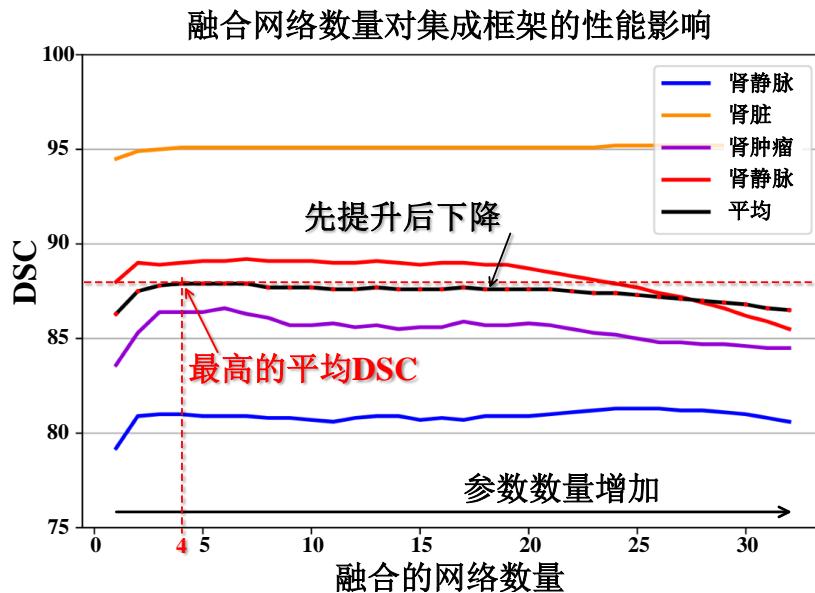


图 4-10: 融合的网络数量折线图。随着融合网络的数量增加，模型性能先增加后减少，并且不同肾脏结构将对网络数量的增加具有不同的敏感性。

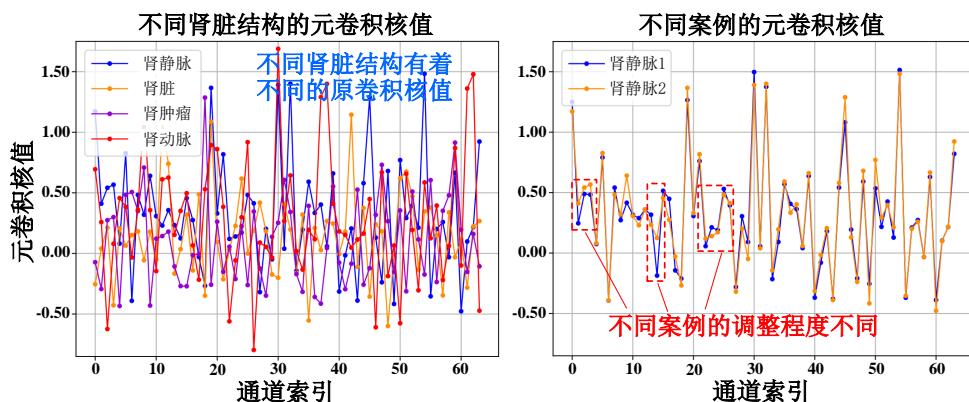


图 4-11: MGA 学习中的元卷积核数值分布图。元卷积核通过调整训练卷积核，动态适应灰度分布的变化。

## 4.5 本章小结

本章研究了基于元学习的肾脏多结构分割，并提出了一种适应图像灰度变化的元灰度自适应网络，MGANet。它设计了一种灰度兴趣搜索算法（GIS）自适应地搜索和扩展图像中任务相关分布以增强任务感兴趣的多个对象特征的鉴别力，并构造了一种图像级的元学习策略——元灰度适应学习，动态调整网络内部参数以适应每个图像的分布变化。实验结果表明，本文的方法能够有效提升肾脏多结构分割性能，在数据有限情况下即可学习获得适应性强的表征。本章的研究成果被发表在MedIA和IPMI 2021上，详见作者简介中的[A2]和[A6]。

尽管本章的研究进一步面向了多结构分割，与上一章相比拥有更强的通用性和面向更复杂的问题，然而本章的研究仍然局限于肾脏多结构分割这一具体医学图像分割

任务，在通用性方面仍旧具有一定的局限性。与此同时，本章的整体学习过程仍旧需要较大的分割标注成本，标签效率较低。在下一章的研究中，我们将始图进一步降低标签数量到个位数，大大减少标签标注成本，并构建通用的医学图像分割学习算法以降低模型的设计成本。



## 第五章 基于分割配准互补性的少标签通用多结构医学图像分割算法

前两章的研究聚焦于肾癌手术辅助这个临床相关的领域，并探索了如何提高对图像的表征效率以及降低对标签数量的依赖。在肾动脉精细分割和肾脏多结构分割这两个单目标和多目标结构的分割任务中，取得了显著的性能提升，展示了算法在提升效率和潜在临床应用方面的潜力。然而，前两章的方法仍然需要相对大量的分割标签，这将消耗较大的标注成本。在标签数量非常少的情况下，这些方法可能无法运行。此外，尽管前两章的研究从单目标结构分割扩展到多目标结构分割，但仍然局限于肾癌影像分割这一特定的临床任务，通用性有限。

因此，本章进一步从分割和配准的互补性角度出发，研究一种少标签通用多结构医学图像分割算法，针对如何进一步减少训练所需标签到个位数，以及如何构建更加通用的多目标分割学习，这两个关键问题展开研究。本章的研究从优化伪标签训练数据生成( $\mathbb{D}$ )和学习策略( $\mathcal{L}$ )的角度，构建了一种可靠的基于配准的伪标签数据生成和鲁棒的分割学习算法，具体如下：首先，本章介绍了一种叫做“学习配准以学习分割”的少标签医学图像分割范式，阐述了配准对分割的促进作用。随后，分析了该范式下伪标签数据生成的真实性、多样性问题，以及学习伪标签数据的鲁棒性问题。进一步，本章通过构建分割和配准的互补性，优化了配准学习过程、伪标签数据生成过程和分割学习过程，实现了一种更加鲁棒和高效的医学图像少标签通用分割学习算法。

### 5.1 研究背景和动机

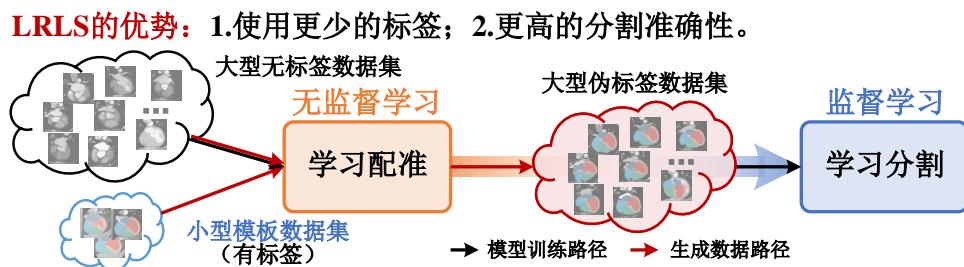


图 5-1: 学习配准以学习分割范式 (LRLS) 用于少标签医学图像分割。

学习配准以学习分割范式 (Learning registration to learn segmentation, LRLS, 图5-1) [214;127;126;125;215;137] 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

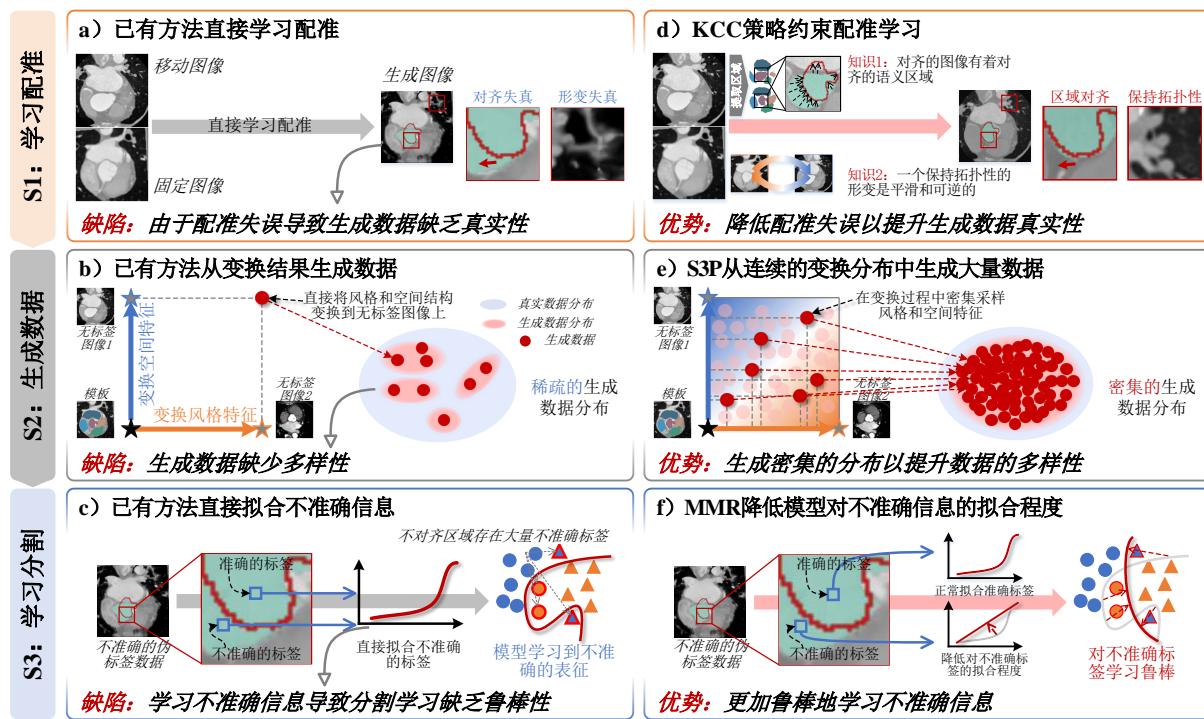


图 5-2: 现有的LRLS框架存在一定的局限性，而本章所提出的BRBS框架突破了已有方法的瓶颈，仅使用非常少标签即可取得强大的医学图像分割性能。a) 到c) 是现有LRLS框架（以DataAug<sup>[127]</sup>为例）的局限性；d) 到f) 是本章得到BRBS框架的设计动机和潜在优势。

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<sup>[127;137;126;125]</sup> 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5-1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5-2a-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<sup>[215;127;214]</sup>。在利用配准生成数据的过程中，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：<sup>[126]</sup>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：<sup>[216]</sup>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<sup>[215;214]</sup>，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<sup>[125;127]</sup>。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<sup>[214]</sup>或风格迁移<sup>[127]</sup>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<sup>[214;127;126;125;137]</sup>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5-2 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5-2d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<sup>[125;217]</sup>，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<sup>[218;219]</sup>，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<sup>[220]</sup>，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5-2e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5-2f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<sup>[221]</sup>，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支持学习基准，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<sup>[90;222;32;6]</sup>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<sup>[214;127;126;125;137]</sup>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<sup>[125]</sup>，BRBS具有更好的简便性。

## 5.2 相关工作

**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<sup>[113;134;135;136]</sup>。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<sup>[136]</sup>、视频分类<sup>[134]</sup>、神经架构搜索<sup>[135]</sup>等。由于医学数据的标注成本高昂且数据稀缺<sup>[62]</sup>，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<sup>[137;126;127;125;138]</sup>。然而，正如 Wang 等人<sup>[113]</sup>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<sup>[214;127;126;125;215;137]</sup>，但该问题仍然是少标签学习的核心瓶颈。

**医学图像配准和分割** 在过去的十年中，医学图像配准<sup>[220]</sup>（medical image registration, MIR）和医学图像分割<sup>[7]</sup>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<sup>[223;220;125;224]</sup>，而分割则密集地预测每个像素位置所属的类别<sup>[16;7;6]</sup>。随着深度学习<sup>[5]</sup>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<sup>[194;220;125;7]</sup>。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<sup>[220;224;125]</sup>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<sup>[7;6]</sup>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。

**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<sup>[215;225;226;227;228;229]</sup>。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<sup>[228]</sup>。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<sup>[215;225;226;227;214;127;126;125]</sup>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<sup>[215;226]</sup>是使用网络无监督地学习配准模型<sup>[220]</sup>，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<sup>[224]</sup>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。

**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<sup>[126]</sup>取得了一系列的成果<sup>[214;127;126;125;137]</sup>，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<sup>[127]</sup>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<sup>[137]</sup>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，

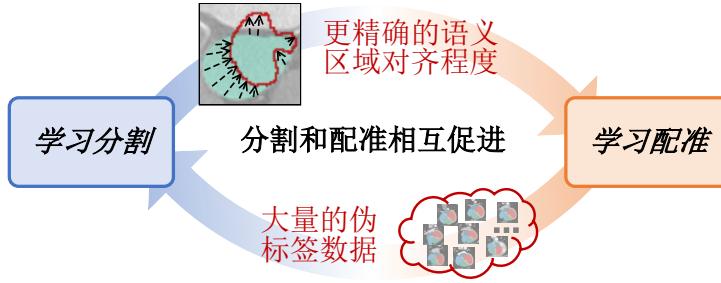


图 5-3: 分割 (S) 和配准 (R) 的互补学习过程。在整个训练过程中，配准和分割网络循环训练，互补优化，从而实现相互促进。

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<sup>[126]</sup>，PC-Reg-RT<sup>[125]</sup>和DeepAtlas<sup>[214]</sup>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<sup>[214;127;126;125;215;137]</sup>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。

### 5.3 方法

如图5-2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5.3.1节），在变换路径上密集采样数据，以获得更好的多样性（S3P，第5.3.2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5.3.3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<sup>[126]</sup>（图5-3），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。

#### 5.3.1 知识一致性约束策略（KCC）

如图5-2d) 和图5-4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：

**双向一致性（BiC）策略** BiC策略约束配准网络 $\mathcal{N}_\xi^R$ 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 $\psi$ 将图像M形变到图像F的空间结构<sup>[220]</sup>，还训练网络学习反向形变，即通过反向形变场 $\psi^{-1}$ 将图像F形变到图像 M 的空间结构：

$$\psi = \mathcal{N}_\xi^R(\mathbf{M}, \mathbf{F}), \psi^{-1} = \mathcal{N}_\xi^R(\mathbf{F}, \mathbf{M}). \quad (5.1)$$

正向和反向形变场还通过反向损失 $\mathcal{L}^I$ 进行相互约束:

$$\begin{aligned}\mathcal{L}_{\xi}^I(\psi, \psi^{-1}) &= \sum_{p \in \psi} \|\psi - I(\psi^{-1})\|^2, \\ I(\psi^{-1}) &= -1 * \psi^{-1}(\psi^{-1}),\end{aligned}\quad (5.2)$$

其中,  $p$ 是 $\psi$ 中的体素位置,  $\psi^{-1}(\psi^{-1})$ 表示使用空间变换操作<sup>[220]</sup>将 $\psi^{-1}$  (括号内) 按照位移向量场 $\psi^{-1}$  (括号外) 形变到目标空间位置,  $I$ 表示将位移向量场取逆变换。因此, 该损失有两个步骤: 1)  $I(\psi^{-1})$ 首先将逆变形场 $\psi^{-1}$ 与正向变形场 $\psi$ 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 $\mathcal{L}_{\xi}^I(\psi, \psi^{-1})$ 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 $\mathbf{M}^{\psi}$ 与固定图像 $\mathbf{F}$ 对之间的相似度 $\mathcal{L}_{\xi}^{sim}$ , 以及形变图像 $\mathbf{F}^{\psi^{-1}}$ 与固定图像 $\mathbf{M}$ 对之间的相似度。因此, 配准模型将学习到更好的可逆性, 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<sup>[220]</sup>作为 $\mathcal{L}_{\xi}^{sim}$ :

$$\mathcal{L}_{\xi}^{sim}(\mathbf{F}, \mathbf{M}^{\psi}) = \sum_{p \in \Omega} \frac{(\sum_{p_i} (\mathbf{F}(p_i) - \hat{\mathbf{F}}(p))(\mathbf{M}^{\psi}(p_i) - \hat{\mathbf{M}}^{\psi}(p)))^2}{(\sum_{p_i} (\mathbf{F}(p_i) - \hat{\mathbf{F}}(p))^2)(\sum_{p_i} (\mathbf{M}^{\psi}(p_i) - \hat{\mathbf{M}}^{\psi}(p))^2)}, \quad (5.3)$$

其中,  $\Omega$ 是图像空间,  $\hat{\mathbf{F}}(p)$  ( $\hat{\mathbf{M}}^{\psi}(p)$ ) 是局部均值强度图像:  $\hat{\mathbf{F}}(p) = \frac{1}{n^3} \sum_{p_i} \mathbf{F}(p_i)$ 。 $p_i$ , 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<sup>[220]</sup> $\mathcal{L}_{\xi}^{smooth}$ 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。

$$\mathcal{L}_{\xi}^{smooth}(\psi) = \sum_{p \in \psi} \|\nabla \psi(p)\|^2, \quad (5.4)$$

因此, BiC的整体损失 $\mathcal{L}^{BiC}$ 是:

$$\begin{aligned}\mathcal{L}_{\xi}^{BiC}(\mathbf{F}, \mathbf{M}, \psi, \psi^{-1}) &= \mathcal{L}_{\xi}^{sim}(\mathbf{F}, \psi(\mathbf{M})) + \mathcal{L}_{\xi}^{sim}(\mathbf{M}, \psi^{-1}(\mathbf{F})) \\ &\quad + \lambda^{smooth}(\mathcal{L}_{\xi}^{smooth}(\psi) + \mathcal{L}_{\xi}^{smooth}(\psi^{-1})) + \lambda^I \mathcal{L}_{\xi}^I(\psi, \psi^{-1}),\end{aligned}\quad (5.5)$$

其中,  $\lambda^{smooth}$ 和 $\lambda^I$ 是 $\mathcal{L}_{\xi}^{BiC}$ 中损失的权重。遵从论文<sup>[220]</sup>和论文<sup>[230]</sup>, 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<sup>[220]</sup>具有更好的拓扑保持能力, 且降低了论文<sup>[230]</sup>中由于过度平滑所导致的配准效果差的风险。

**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 ( $\mathcal{N}_{\theta}^S$ ) 对两幅图像 ( $\mathbf{M}$ 、 $\mathbf{F}$ ) 进行分割来提取语义区域, 并通过SeC损失 $\mathcal{L}_{\xi}^{SeC}$ 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<sup>[214]</sup>作为度量语义区域对齐程度的损失函数:

$$\mathcal{L}_{\xi}^{SeC}(\mathbf{F}, \mathbf{M}, \psi, \psi^{-1}) = \mathcal{L}_{\xi}^{DSC}(\mathcal{N}_{\theta}^S(\mathbf{F}), \psi(\mathcal{N}_{\theta}^S(\mathbf{M}))) + \mathcal{L}_{\xi}^{DSC}(\mathcal{N}_{\theta}^S(\mathbf{M}), \psi^{-1}(\mathcal{N}_{\theta}^S(\mathbf{F}))). \quad (5.6)$$

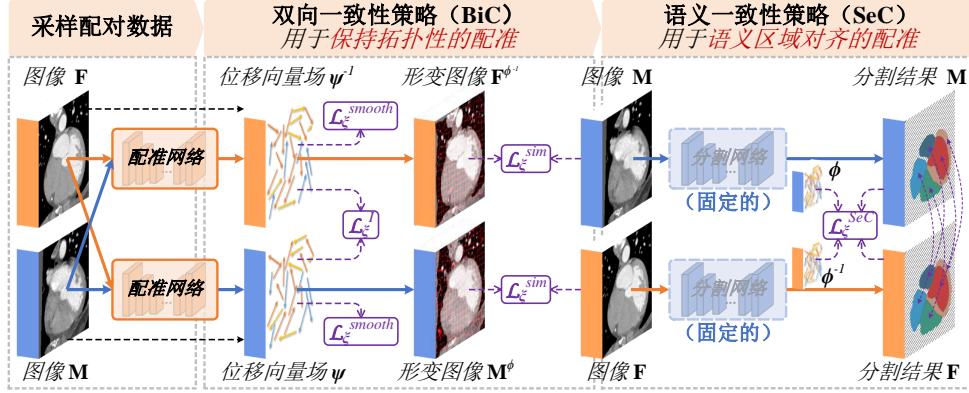


图 5-4: KCC 通过 BiC 和 SeC 策略来训练拥有保持拓扑性和语义区域对齐能力的配准模型, 从而生成符合领域知识伪标签数据生成过程, 提高生成数据的真实性。

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  $\mathcal{L}_\xi^R$ :

$$\hat{\xi} = \arg \min_{\xi} [\mathbb{E}_{(\mathbf{F}, \mathbf{M}) \sim \mathbb{D}_{train}} [\mathcal{L}_\xi^R(\mathbf{F}, \mathbf{M}, \mathcal{N}_\xi^R(\mathbf{F}, \mathbf{M}), \mathcal{N}_\xi^R(\mathbf{M}, \mathbf{F}))]] \quad (5.7)$$

其中  $\mathbf{F}$  和  $\mathbf{M}$  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:

$$\mathcal{L}_\xi^R(\mathbf{F}, \mathbf{M}, \mathcal{N}_\xi^R(\mathbf{F}, \mathbf{M}), \mathcal{N}_\xi^R(\mathbf{M}, \mathbf{F})) = \mathcal{L}_\xi^{BiC}(\mathbf{F}, \mathbf{M}, \psi, \psi^{-1}) + \lambda^{SeC} \mathcal{L}_\xi^{SeC}(\mathbf{F}, \mathbf{M}, \psi, \psi^{-1}), \quad (5.8)$$

其中,  $\lambda^{SeC}$  是  $\mathcal{L}_\xi^{SeC}$  的权重。

### 5.3.2 空间-风格采样过程 (S3P)

如图5-5, S3P构建了一种在基于配准的生成过程中密集采样的策略, 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<sup>[127]</sup>, 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  $\mathbb{D}_{train}^U$ , 以及一个拥有  $N$  ( $N, N \ll M$ ) 个有标签图像的模板数据集  $\mathbb{D}_{train}^A$ 。因此, 训练数据集  $\mathbb{D}_{train}$  可以描述为:

$$\mathbb{D}_{train} = \{ \{(\mathbf{A}, \mathbf{y})\}^N, \{\mathbf{U}\}^M \} (N \ll M), \quad (5.9)$$

S3P从无标签数据集  $\mathbb{D}_{train}^U$  中随机采样两个拥有新的空间和风格特征的无标签图像  $\mathbf{U}_j$ ,  $\mathbf{U}_k$ , 然后从模板数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  $(\mathbf{A}_i, \mathbf{y}_i)$ :  $\{(\mathbf{A}_i, \mathbf{y}_i), \mathbf{U}_j, \mathbf{U}_k\} \sim \mathbb{D}_{train}$ 。

对于生成更加多样的空间特征, S3P将一个无标签图像  $\mathbf{U}_j$  和一个模板图像  $\mathbf{A}_i$  输入到配准网络  $\mathcal{N}_\xi^R$  中, 得到一个形变场  $\psi_{ij}$  (公式5.10), 该形变场能够将模板图像的空间特

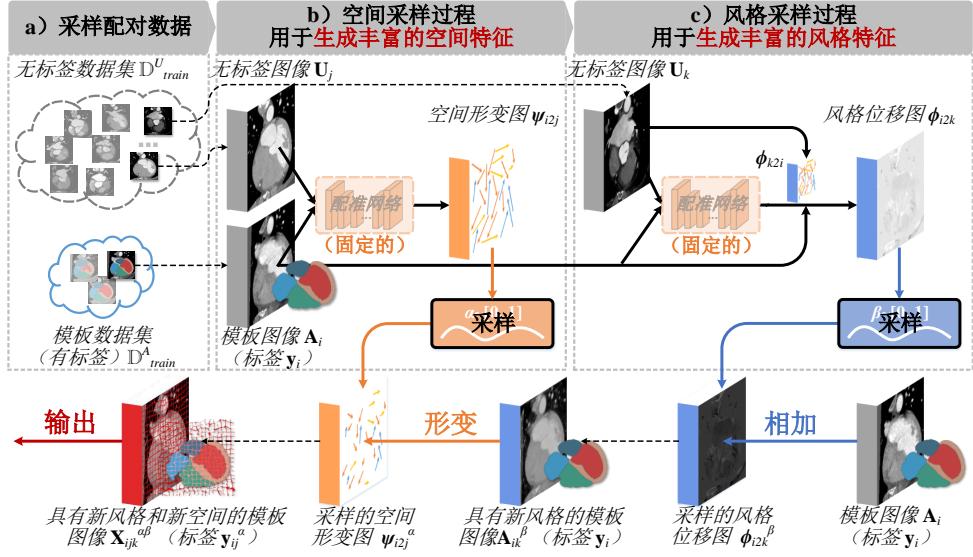


图 5-5: S3P构建了一个基于配准概率生成过程，通过对流形上图像间空间和风格路径进行采样，生成更加丰富的空间和风格特征，提高了生成数据的多样性。

征转化为无标签图像  $\mathbf{U}_j$  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：

$$\psi_{i2j} = \mathcal{N}_{\xi}^R(\mathbf{A}_i, \mathbf{U}_j), \psi_{i2j}^{\alpha} = \alpha \psi_{i2j}. \quad (5.10)$$

对于生成更加多样的风格特征，S3P将另一个无标签图像  $\mathbf{U}_k$  通过配准网络对齐到模板图像  $\mathbf{A}_i$  上，得到形变后的无标签图像  $\mathbf{U}_{k2i} = \mathcal{N}_{\xi}^R(\mathbf{U}_k, \mathbf{A}_i)(\mathbf{U}_k)$ 。然后，通过对齐的两个图像 ( $\mathbf{U}_{k2i}, \mathbf{A}_i$ ) 之间进行像素值相减（公式5.11），获得模板图像的风格位移图  $\phi_{i2k}$ ，从而实现了从模板图像到无标签图像  $\mathbf{U}_k$  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风格之间流形上的采样，获得更加多样的风格信息：

$$\phi_{i2k} = \mathbf{U}_{k2i} - \mathbf{A}_i, \phi_{i2k}^{\beta} = \beta \phi_{i2k}. \quad (5.11)$$

最后，采样获得的风格位移图  $\phi_{i2k}^{\beta}$  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  $\psi_{i2j}^{\alpha}$  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 ( $\mathbf{X}_{ijk}^{\alpha\beta}, \mathbf{y}_{ij}^{\alpha}$ )：

$$\mathbf{X}_{ijk}^{\alpha\beta} = \psi_{i2j}^{\alpha}(\mathbf{A}_i + \phi_{i2k}^{\beta}), \mathbf{y}_{ij}^{\alpha} = \psi_{i2j}^{\alpha}(\mathbf{y}_i). \quad (5.12)$$

如果  $\beta = 0$ ，则生成的图像是仅空间形变后的模板图像  $\mathbf{A}_{ij}^{\alpha}$ ；如果  $\alpha = 0$ ，则生成的图像是仅风格变换后的模板图像  $\mathbf{A}_{ik}^{\beta}$ 。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5.1所示，与其他典型的LRLS方法相比<sup>[214;127;126;125]</sup>，BRBS具有包括图像级、空间级和风格级的三个级别采样，拥有更多样化的数据生成能力。它使用三

表 5.1: 不同典型LRLS方法的概率生成过程的概率分布表明, 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	生成随机性
DeepAtlas <sup>[214]</sup>	$p((\mathbf{A}_i, \mathbf{y}_i))p(\mathbf{U}_j)$
PC-Reg-RT <sup>[125]</sup>	$p(\psi_{i2j} \mathbb{R}, \mathbf{A}_i, \mathbf{U}_j)p((\mathbf{A}_i, \mathbf{y}_i))p(\mathbf{U}_j)$
DataAug <sup>[127]</sup>	$p((\mathbf{A}_i, \mathbf{y}_i))p(\mathbf{U}_j)p(\mathbf{U}_k)$
DeepRS <sup>[126]</sup>	$p(\alpha)p(\psi_{i2j} \mathbb{R}, \mathbf{A}_i, \mathbf{U}_j)p((\mathbf{A}_i, \mathbf{y}_i))p(\mathbf{U}_j)$
<b>Our BRBS</b>	$p(\beta)p(\phi_{i2k} \mathbb{R}, \mathbf{A}_i, \mathbf{U}_k)p(\alpha)p(\psi_{i2j} \mathbb{R}, \mathbf{A}_i, \mathbf{U}_j)p((\mathbf{A}_i, \mathbf{y}_i))p(\mathbf{U}_j)p(\mathbf{U}_k)$

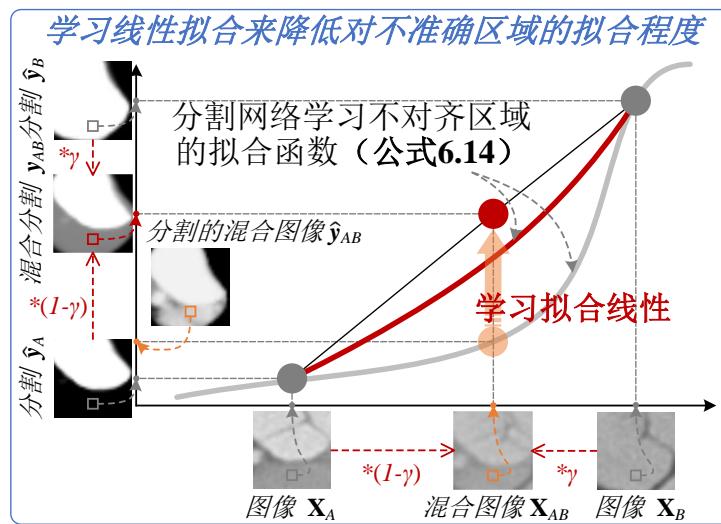


图 5-6: MMR提高了分割网络训练的鲁棒性。从优化的视角来看, 它将两个图像混合以模拟不对齐失真, 并训练深度学习模型在不对齐区域学习线性函数, 从而减少对这些不准确信息的拟合。

种随机性, 包括图像间空间和风格配对的随机性(DataAug,  $p((\mathbf{A}_i, \mathbf{y}_i))p(\mathbf{U}_j)p(\mathbf{U}_k)$ ), 配准网络训练过程的随机性(DeepRS, PC-Reg-RT,  $p(\psi_{i2j}|\mathcal{N}_{\xi}^R, \mathbf{A}_i, \mathbf{U}_j)$ ), 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。

### 5.3.3 混合不对齐正则化 (MMR)

如图5-7所示, MMR策略在使用生成的伪标签数据 $(\mathbf{X}_{ijk}^{\alpha\beta}, \mathbf{y}_{ij}^{\alpha})$ 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 $\mathbf{X}_A$ 和 $\mathbf{X}_B$ , 生成一种混合图像 $\mathbf{X}_{AB}$ 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5-7)。该策略首先从训练数据集 $\mathbb{D}_{train}$ 中采样两个图像 $\mathbf{X}_A$ 和 $\mathbf{X}_B$ , 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 $\mathbf{X}_{AB}$ 以模拟生成数据中由于配准

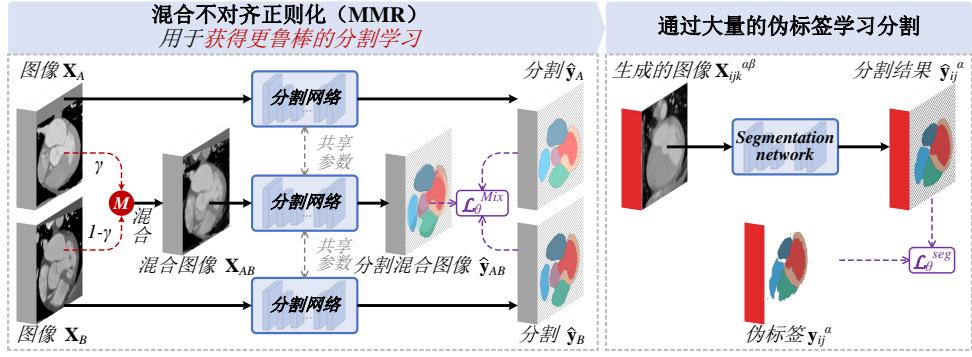


图 5-7: 分割网络的整体学习过程。它通过MMR的Mix损失 $\mathcal{L}^{Mix}$ 和分割损失 $\mathcal{L}^{seg}$ 来学习医学图像分割。

不对齐导致的风格失真。接着，分割模型 $\mathcal{N}_\theta^S$ 对两个原始图像 $\mathbf{X}_A, \mathbf{X}_B$ 和混合图像 $\mathbf{X}_{AB}$ 分别进行推理预测，获得它们的分割结果 $\hat{\mathbf{y}}_A = \mathcal{N}_\theta^S(\mathbf{X}_A)$ ,  $\hat{\mathbf{y}}_B = \mathcal{N}_\theta^S(\mathbf{X}_B)$ 和分割混合图 $\hat{\mathbf{y}}_{AB} = \mathcal{N}_\theta^S(\mathbf{X}_{AB})$ 。然后，分割结果图 $\hat{\mathbf{y}}_A, \hat{\mathbf{y}}_B$ 也通过随机值 $\gamma$ 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 $\mathbf{y}_{AB}$ 。最后，通过混合损失 $\mathcal{L}_\theta^{Mix}$ ，将混合图像的分割结果图 $\hat{\mathbf{y}}_{AB}$ 与混合分割结果图 $\mathbf{y}_{AB}$ 约束抑制，从而提高分割网络应对生成数据失真的鲁棒性。

$$\mathcal{L}_\theta^{Mix}(\hat{\mathbf{y}}_A, \hat{\mathbf{y}}_B, \hat{\mathbf{y}}_{AB}) = \| \gamma \hat{\mathbf{y}}_A + (1 - \gamma) \hat{\mathbf{y}}_B - \hat{\mathbf{y}}_{AB} \|^2. \quad (5.13)$$

如图5-6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 $(\mathbf{X}_A, \hat{\mathbf{y}}_A), (\mathbf{X}_B, \hat{\mathbf{y}}_B)$ 中不对齐区域是图中分割网络函数 $\mathcal{N}^S$ 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 $(\mathbf{X}_{AB}, \mathbf{y}_{AB})$ ）。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 $\hat{\mathbf{y}}_{AB}$ 和混合分割结果图 $\mathbf{y}_{AB}$ 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：

$$\mathcal{N}_\theta^S(\gamma \mathbf{X}_A + (1 - \gamma) \mathbf{X}_B) \Leftrightarrow \gamma \mathcal{N}_\theta^S(\mathbf{X}_A) + (1 - \gamma) \mathcal{N}_\theta^S(\mathbf{X}_B). \quad (5.14)$$

在实验中，随机值 $\gamma$ 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 $\mathcal{L}_\theta^{seg}$ 学习生成的伪标签数据 $(\mathbf{X}_{ijk}^{\alpha\beta}, \mathbf{y}_{ij}^\alpha)$ 。如图5-7所示，分割损失 $\mathcal{L}_\theta^{seg}$ 计算生成图像的分割结果 $\hat{\mathbf{y}}_{ij}^\alpha$ 与生成标签 $\mathbf{y}_{ij}^\alpha$ 之间的距离以训练分割模型。在本文中，我们采用DSC损失<sup>[214]</sup>作为分割损失。

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 $\mathcal{L}^S$ :

$$\hat{\theta} = \arg \min_{\theta} [\mathbb{E}_{((\mathbf{A}_i, \mathbf{y}_i), \mathbf{U}_j, \mathbf{U}_k, \mathbf{X}_A, \mathbf{X}_B) \sim \mathbb{D}_{train}} [\mathcal{L}_\xi^S(\hat{\mathbf{y}}_{ij}^\alpha, \mathbf{y}_{ij}^\alpha, \hat{\mathbf{y}}_{AB}, \hat{\mathbf{y}}_A, \hat{\mathbf{y}}_B)]]. \quad (5.15)$$

分割网络的完整损失为：

$$\mathcal{L}_{\theta}^S(\hat{\mathbf{y}}_{ij}, \mathbf{y}_{ij}, \hat{\mathbf{y}}_{AB}, \hat{\mathbf{y}}_A, \hat{\mathbf{y}}_B) = \omega^{Mix} \mathcal{L}_{\theta}^{Mix}(\hat{\mathbf{y}}_A, \hat{\mathbf{y}}_B, \hat{\mathbf{y}}_{AB}) + \mathcal{L}_{\theta}^{seg}(\hat{\mathbf{y}}_{ij}^{\alpha}, \mathbf{y}_{ij}^{\alpha}) \quad (5.16)$$

其中， $\omega^{Mix}$ 是完整的分割网络损失中混合损失的权重。

## 5.4 实验设置与结果分析

### 5.4.1 实验方案

**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**<sup>[231]</sup>：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<sup>[231]</sup>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<sup>[232]</sup>数据集（60张无心脏标签图像），CAT08挑战赛<sup>[233]</sup>数据集（32张带有心脏结构标签的图像，标签来自该连接<sup>1</sup>）。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1一个标签情况的验证，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1一个标签（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**<sup>[234]</sup>：该任务在T1 MRI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<sup>[235]</sup>（包含103个带有脑组织标签的T1脑MR图像）。遵从论文<sup>[137]</sup>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。

**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）<sup>[90]</sup>, SegNet（2017）<sup>[222]</sup>, U-Net++（2019）<sup>[32]</sup>, DBN（2020）<sup>[6]</sup>）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）<sup>[236]</sup>, MASSL（2019）<sup>[52]</sup>, DPA-DBN（2020）<sup>[6]</sup>, CPS（2021）<sup>[182]</sup>）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）<sup>[220]</sup>, 标签约束的VM（LC-VM, 2019）<sup>[224]</sup>, LT-Net（2020）<sup>[215]</sup>）以显示由于模板和目标图像之间的

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

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）<sup>[214]</sup>, DataAug（2019）<sup>[127]</sup>, DeepRS（2020）<sup>[126]</sup>, PC-Reg-RT（2021）<sup>[125]</sup>, VAEAug（2021）<sup>[137]</sup>）进而展示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<sup>[90]</sup>，以避免网络架构的不同对性能造成干扰而破坏公平性。

**实施细节和评价指标** 我们基于PyTorch<sup>[213]</sup>实施BRBS框架，在24GB显存的NVIDIA GeForce RTX 3090 GPU上训练和测试。我们将 $\alpha$ ,  $\beta$ 和 $\gamma$ 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<sup>[224]</sup>和论文<sup>[230]</sup>，我们设置 $\lambda^{smooth} = 1$ ,  $\lambda^I = 0.1$ ,  $\lambda^{SeC} = 100$ （参照5.4.3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5.4.3节）以平衡分割网络的训练。训练过程采用Adam优化器<sup>[183]</sup>进行优化，学习率设置为 $1 \times 10^{-4}$ ，批大小为1，以实现快速收敛<sup>[125]</sup>。我们通过AntsPy<sup>2</sup>对所有图像执行仿射变换以统一图像的空间位置。

本章实验从两个方面评估模型的分割性能<sup>[184]</sup>: 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ASD来评估表面的重合度。ASD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ASD，还计算每个变形体素位置上位移向量场的雅可比矩阵 $J_\psi < 0$ 的位置站全部位置的百分比 $J_\psi < 0$ ，从而评估形变的失真程度（越低越好）。

### 5.4.2 对比分析

本实验进行了定量对比分析和定性对比分析以验证BRBS方法的优越性。

**定量对比分析** 如表5.2所示，基于LRLS的方法在少样本医学图像分割任务中表现出显著的优越性。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。这是因为非常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

<sup>2</sup><https://github.com/ANTsX/ANTsPy>

表 5.2: 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ASD无法计算。

方法	类型	a) CT图像上的心脏结构分割				b) MRI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% $\uparrow$	ASD <sub>mm</sub> $\downarrow$						
3D U-Net <sup>[90]</sup>	LS	63.8 $\pm$ 16.3	6.13 $\pm$ 3.46	84.3 $\pm$ 9.6	2.43 $\pm$ 2.14	54.4 $\pm$ 10.8	2.94 $\pm$ 1.23	69.5 $\pm$ 8.8	1.59 $\pm$ 0.84
SegNet <sup>[222]</sup>		57.5 $\pm$ 17.4	7.01 $\pm$ 4.53	78.8 $\pm$ 10.5	2.68 $\pm$ 1.72	52.3 $\pm$ 4.9	3.18 $\pm$ 0.37	62.7 $\pm$ 7.0	1.98 $\pm$ 0.72
U-Net++ <sup>[32]</sup>		42.9 $\pm$ 20.5	9.18 $\pm$ 3.78	84.0 $\pm$ 8.6	2.51 $\pm$ 2.26	51.2 $\pm$ 10.6	2.33 $\pm$ 1.06	66.4 $\pm$ 12.7	2.02 $\pm$ 1.62
DBN <sup>[6]</sup>		48.8 $\pm$ 16.5	10.70 $\pm$ 4.10	78.9 $\pm$ 12.0	3.90 $\pm$ 3.12	23.5 $\pm$ 15.9	13.83 $\pm$ 7.26	80.2 $\pm$ 5.6	0.92 $\pm$ 0.30
UA-MT <sup>[236]</sup>		54.8 $\pm$ 17.0	9.44 $\pm$ 4.77	66.4 $\pm$ 16.2	4.69 $\pm$ 2.27	36.7 $\pm$ 8.4	8.69 $\pm$ 2.29	75.5 $\pm$ 3.4	1.31 $\pm$ 0.95
CPS <sup>[182]</sup>		70.7 $\pm$ 9.4	4.01 $\pm$ 1.73	87.4 $\pm$ 5.4	1.40 $\pm$ 0.76	25.3 $\pm$ 1.2	unable	37.1 $\pm$ 1.8	unable
MASSL <sup>[52]</sup>	SLS	57.2 $\pm$ 12.5	13.86 $\pm$ 3.16	77.4 $\pm$ 8.7	9.07 $\pm$ 3.11	74.0 $\pm$ 3.1	1.32 $\pm$ 0.35	80.5 $\pm$ 3.1	0.92 $\pm$ 0.43
DPA-DBN <sup>[6]</sup>		49.0 $\pm$ 14.4	10.47 $\pm$ 3.81	68.0 $\pm$ 14.5	5.75 $\pm$ 3.89	28.1 $\pm$ 7.6	7.75 $\pm$ 1.78	68.7 $\pm$ 8.2	3.90 $\pm$ 2.39
VM <sup>[220]</sup>		77.6 $\pm$ 6.0	2.49 $\pm$ 0.73	81.0 $\pm$ 6.1	2.13 $\pm$ 0.78	78.7 $\pm$ 1.8	0.73 $\pm$ 0.07	83.1 $\pm$ 1.8	0.56 $\pm$ 0.08
LC-VM <sup>[224]</sup>		-	-	81.7 $\pm$ 6.0	2.04 $\pm$ 0.77	-	-	83.0 $\pm$ 1.8	0.56 $\pm$ 0.07
LT-Net <sup>[215]</sup>		67.2 $\pm$ 6.5	3.55 $\pm$ 0.90	77.8 $\pm$ 7.8	2.25 $\pm$ 0.95	76.9 $\pm$ 1.5	0.75 $\pm$ 0.51	82.6 $\pm$ 1.2	0.57 $\pm$ 0.05
DeepAtlas <sup>[214]</sup>		85.4 $\pm$ 4.5	1.59 $\pm$ 0.56	87.9 $\pm$ 4.3	1.30 $\pm$ 0.57	73.0 $\pm$ 2.4	1.02 $\pm$ 0.10	79.3 $\pm$ 2.6	0.74 $\pm$ 0.12
DataAug <sup>[127]</sup>	LRLS	81.4 $\pm$ 5.2	2.23 $\pm$ 0.67	82.2 $\pm$ 5.2	2.04 $\pm$ 0.73	81.3 $\pm$ 1.4	0.69 $\pm$ 0.06	83.9 $\pm$ 1.2	0.55 $\pm$ 0.06
DeepRS <sup>[126]</sup>		73.4 $\pm$ 12.3	3.40 $\pm$ 1.92	87.0 $\pm$ 5.0	1.60 $\pm$ 0.90	55.9 $\pm$ 12.0	1.81 $\pm$ 0.91	73.0 $\pm$ 5.9	0.93 $\pm$ 0.25
PC-Reg-RT <sup>[125]</sup>		85.5 $\pm$ 4.7	1.55 $\pm$ 0.63	88.5 $\pm$ 4.9	1.23 $\pm$ 0.72	66.9 $\pm$ 3.6	1.38 $\pm$ 0.19	73.1 $\pm$ 3.1	1.09 $\pm$ 0.17
VAEAug <sup>[137]</sup>		75.5 $\pm$ 11.0	4.29 $\pm$ 2.12	-	-	74.8 $\pm$ 12.2	1.71 $\pm$ 2.71	-	-
<b>BRBS</b>	LRLS	<b>89.2<math>\pm</math>3.4</b>	<b>1.24<math>\pm</math>0.50</b>	<b>91.1<math>\pm</math>3.9</b>	<b>0.93<math>\pm</math>0.57</b>	<b>85.7<math>\pm</math>1.0</b>	<b>0.49<math>\pm</math>0.04</b>	<b>87.2<math>\pm</math>1.0</b>	<b>0.43<math>\pm</math>0.05</b>

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ASD(0.93mm, 0.43mm)。虽然PC-Reg-RT在大结构分割(任务(a))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大大限制了配准性能, 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))上非常受限。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-

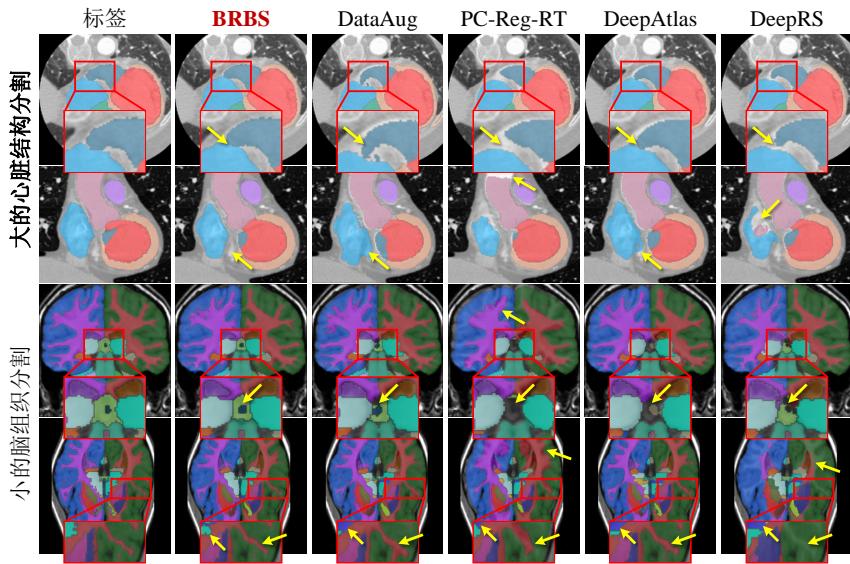


图 5-8: 定性评估显示, 与四种典型的LRLS方法相比, BRBS在5-shot设置下的两个少标签医学图像分割任务上都具有更好的视觉效果。

Reg-RT中的配准网络的输入基于其分割网络的输出, 当有标签图像数量非常少时, 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。

**定性对比分析** 图5-8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, 这些小结构的特征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。

表 5.3: 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% $\uparrow$	ASD $mm \downarrow$	DSC% $\uparrow$	ASD $mm \downarrow$	$ J_\psi  \leq 0\% \downarrow$	
					$84.3 \pm 9.6$	$2.43 \pm 2.14$	-	-	-	
	✓				$80.7 \pm 9.6$	$2.52 \pm 1.52$	$72.6 \pm 13.8$	$2.89 \pm 1.18$	$3.3 \pm 0.7$	
	✓		✓		$83.7 \pm 8.0$	$2.33 \pm 2.03$	$73.2 \pm 13.8$	$2.84 \pm 1.17$	$3.2 \pm 0.7$	
	✓		✓	✓	$88.1 \pm 4.7$	$1.25 \pm 0.63$	$73.0 \pm 13.9$	$2.87 \pm 1.20$	$3.5 \pm 0.8$	
	✓		✓		✓	$84.4 \pm 6.5$	$1.85 \pm 0.89$	$73.5 \pm 13.8$	$2.84 \pm 1.18$	$3.7 \pm 0.8$
✓	✓		✓	✓		$90.0 \pm 3.8$	$1.04 \pm 0.49$	$85.9 \pm 13.5$	$1.33 \pm 0.67$	$6.2 \pm 1.2$
✓	✓	✓	✓	✓		$90.4 \pm 3.4$	$1.00 \pm 0.44$	$86.0 \pm 13.5$	$1.31 \pm 0.64$	$2.5 \pm 1.1$
✓	✓	✓	✓	✓	✓	<b><math>91.1 \pm 3.9</math></b>	<b><math>0.93 \pm 0.57</math></b>	<b><math>86.7 \pm 13.6</math></b>	<b><math>1.22 \pm 0.62</math></b>	<b><math>1.7 \pm 0.8</math></b>

#### 5.4.3 消融实验和模型分析

**消融实验** 如表5.3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ASD。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。加入风格采样后显著地提升了分割模型性能，获得了88.1%和1.25mm的DSC与ASD。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，有效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 $|J_\psi| \leq 0$ 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 $|J_\psi| \leq 0$ 。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ASD。

**超参数分析** 如图5-9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。1) MMR显著提高了DataAug<sup>[127]</sup>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这种非常容易受到配准对齐失真干扰的模型上评估MMR的影响。由

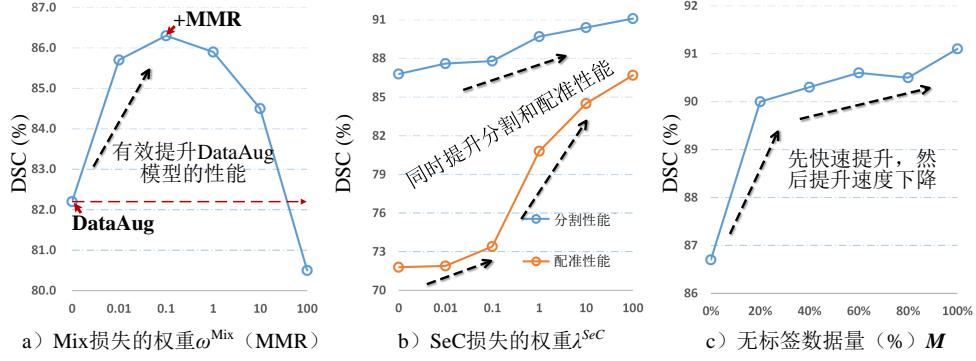


图 5-9: 该实验在5-shot心脏结构分割任务上进行了本章中新引入的超参数的分析。分析了三个超参数: **a)** MMR中Mix损失的权重  $\omega^{Mix}$ , **b)** SeC损失的权重  $\lambda^{SeC}$ , **c)** 无标签数据集的大小  $M$ 。

表 5.4: 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, ASD) 和更低的扭曲失真 ( $|J_\psi| \leq 0$ )。“Trad”表示传统方法, “DL”表示深度学习方法。

方法	类型	DSC $_{\% \pm std} \uparrow$	ASD $_{mm \pm std} \downarrow$	$ J_\psi  \leq 0 \%, \pm std \downarrow$
Initial		$62.1 \pm 8.7$	$4.37 \pm 1.34$	-
B样条法 <sup>[237]</sup>	Trad	$77.0 \pm 11.2$	$2.61 \pm 1.57$	$8.4 \pm 9.7$
SyN <sup>[216]</sup>	Trad	$79.2 \pm 8.9$	$2.33 \pm 0.92$	<b>0</b>
VM <sup>[220]</sup>	DL	$72.7 \pm 13.9$	$2.89 \pm 1.20$	$3.5 \pm 0.7$
LC-VM <sup>[224]</sup>	DL	$73.0 \pm 13.9$	$2.85 \pm 1.19$	$3.1 \pm 0.6$
DeepAtlas <sup>[214]</sup>	DL	$79.7 \pm 13.6$	$2.10 \pm 0.96$	$2.5 \pm 0.6$
PC-Reg-RT <sup>[125]</sup>	DL	$80.8 \pm 13.6$	$1.94 \pm 0.91$	$0.5 \pm 0.3$
DeepRS <sup>[126]</sup>	DL	$76.9 \pm 13.3$	$2.39 \pm 0.99$	$5.3 \pm 1.3$
LT-Net <sup>[215]</sup>	DL	$68.2 \pm 13.9$	$3.26 \pm 1.29$	$2.5 \pm 0.3$
<b>BRBS</b>	DL	<b><math>86.7 \pm 13.6</math></b>	<b><math>1.22 \pm 0.62</math></b>	$1.7 \pm 0.8$

于MMR对不准确信息的抑制作用, 当权重  $\omega^{Mix}$  为 0.1 时, DataAug 模型取得了最高和显著的提升。但当将权重  $\omega^{Mix}$  增加到 100 时, MMR 过于强大的正则化作用导致模型欠拟合, 削弱了性能 (甚至低于原始的 DataAug)。2) 随着其权重  $\lambda^{SeC}$  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。3) 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 有效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。

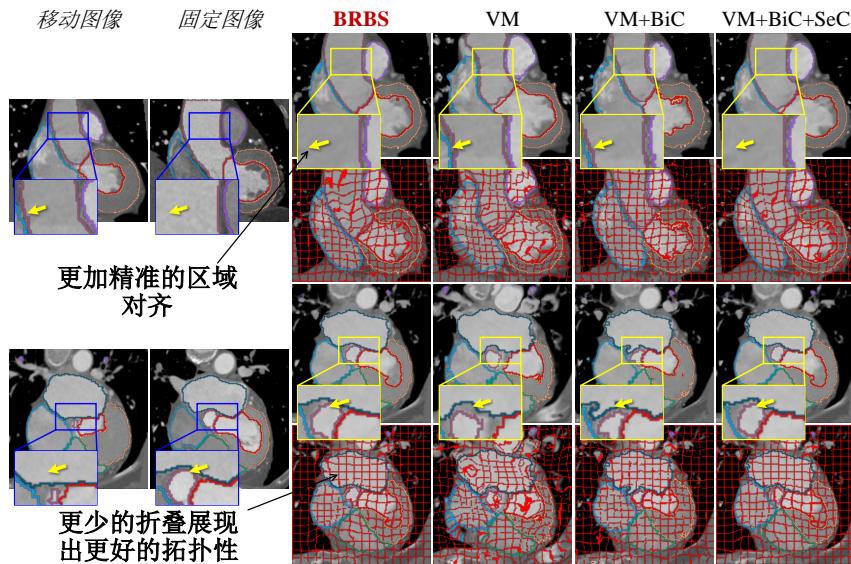


图 5-10: KCC通过BiC和SeC策略优化配准网络的学习，使其具有更好的保持拓扑性和语义区域对齐的能力。

**配准性能分析** 如表5.4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<sup>[237]</sup>和SyN<sup>[216]</sup>）、两个基于深度学习的配准模型（VM<sup>[220]</sup>和LC-VM<sup>[224]</sup>）以及三个LRLS模型的配准部分（DeepAtlas<sup>[214]</sup>，PC-Reg-RT<sup>[125]</sup>和DeepRS<sup>[126]</sup>）。B样条法在DSC、ASD和 $|J_\psi| \leq 0$ 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 $|J_\psi| \leq 0$ ，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 $|J_\psi| \leq 0$ 。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ASD（1.22mm）和相对较低的 $|J_\psi| \leq 0$ （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。

**KCC分析** 如图5-10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。1) 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5.4所示，它的 $|J_\psi| \leq 0$ 较高（3.5%）而准确性较低（72.7%）。2) BiC利

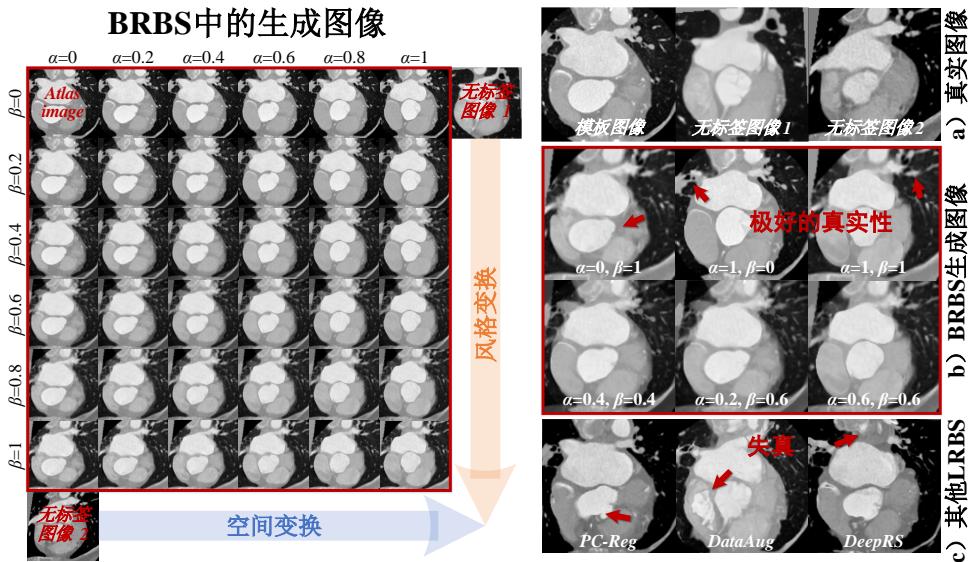


图 5-11: BRBS 中生成的图像具有很高的多样性和真实性。(左) S3P 通过空间和风格密集采样,使得生成的图像特征覆盖图像间的风格流形和空间结构流形上的位移路径,从而获得更加多样的空间和风格特征的数据。(右) 从 a) 到 c) 分别是来自真实世界的图像、来自 S3P 生成图像和来自其他 LRLS 方法的图像。

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。3) 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。

**生成数据的真实性分析** 如图 5-11 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 生成的图像具有良好的真实性。1) 与其他基于 LRLS 的方法 ([125;126;127]) 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。2) 与真实世界的图像 (图 5-11 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。

**生成数据的多样性分析** 如图 5-12 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。1) 在没有 S3P 的情况下 [214], 所生成的图像仅具有

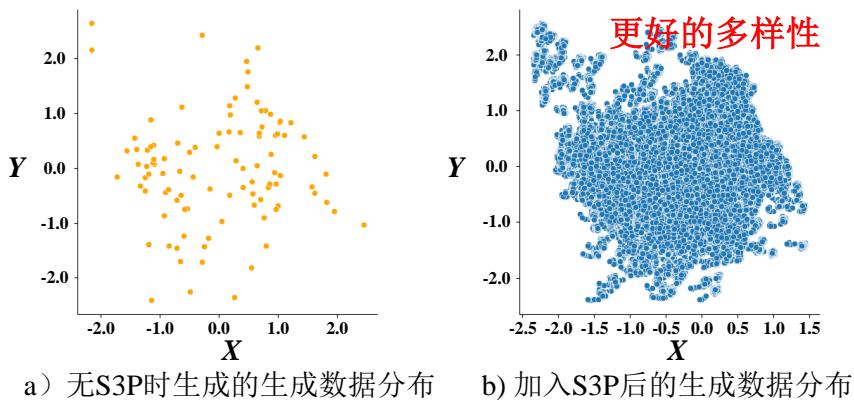


图 5-12: 从风格网络瓶颈层提取的特征降维 (PCA) 后的散点图展示了 S3P 对生成数据多样性的促进作用。**a)** 当没有 S3P 时<sup>[214]</sup>, 生成数据分布稀疏离散, 多样性较低。**b)** 当加入 S3P 时, BRBS 生成了具有良好多样性的数据分布。

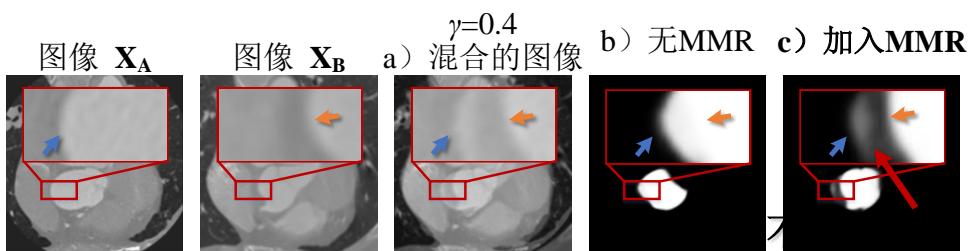


图 5-13: MMR 约束分割网络感知不对齐的区域, 降低这些不准确信息的干扰。**a)** 它通过加权融合两个图像来模拟生成具有不对齐区域的图像。**b)** 当没有 MMR 的时候, 3D U-Net 无法感知不对齐的区域, 使其学习到不对齐区域中不准确的信息。**c)** 加入 MMR 后, BRBS 拥有感知这些不对齐区域的能力, 并且学习降低对这些区域的拟合程度, 获得了更好的鲁棒性。

稀疏的特征分布, 多样性较差。这是因为数据生成过程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。2) S3P构建了一种空间和风格连续的分布并在该分布上密集采样, 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。3) 从视觉上看, 图5-11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 $\alpha$ 和 $\beta$ 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。

**MMR的正则化能力分析** 如图5-13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 $\mathbf{X}_A$ 和 $\mathbf{X}_B$ 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。

## 5.5 本章小结

本章研究了一种新颖的少标签通用医学图像分割学习框架，即BRBS框架。该框架面向人体解剖结构的分割任务，仅需要少量标签即可获得有效的分割结果，有效地提升了标签使用效率。BRBS框架的创新之处在于融合了分割配准任务的互补性，在多样性、真实性和鲁棒性方面推进了原始LRLS范式，进一步突破这类方法的瓶颈。其中KCC策略，用于训练配准网络，实现语义对齐和拓扑保持的配准，生成具有更高的空间和风格真实性的新数据；S3P策略，建模了模板与大量无标签图像之间的空间和风格变换路径，从而在生成数据时实现更多样化的空间和样式特征采样；MMR策略，模拟了潜在的配准失真并通过训练网络减少对不准确信息的拟合程度，从而提高分割学习的鲁棒性。本章的相关研究成果分别被发表在TNNLS和ECCV 2020上，详见作者简介中的文献[A3]和[A5]，相关源代码已开源在<https://github.com/YutingHe-list/BRBS>。

尽管本章的研究已在多种不同多结构分割任务中得到验证，并且成功将模型训练的标签需求降低到仅需要几个标签，具有更强的通用性和更高的标签效率，但其局限性在于仅适用于要求人体解剖结构一致性的多结构分割任务。一旦分割目标不符合这一前提条件，例如包含肿瘤等情况，配准可能无法进行，从而限制了该模型的通用性。此外，本章的研究是在已有少量标签的前提下进行的。然而，在实际情况中，医学图像分割任务的标注通常滞后，最初只有无标签的医学图像数据可用。因此，在下一章的研究中，我们将进一步探讨如何在无标签情况下进行医学图像的自监督学习，以利用学到的通用特征来促进对下游分割任务的学习。



## 第六章 基于自监督学习的三维医学图像表征预训练算法

上一章的研究探索了一种少标签医学图像通用分割算法，该算法不受限于特定的临床分割任务，而是适用于满足解剖结构一致性的分割对象，例如心脏结构分割和脑组织分割。在上一章的实验中，该算法仅使用一个或五个标签就能够取得非常有效的分割性能，展现该算法良好的通用性和强大的标签效率。然而，上一章的研究仍然受到特定的条件限制，一旦分割目标不满足前提条件，那么该算法将无法应用。此外，前三章的研究都依赖于一定数量分割标签来驱动医学图像分割学习，然而，在实际场景中，分割标签的标注通常存在滞后性，往往在医学图像数据收集完之后，需要花费大量时间来完成分割标注工作，因此在没有任何标注信息的情况下如何进行预学习是一个值得研究的问题。

因此，本章进一步从自监督医学图像表征预训练算法出发，针对如何在没有标签的情况下学习通用表征，以促进下游潜在的三维医学图像分割任务的学习，这个关键问题展开研究。本章从优化网络初始化参数 ( $\theta$ ) 的角度出发，首先分析了人体间的解剖结构一致性，以及学习图像间相似性在医学图像自监督学习中的重要性。随后，设计了一种叫做几何视觉相似性学习的自监督学习代理任务，该任务通过学习图像中不同语义区域的对应关系，驱动网络学习对应区域的特征表征一致，非对应区域的特征表征不一致，从而预训练一个具有强大表征能力的基础模型，从而促进任何潜在的三维医学图像分割任务的学习。

### 6.1 研究背景和动机

学习图像间相似性<sup>[238;239;240]</sup>对于三维医学图像（如CT、MRI）的自监督预训练（self-supervised pre-training, SSP）<sup>[112]</sup>至关重要。如图6-1所示，与在SSP中被广泛研究的自然图像不同，由于人体解剖结构的一致性<sup>[241]</sup>和3D视觉中完整的空间信息<sup>[242]</sup>，三维医学图像间共享大量相同的语义区域，为有效的SSP带来了强有力的先验。因此，其目标在于约束预训练网络以在没有注释的图像之间一致地表示这些相同的语义区域。这将对相同的语义特征带来强大的聚类效应，使得预训练网络具有良好的表征能力，最终能够进行有效地迁移学习潜在的下游具体医学图像分割任务。因此，它将有效地训练数据不足的巨大挑战<sup>[148]</sup>，即使在目标任务数据量较少的情况下依旧能够获得更强大的模型。

虽然现有的SSP的相关研究工作已经在其各自的目标任务中取得了良好的效果，但在三维医学图像的图像间相似性学习方面仍存在一定的局限性。1) 基于聚类的方法<sup>[243;244]</sup>根据图像在特征空间中的聚类模式来度量图像间特征的相似程度，并学习聚合

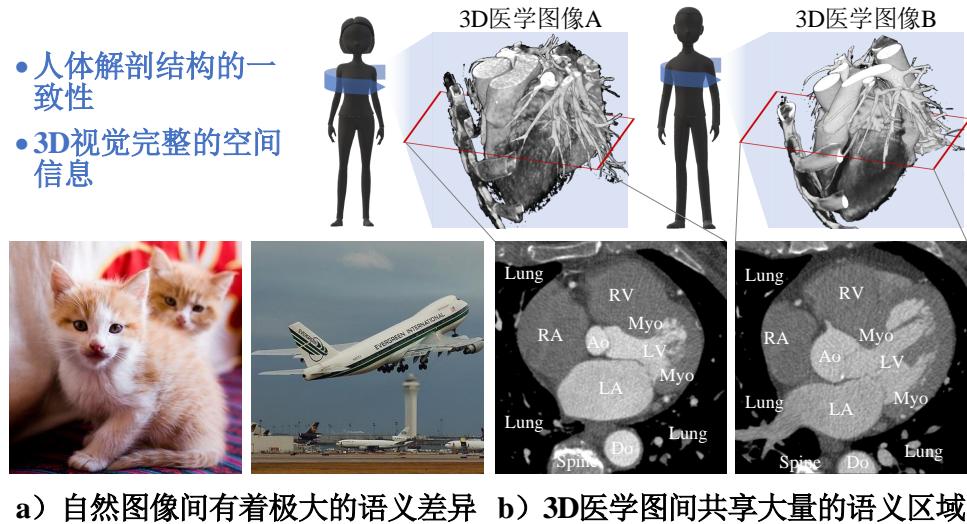


图 6-1: 三维医学图像与自然图像的差异图。a) 自然图像之间相似性较弱，具有较大的语义差异。b) 由于人体解剖结构的一致性和空间信息的完备性，三维医学图像之间共享大量相同的语义区域，图像间相似性大。

相同潜在类别的特征，分离不同潜在类别的特征。然而，基于聚类的SSP方法通常采用马氏距离或欧氏距离<sup>[245]</sup>作为距离度量函数，使距离度量受到图像内语义无关变化（如外观变化）的干扰（图6-2），大大限制了对潜在类别判断的准确性，限制最终的自监督学习。2) 基于对比学习的方法<sup>[153;152]</sup>直接通过对比来分离来自不同图像的特征，试图学习图像间的特征相异性。但是过度学习图像间的不同将使得网络容易把相同的语义区域表征为不同的特征，使得模型无法学习图像间的相似性。虽然其他一些对比学习的研究<sup>[151;153;246]</sup>通过去除了对比学习中的负对学习来避免这种情况，仅依靠学习同一图像变换前后的一致性来驱动SSP，但使这些方法仍然无法学习图像间相同语义的一致性。3) 基于生成的方法<sup>[147;150;177;148]</sup>通过人工设计的变换方法（例如，旋转<sup>[150]</sup>）来生成代理标签（pretext label），然后训练网络来预测这些代理标签以驱动SSP。但是这些通过人工设计的变换方法隐式地将人的偏好强加到SSP学习中，使得预训练的网络参数关注于代理标签所偏好特征，导致获得的预训练模型对场景变化敏感<sup>[147]</sup>。例如，旋转生成的代理标签<sup>[150]</sup>将使网络偏向于图像中对象的位置和姿态特征，当某些场景中对象的位置和姿态是一种与语义无关的信息时，这些图像将被错误表征。

让我们重新思考上述已有工作的局限性，大范围图像间相似性的错误度量是三维医学图像SSP面临的关键挑战。该挑战干扰了在自监督学习过程中对潜在相同语义对象的发现，阻碍了模型对相同语义区域学习一致的表征。如图6-2所示，与语义无关的图像特征变化使得三维医学图像间具有不同的外观，导致不同的语义区域具有相似的外观（图A和B之间的左心肌和右心房），而相同的语义区域在图像之间具有不同的外观（图B和C之间的右心房）。类似于基于聚类<sup>[243;244]</sup>这种利用欧氏距离或马氏距离在特征空间中的直接相似性度量，由于其度量过程中缺少语义相关的先验知识而变得敏感，容易受到语义无关变化的干扰。因此，在无监督学习的情况下，一旦被提取的特征受

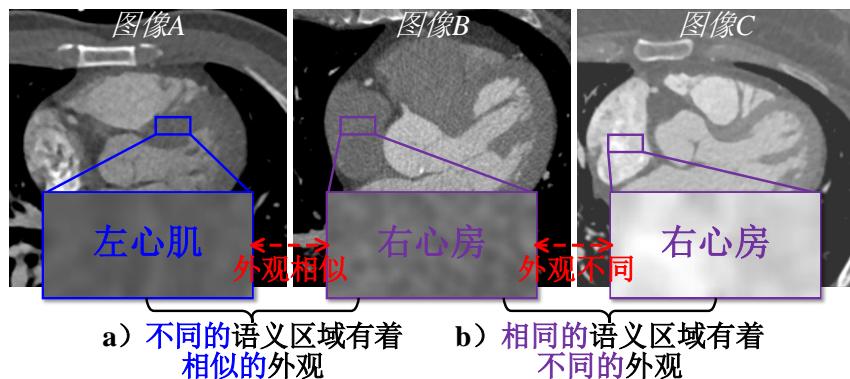


图 6-2: 度量图像间相似性的挑战。a) 图A的左心肌和图B的右心房之间存在着很大的相似性。b) 由于采集仪器扫描协议的变化，在图B和图C的两个右心房之间存在着很大的外观差异。

到外观干扰发生变化，就会导致大规模的语义相似性的错误测量，从而错误地聚合或分离语义对象，最终使得预训练模型学习到的表示偏离真实的语义分布。

三维医学图像中语义区域的拓扑不变性<sup>[247]</sup>为SSP提供了一种构建可靠的图像间相似性度量的潜在方法。如图6-3所示，由于人体内部解剖结构的一致性<sup>[241]</sup>，三维医学图像的语义区域之间具有一致的上下文拓扑（例如人体心脏的四个腔室具有固定的上下空间关系），并且不同图像中的相同语义区域在也具有相似的形状（例如图中的主动脉具有稳定的管状结构），因此图像中的语义区域之间能够构建稳定的拓扑结构，形成了一种可靠的先验知识。根据该语义区域拓扑不变性先验，我们能够将图像间的区域经由某种拓扑不变的映射<sup>[248]</sup>在空间中对齐，从而即使图像在外观上有大的变化也能够可靠地构建特征间的对应关系。一种直观的策略是使用配准（Registration）或几何匹配（Geometric Matching, GM）方法<sup>[125;249;250;223;251]</sup>来发现图像像素间的对应关系，并使用这些对应索引来约束网络输出对应像素位置区域的特征一致。然而，当配准或几何匹配出现错误时，这些索引的误差会带来误的对应关系，学习到错误的表征。在本章中我们提出了高效三位医学图像自监督表征预训练的新型假说：三维医学图像之间语义区域的拓扑不变性先验能够驱动模型发现图像间特征对应关系，从而引导网络可靠学习图像间的相似性。

在本章中，我们提出了一种新的自监督表征预训练学习方法，几何视觉相似性学习（Geometric Visual Similarity Learning, GVSL），用于可靠学习三维医学图像间的相似性。该算法将拓扑不变性先验嵌入到相似性度量和学习过程中，训练网络利用从两张三维医学图像上分别提取的特征来预测语义区域的对应关系，即配准，从而驱动学习图像间的一致性表征。由于拓扑不变性这种有效的先验知识的帮助，图像间相似性将在医学图像中语义区域固有的拓扑结构下被度量，从而避免了与语义无关的外观变化的干扰。当模型学习更准确地估计对应关系时，反向传播的梯度将训练神经网络在特征空间中聚合对应的特征以获得更一致的表示。为了有效地同时学习全局和局部的表征，我们进一步提出了新型的自监督学习模块，Z形匹配模块。该学习模块通过同时学习仿射变换（全局配准）和弹性形变（局部配准）<sup>[223]</sup>来驱动模型同时学习图像间全

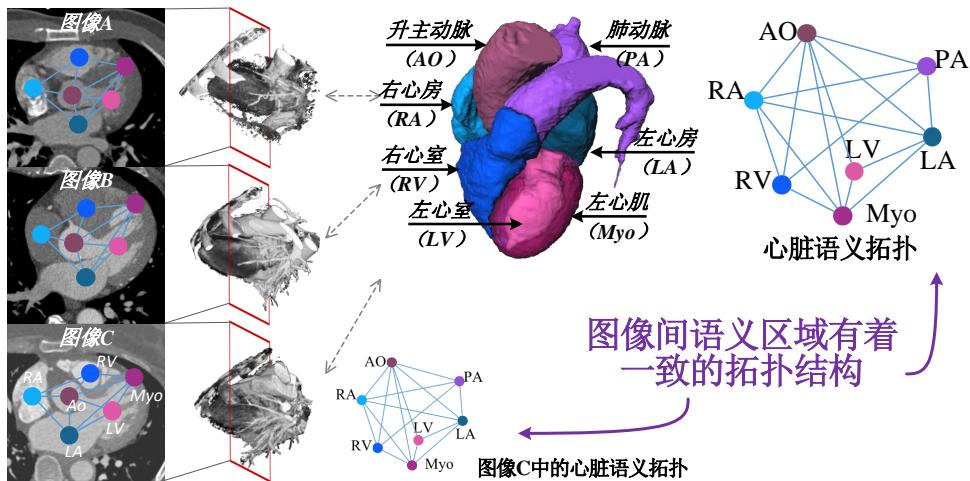


图 6-3: 三维医学图像之间语义区域的拓扑不变性。

局和局部相似性，从而使网络拥有同时高效学习全局和局部表征的能力，为对不同尺度目标结构的医学图像分割任务都提供强大的迁移能力。本章的贡献总结如下：

- 将图像间的相似性学习应用于三维医学图像自监督表示学习中，从而预训练神经网络以学习语义一致的表示，在无需标注的情况下，使模型拥有强大迁移能力的预训练参数。
- 提出了几何视觉相似性学习 (GVSL)，该方法将三维医学图像的拓扑不变性先验嵌入到相似性度量中，从而可靠地度量图像间的相似性，促进模型学习图像间相同语义区域的一致表示。
- 提出了一种新型的自监督学习模块，Z形匹配模块，该模块同时学习仿射变换和弹性形变配准，同时学习强大的全局和局部表示，最终获得了对不同语义粒度表示的有效优化，为不同尺度目标结构的医学图像分割任务提供了强大的迁移学习能力。

## 6.2 相关工作

**自监督预训练中的相似性学习** 相似性学习<sup>[238;252]</sup>的目标是学习图像中相似的对象以获得一致的表示，同时学习相异的对象以获得有鉴别力的表示。它作为视觉自监督表征学习的一个基本任务<sup>[112]</sup>，能够提升模型在学习下游任务时初始化的表征能力。正如我们已经在第6.1节所介绍的，学习图像间的相似性有三个主要范式。对比学习<sup>[153;152;253;254]</sup>范式认为同一图像的不同数据增强视图是相似的对象，而不同图像为相异的对象，从而学习他们的一致性和鉴别性来获得有效表征。但是，这类方法无法学习图像间的相似性，简单地认为来自不同图像的特征为相异对象并分离这些特征，最终将干扰三维医学图像的表示。基于聚类的方法<sup>[243;244]</sup>利用度量的方法来获得相似与相异对象簇，从

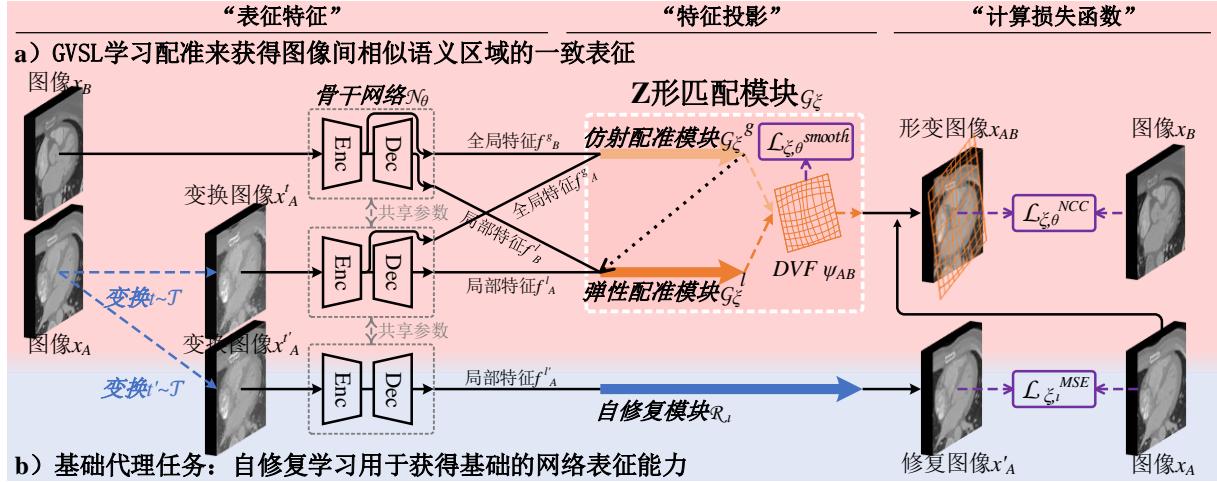


图 6-4: GVSL框架细节。a) GVSL从图像的特征中学习配准，通过反向传播的梯度驱动图像之间的相似性学习。b) 为了有效学习，在学习配准时，它还进一步融合了一个基础的自我修复任务，为网络的基础表征能力提供了一种预热，从而更加高效地驱动配准学习。

而聚集相似对象以学习图像间的相似性，这类方法很容易发生聚类错误的问题，从而将本身不同类别的图像认为是相似对象而聚集，相似性干扰了模型学习有效的表征。基于生成的方法<sup>[147;150;177;148;255;256]</sup>通过人工设计获得生成的代理标签，并约束网络预测这些标签。然而，它隐含地嵌入了手动设计的偏见，使网络忽略一些潜在的方面，并限制了在某些特定情况下的迁移。

**几何匹配与配准** 配准（或称为几何匹配）<sup>[125;126;107;249;250;223]</sup>将图像的语义区域对齐到同一空间坐标系中，从而提供两个图像之间的对应索引。它具有两个级别的变换：1) 仿射变换<sup>[257;223]</sup>能够在全局上对齐图像分布。它计算一个由旋转、缩放、平移和剪切操作组成的变换矩阵，将图像转换为在全局上对齐的形式。2) 弹性形变<sup>[125;223;107;251]</sup>能够在局部对齐图像细节。它预测一个体素级位移矢量场（Displacement vector field, DVF），获得图像间体素的对应关系，并通过空间变换操作<sup>[258]</sup>将图像局部对齐。最近，由于深度学习的发展，基于深度学习的配准<sup>[249;125;223]</sup>利用学习的方式获得端到端的对应预测能力，为我们的框架设计提供了一种潜在的方案。

### 6.3 方法

GVSL框架（图6-4）能够从无标签的三维医学图像中学习具有图像间相似性的通用视觉表征。它将拓扑不变性先验嵌入到相似度度量过程中，从而驱动模型学习可靠的相似语义区域的一致表征。

### 6.3.1 几何视觉相似性学习 (GVSL)

如图6-4 a) 所示, GVSL将学习图像间相似性的过程建模为网络所表征的特征中预测图像间配准的过程, 嵌入图像中语义区域的拓扑不变性先验, 从而在反向传播中训练网络来为相同的语义区域表征一致的特征。

**GVSL的方法细节描述** GVSL的目标是学习一个相同语义区域的特征在不同的图像间也具有的强大聚类效应的通用网络表征。如图6-4所示, 它使用两个共享权重的神经网络 $\mathcal{N}_\theta$ 来表征来自两个图像 $x_A, x_B$ 的特征 $f_A, f_B$ , 其中 $\theta$ 是网络内部的可学习权重。这些特征被进一步输入配准模块 $\mathcal{G}_\xi$  (即, 我们框架中的Z形匹配模块, Sec.6.3.2), 以学习图像之间语义区域的对应关系, 从而使 $\mathcal{N}_\theta$ 获得这些相同语义特征的一致表示。

给定一组三维医学图像 $\mathbb{D}_{train}$ , 从 $\mathbb{D}_{train}$ 中均匀随机采样两个图像 $\{x_A, x_B\} \sim \mathbb{D}_{train}$  ( $A$ 和 $B$ 代表不同的图像), 并从一个变换集 $\mathcal{T}$ 中采样一个变换操作 $t \sim \mathcal{T}$ 。在本研究中, 我们使用修补、体素局部抖动和图像强度非线性作为变换集 $\mathcal{T}$ 。GVSL通过将变换 $t$ 作用于图像上, 使 $x_A$ 变换为一个变化视图 $x'_A \triangleq t(x_A)$ , 从而提高了训练图像的多样性。网络输出 $x'_A$ 和 $x_B$ 的全局和局部特征 $\{f_A^g, f_A^l\} \triangleq \mathcal{N}_\theta(x'_A), \{f_B^g, f_B^l\} \triangleq \mathcal{N}_\theta(x_B)$ 。接着这些特征输入配准模块 $\mathcal{G}_\xi$ 以输出一个位移向量场(DVF) $\psi_{AB} \triangleq \mathcal{G}_\xi(f_A^l, f_B^l, f_A^g, f_B^g)$ , 该向量场中的每个向量对应于移动图像中的每个像素移动到固定图像空间中的位置, 即是两个图像之间体素的对应关系。最后通过空间变换, 将图像 $x_A$ 变换到图像 $x_B$ 以配准两个图像, 得到一个被几何匹配的形变图像 $x_{AB} \triangleq \psi_{AB}(x_A)$ 。其中, 空间变换操作我们采用了<sup>[258]</sup>的实现方式, 即对于图像 $x_A$ 中的每个体素 $p$ , 位移向量场 $\psi_{AB}$ 将 $p$ 位移至图像空间中的一个新的(亚像素)体素位置 $\psi_{AB}(x_A(p))$ 。然后, 线性插值该位移后在亚像素位置上的体素到八个相邻体素的整数位置。该过程表示为:

$$\psi_{AB}(x_A(p)) = \sum_{q \in \psi_{AB}(\mathbb{Z}(p))} x_A(q) \prod_{d \in x, y, z} (1 - |\psi_d(x_A(p)) - q_d|), \quad (6.1)$$

其中,  $\psi_{AB}(\mathbb{Z}(p))$ 是 $\psi(x_A(p))$ 的相邻体素,  $x, y, z$ 是三维图像的 $x, y, z$ 轴。

我们利用归一化互相关(NCC)来评估两个图像之间的对齐程度, 从而间接评估所预测的对应关系(DVF)的准确性。我们训练框架来最小化这个损失函数 $\mathcal{L}^{NCC}$ , 从而驱动模型优化这个对应关系以获得更好的配准:

$$\mathcal{L}_{\theta, \xi}^{NCC} \triangleq \sum_{p \in \Omega} \frac{(\sum_{p_i} (x_B(p_i) - \hat{x}_B(p))(x_{AB}(p_i) - \hat{x}_{AB}(p)))^2}{(\sum_{p_i} (x_B(p_i) - \hat{x}_B(p))^2)(\sum_{p_i} (x_{AB}(p_i) - \hat{x}_{AB}(p))^2)}, \quad (6.2)$$

其中 $p$ 是图像空间 $\Omega$ 中体素的位置,  $i$ 是 $p$ 的索引。 $\hat{x}_B(p)$ 是图像强度的局部均值:  $\hat{x}_B(p) = \frac{1}{n^3} \sum_{p_i} x_B(p_i)$  ( $\hat{x}_{AB}(p)$ 与其计算方式相同)。为了保持图像配准前后的拓扑不变性, 我们进一步在DVF上计算平滑损失 $\mathcal{L}^{smooth}$ , 以训练网络在保持拓扑一致的条件下学习语义区域的对应关系:

$$\mathcal{L}_{\theta, \xi}^{smooth} \triangleq \sum_{p \in \Omega} \|\nabla \psi_{AB}(p)\|^2. \quad (6.3)$$

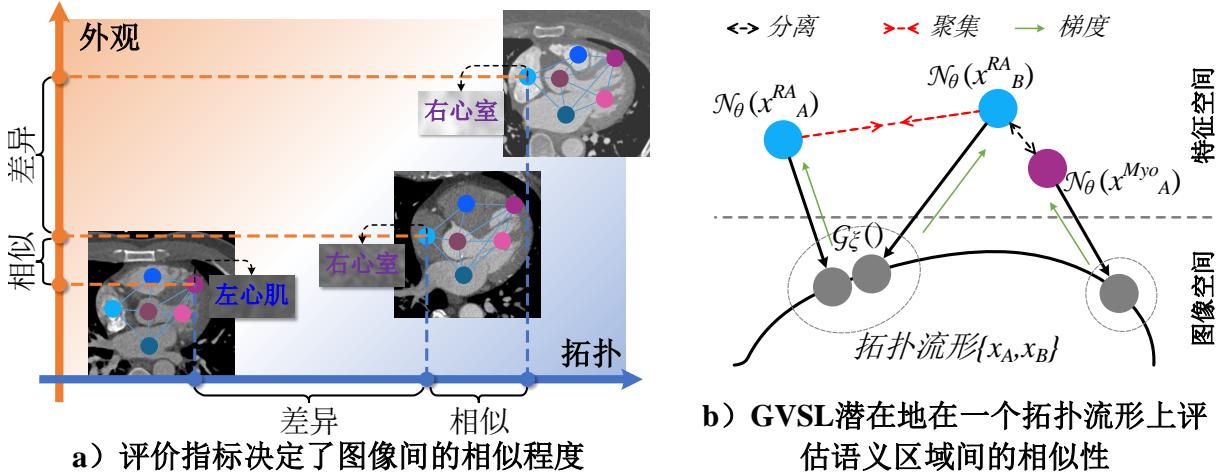


图 6-5: GVSL背后的直觉解释。配准中的拓扑不变性先验将拓扑流形嵌入到度量过程中，从而为图像之间的相似性提供了有效的度量，从而反向通过梯度引导模型学习相同语义的特征拥有一致的表征。

在每个训练步骤中，我们使用梯度进行反向传播，通过最小化  $\mathcal{L}_{\theta,\xi}^{GVSL} = \mathcal{L}_{\theta,\xi}^{NCC} + \mathcal{L}_{\theta,\xi}^{smooth}$  来优化网络权重  $\theta$  和配准模块的权重  $\xi$ 。因此，框架的总优化过程可被描述为一个基于梯度下降的过程  $\theta, \xi \leftarrow optimizer(\theta, \xi, \nabla_{\theta, \xi} \mathcal{L}_{\theta, \xi}^{GVSL}, \eta)$ ，其中 *optimizer* 是训练中所使用的优化器， $\eta$  是学习率。

**GVSL背后的直觉解释** GVSL中配准的学习过程、相似性损失  $\mathcal{L}^{NCC}$  和用于拓扑保持的平滑损失  $\mathcal{L}^{smooth}$  共同组成的模型学习过程是一种带有拓扑不变先验的隐式度量[259]。如图6-5 a) 所示，相似性度量的依据决定了图像之间的相似度大小。由于外观上与语义无关的变换，仅使用图像外观的相似性将大大限制度量的可靠性。而配准约束度量相似性的过程在拓扑不变条件下进行，避免了由于外观造成的误度量（如图6-2所示）。我们可以将其描述为在  $\{x_A, x_B\}$  的流行平面上进行的相似性度量：

$$\min_{\theta, \xi} \mathcal{L}(G_\xi(\mathcal{N}_\theta(x_A), \mathcal{N}_\theta(x_B)); \{x_A, x_B\}) \quad (6.4)$$

如图 6-5 b) 和式 6.4 所示，它隐式地将图像  $x_A, x_B$  中语义结构内在拓扑流形嵌入到相似性度量过程中 ( $\mathcal{L}^{NCC}$ )，从而获得更加可靠的相似性评价。 $x_A^{RA}$ ,  $x_B^{RA}$  和  $x_A^{Myo}$  分别为图像  $x_A$  和  $x_B$  上的右心房区域和图像  $x_A$  上的左心肌区域。由于图像  $x_B$  的右心房和图像  $x_A$  的左心肌区域在外观上相似，和图像  $x_A$  右心房外观上不同，在特征空间中，两个不同语义的特征将更加接近而相同语义的特征将更远。因此某些直接的特征距离度量，如常用的欧氏距离和马氏距离，将非常容易发生错误，最终导致错误的表征学习。而我们的 GVSL 使用配准模块  $G_\xi$  将特征空间中的特征 ( $\mathcal{N}_\theta(x_A^{RA})$ 、 $\mathcal{N}_\theta(x_B^{RA})$ 、 $\mathcal{N}_\theta(x_A^{Myo})$ ) 进一步映射到图像空间中的拓扑流形上。因此，由于有效的拓扑不变先验，相同语义区域的特征之间的距离将变得更近，从而通过梯度来学习有效的图像相似性。

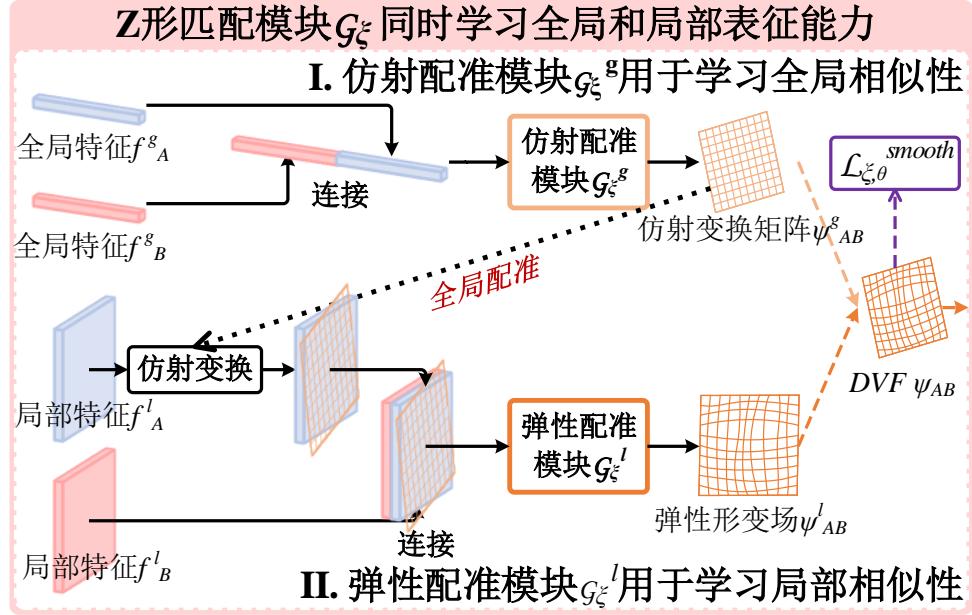


图 6-6: Z形匹配模块的细节。该方法同时学习仿射和弹性形变配准，以获得强大的全局和局部表征能力。

在整个训练过程中，网络 $\mathcal{N}_\theta$ 中的特征表征学习和配准模块 $G_\xi$ 中的语义区域对应学习是一个博奕关系<sup>[260]</sup>。配准模块 $G_\xi$ 学习从表征的特征 $f_A, f_B$ 中预测图像空间中语义区域的对应关系。而网络 $\mathcal{N}_\theta$ 学习向配准模块 $G_\xi$ 提供语义区域的特征，来进行区域对应的学习。而为了获得更精确的对应关系，配准模块也将通过梯度驱动网络输出更具有代表性的语义区域的特征。在这种两方的相互博奕下，网络 $\mathcal{N}_\theta$ 将为配准模块提供更具代表性的特征，以获得更好的对应关系的预测能力，而配准模块 $G_\xi$ 也将用有更强大的能力来学习图像之间的相似性，从而反向促进网络表征。

### 6.3.2 Z形匹配模块

如图6-6所示，Z形匹配模块 $G_\xi$ 是一种新型的代理任务学习模块，它协同学习仿射和弹性形变配准，同时学习全局和局部的表征能力。它有两个子模块，包括用于学习全局相似性的仿射配准模块 $G_\xi^g$ 和用于学习局部相似性的弹性配准模块 $G_\xi^l$ 。

**仿射配准模块用于学习全局视觉相似性** 这部分首先将两张图像的编码器部分输出的全局特征（ $f_A^g, f_B^g$ ）连接起来，然后输入仿射配准模块 $G_\xi^g$ 中以预测仿射变换矩阵数值。在我们的研究中，这些数值包括3个与旋转相关的数值（ $\theta_x, \theta_y, \theta_z$ ），3个与平移相关的数值（ $t_x, t_y, t_z$ ），3个与缩放相关的数值（ $s_x, s_y, s_z$ ）以及6个与剪切相关的数值（ $sh_{xy}, sh_{xz}, sh_{yx}, sh_{yz}, sh_{zx}, sh_{zy}$ ）。如公式6.5所示，这些数值一起计算获得仿射变换矩阵 $\psi_{AB}^g \triangleq G_\xi^g(f_A^g, f_B^g)$ ，该矩阵表示在空间坐标系中对 $x_A$ 进行全局的变换，从而与 $x_B$ 在全局上对齐。因此，该仿射配准模块能够促进模型感知全局对应关系，训练编码器提取

相同全局语义区域间一致且具有代表性的特征。

$$\psi_{AB}^g = \underbrace{\begin{bmatrix} \cos \theta_y \cos \theta_z & \sin \theta_x \sin \theta_y \cos \theta_z - \sin \theta_z \cos \theta_x & \sin \theta_y \cos \theta_x \cos \theta_z + \sin \theta_x \cos \theta_z & 0 \\ \cos \theta_y \sin \theta_z & \cos \theta_x \cos \theta_z + \sin \theta_x \sin \theta_y \sin \theta_z & -\sin \theta_x \cos \theta_z + \sin \theta_z \sin \theta_y \cos \theta_x & 0 \\ -\sin \theta_y & \sin \theta_x \cos \theta_y & \cos \theta_x \cos \theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Rotation}} \underbrace{\begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Scaling}} \underbrace{\begin{bmatrix} 1 & sh_{yx} & sh_{zx} & 0 \\ sh_{xy} & 1 & sh_{zy} & 0 \\ sh_{xz} & sh_{yz} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Shearing}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Translation}}, \quad (6.5)$$

**弹性配准模块用于学习局部视觉相似性** 这部分首先利用仿射变换矩阵  $\psi_{AB}^g$  将网络  $\mathcal{N}_\theta$  解码器部分输出的局部特征  $f_A^l$  在全局上与另一张图像的局部特征  $f_B^l$  对齐，接着将其与  $f_B^l$  在特征维度连接起来输入到弹性配准模块  $\mathcal{G}_\xi^l$  中以预测用于局部对齐的弹性形变场  $\psi_{AB}^l \triangleq \mathcal{G}_\xi^l(\psi_{AB}^g(f_A^l), f_B^l)$ 。从而将图像A形变以实现将其体素对齐到图像B。在整个过程中，弹性配准模块为了学习精准的体素级配准，优化器将训练整个网络  $\mathcal{N}_\theta$  对相同局部语义区域提取一致且有代表性的特征。最终，如公式6.6所示，仿射变换矩阵  $\psi_{AB}^g$  和弹性形变场  $\psi_{AB}^l$  通过  $\odot$  操作融合，得到位移向量场  $\psi_{AB} \triangleq \psi_{AB}^l \odot \psi_{AB}^g$  对每个像素进行位移。

$$\begin{aligned} \psi_{AB}(p) &\triangleq \psi_{AB}^l(p) \odot \psi_{AB}^g = \begin{bmatrix} \psi_{AB}(p_x) \\ \psi_{AB}(p_y) \\ \psi_{AB}(p_z) \end{bmatrix} \\ &= \begin{bmatrix} \psi_{AB}^l(\hat{p}_x) \\ \psi_{AB}^l(\hat{p}_y) \\ \psi_{AB}^l(\hat{p}_z) \end{bmatrix} + \begin{bmatrix} \hat{p}_x \\ \hat{p}_y \\ \hat{p}_z \end{bmatrix} - \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix}, \begin{bmatrix} \hat{p}_x \\ \hat{p}_y \\ \hat{p}_z \\ 1 \end{bmatrix} = \psi_{AB}^g \times \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} \end{aligned} \quad (6.6)$$

其中，仿射矩阵  $\psi_{AB}^g$  将图像中体素的位置索引  $p = p_x, p_y, p_z$  乘以仿射变换矩阵，获得变换后的位置索引  $\hat{p} = p_x, p_y, p_z$ 。仿射变换后的位置索引  $\hat{p}$  减去原始位置索引  $p$ ，得到仿射变换的位移向量，并且将该向量进一步添加到弹性形变场  $\psi_{AB}^l$  对应向量所在的位置  $\hat{p}$  处的变形向量中，从而获得该位置处的仿射变换和弹性形变位移向量  $\psi_{AB}(p)$ 。对整个图像空间中的所有位置都执行此操作，即可将仿射矩阵和变形场融合为位移向量场  $\psi_{AB}$ ，即为两个图像之间语义区域的对应关系，利用相似性损失（NCC）和平滑损失优化该对应关系，从而进一步约束网络  $\mathcal{N}_\theta$  学习有效的语义特征表征。

### 6.3.3 基础代理任务

模型对语义区域的基本表征能力对于GVSL很重要，因此我们将自修复<sup>[148]</sup>作为框架中的基本预训练任务，与GVSL一同进行。这是因为在学习配准的过程中，对应关系的学习 $\mathcal{S}_\xi$ 依赖于网络 $\mathcal{N}_\theta$ 对于两个图像的特征提取能力。然而，如果初始的网络 $\mathcal{N}_\theta$ 对于特征的表征能力较弱，那么将无法为配准模块提供学习发现语义区域间对应关系的特征基础，最终导致发掘可靠的优化目标来训练网络对齐相同的语义区域，阻碍了图像间相似性的学习。因此，我们在框架中融合了一个自修复代理任务<sup>[148]</sup>作为基础学习任务来预热网络。

如图6-4 b) 所示，自修复任务从我们的变换集中采样一个变换操作 ( $t' \sim \mathcal{T}$ ) 来随机变换图像的外观 ( $x_A$ ) 以生成变换后的图像 ( $x'_A \triangleq t'(x_A)$ )，然后将其输入到网络 $\mathcal{N}_\theta$ 中，并从解码器中输出图像的局部特征 $f'_A$ 。（为了节省计算资源，在我们的实验中，GVSL共享此操作，即 $\mathcal{N}_\theta(x'_A) = \mathcal{N}_\theta(x_A)$ 。）然后，将特征 $f'_A$ 输入到一个自修复模块 $\mathcal{R}_t$ 中，生成修复后的图像 ( $x'_A$ )，并与原始图像 $x_A$ 一起计算均方误差作为损失函数 $\mathcal{L}_{\theta,t}^{MSE} = |\mathcal{R}_t(\mathcal{N}_\theta(t'(x_A))) - x_A|^2$ <sup>[148]</sup>，以学习从上下文中修复语义区域原本的特征。因此，网络将在训练过程中学习基本的语义表征来进行预热，即

$$\theta, \iota \leftarrow \text{optimizer}(\theta, \iota, \nabla_{\theta,\iota} \mathcal{L}_{\theta,t}^{MSE}, \eta), \quad (6.7)$$

从而避免仅学习GVSL时产生无效优化的风险。

## 6.4 实验设置与结果分析

### 6.4.1 实验方案

**数据集** 本实验使用了四个数据集，包括一个私有的预训练数据集和三个公开的下游任务数据集。

**预训练数据集：** 我们从东部战区总医院收集了302位患者的无标注心脏CTA图像作为自监督预训练数据集。这些图像通过SOMATOM Definition Flash CT扫描仪上采集获得，在扫描过程中病人被注射造影剂以获得清晰的心脏内部结构。图像的x轴和y轴分辨率在0.25至0.57mm/体素之间，大小为512个体素，z轴层厚在0.75至3mm/体素之间，大小在128至994个体素之间。在于处理过程种，我们首先将这些图像的分辨率重新采样为 $1mm \times 1mm \times 1mm$ ，以获得统一的分辨率，然后将它们的灰度值阈值设置为[0, 2048]后归一化为[0, 1]，以获得统一的强度范围。

**下游任务数据集：** 我们使用三个公共数据集（MM-WHS<sup>[231]</sup>、ASOCA<sup>[261]</sup>、CANDI<sup>[235]</sup>）来验证我们框架的优越性。根据数据种类，我们将其分为场景内迁移和场景间迁移两种不同的研究类型。场景内迁移评估旨在利用心脏CTA图像上的心脏腔室分割（MM-WHS）<sup>[231]</sup>和心脏CTA图像上的冠状动脉分割（ASOCA）<sup>[261]</sup>来评估在与预训练

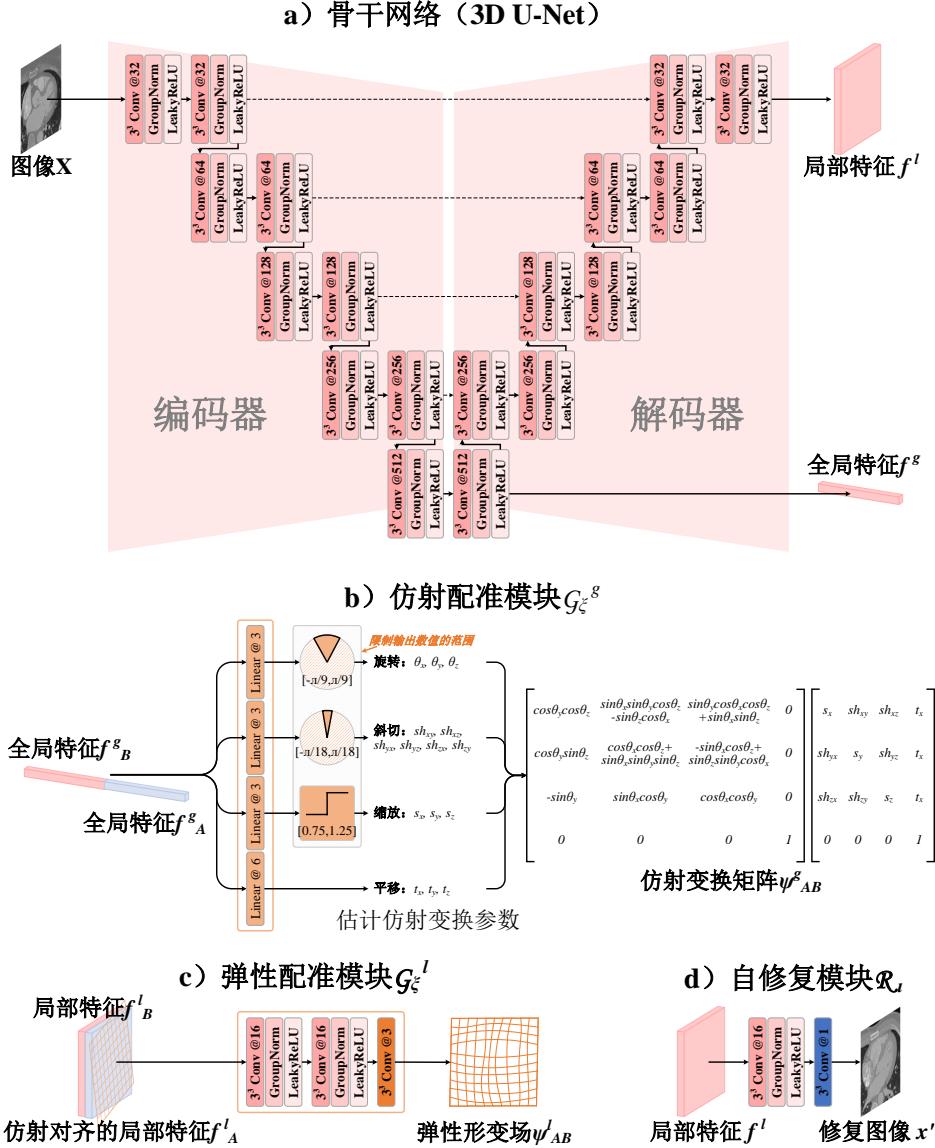


图 6-7: GVSL框架的实施细节。

数据集相同/相近图像场景上的迁移能力。场景间迁移评估旨在利用脑MRI图像上的脑组织分割 (CANDI)<sup>[235]</sup>来评估在与源数据集不同图像场景上的迁移能力。具体的：

a) 心脏CTA图像上的心脏腔室分割 (MM-WHS): 该任务旨在利用MM-WHS 2017数据集<sup>[231]</sup>中的CTA图像，实现七个心脏结构分割。该数据集包含20个被标注的训练图像和40个无标注的测试图像。我们随机选择15个训练图像作为训练集，5个训练图像作为验证集，40个无标签图像作为测试集，并在数据集官方提供的软件上测试结果。在预处理过程中，我们首先在原图上裁剪出心脏感兴趣区域 (ROIs) 以减小显存的占用，然后将这些图像的分辨率重新采样为 $1mm \times 1mm \times 1mm$ ，以获得统一的分辨率。接着，我们将这些图像的灰度值阈值化为[0, 2048]后归一化到[0, 1]，以获得统一的强度范围。此任务评估了模型在针对大结构的密集预测任务中的场景内迁移能力。

b) 心脏CTA图像上的冠状动脉分割 (ASOCA): 该任务旨在利用ASOCA数据集<sup>[261]</sup>中的CTA图像，实现细小的冠状动脉分割，该数据集包含40张被标注的图像。我们随机选择15张作为训练集，5张作为验证集，20张作为测试集。与MM-WHS任务类似，我们在原图上裁剪心脏ROIs，然后将其分辨率重新采样为 $1mm \times 1mm \times 1mm$ ，将灰度值阈值化为[0, 2048]，然后归一化为[0, 1]。该任务评估了模型在对细小结构进行密集预测的任务上的场景内迁移能力。

c) 脑MRI图像上的脑组织分割 (CANDI): 该任务旨在利用CANDI数据集<sup>[235]</sup>中的T1加权脑MRI图像，实现28个脑组织区域分割，该数据集包含103张被标注的图像。我们随机选择40张作为训练集，20张作为验证集，43张作为测试集。与一些在此数据集上的研究工作<sup>[215;137]</sup>相同，我们先在整个脑区域中心向外裁剪 $160 \times 160 \times 128$ 大小的区域，从而去除无关区域使图像中包含整个脑部。将图像中小于0的灰度值阈值化为0，然后通过最小-最大归一化将所有图像归一化到[0, 1]来统一图像强度。该任务评估了模型在对多个（28个）结构进行密集预测的任务上的场景间迁移能力。

**对比设置** 在这项研究中，我们对比了八个相关的工作来验证我们框架的前沿性与优越性，其中包括基于生成的方法<sup>[177;262;148;150]</sup>，基于对比的方法<sup>[152;151;153]</sup>，以及基于聚类的方法<sup>[243]</sup>从而全面分析和展示我们方法在相关研究中的优越性。在所有的实验和对比模型中，我们统一使用3D U-Net<sup>[90]</sup>作为所有方法的骨干网络进行公平的比较。我们同时进行了线性评估和微调评估两种评估方式来验证预训练方法的表征能力和迁移能力，具体的：

a) 线性评估：我们利用线性评估来评价所提取特征的可分类性，进而展示预训练网络的表征能力。我们使用整个预训练网络 $\mathcal{N}_\theta$ 作为下游任务的特征提取器，并固定其参数，使其不被训练。然后，提取图像中的局部特征 $f^l$ ，进一步输入卷积层和一个Softmax激活层，输出分割预测结果。整个模型仅优化最后一层作为预测器的卷积层以判断模型的密集表征能力。

b) 微调评估：我们利用微调评估来评价模型参数的迁移学习能力，从而展示预训练网络作为下游任务初始化参数所带来的性能提升。在这项实验中，我们遵循Models Genesis<sup>[148]</sup>的微调方法。我们将整个3D U-Net网络 $\mathcal{N}_\theta$ 连接到一个卷积层和一个Softmax激活函数，并进行端到端训练，利用反向传播优化网络中的所有参数。

**实施细节和评价指标** 为了进行公正的比较，所有网络都采用Adam优化器<sup>[183]</sup>来进行训练，学习率设置为 $10^{-4}$ 。所有的对比试验都使用Pytorch<sup>[213]</sup>实践，并在一张显存为24GB的NVIDIA Ge-Force RTX 3090 GPU上进行训练。在预训练任务训练过程中，所有方法都执行了 $2 \times 10^5$ 次迭代，在下游任务训练过程中，所有方法将进行 $4 \times 10^4$ 训练次迭代，并且每执行200次迭代，模型都将在验证集上测试性能并保存最佳模型。我们使用平均 Dice系数 (DSC) 来评估分割任务性能<sup>[184]</sup>。

如图6-7所示，在实施过程种，我们的GVSL框架采用3D U-Net作为我们的骨干网

络，瓶颈输出全局特征 $f^g$ ，网络最后一层输出局部特征 $f^l$ 。仿射配准模块利用四个全连接层来分别估计旋转、剪切、缩放和平移的参数，这些参数最终被用于生成仿射变换矩阵 $\psi_{AB}^g$ 。弹性配准模块使用两个卷积-组归一化-LeakyReLU层组和模块和一个卷积层来从局部特征中估计弹性形变场 $\phi_{AB}^l$ 。自修复模块使用一个卷积-组归一化-LeakyReLU层组和模块和一个卷积层来修复图像。源代码发布于<https://github.com/YutingHe-list/GVSL>。

#### 6.4.2 对比分析

表 6.1: 线性评估 (a) 和微调评估 (b) 展示了GVSL强大的表征和迁移能力。表格中加粗的数值代表是该列中的最高值。

Pre-training	a) 线性评估展现出GVSL强大的表征能力			b) 微调评估展现出GVSL强大的迁移能力		
	MM-WHS <sub>DSC%</sub>	ASOCA <sub>DSC%</sub>	CANDI <sub>DSC%</sub>	MM-WHS <sub>DSC%</sub>	ASOCA <sub>DSC%</sub>	CANDI <sub>DSC%</sub>
			场景内迁移			场景间迁移
随机初始化	21.9	10.0	56.4	87.8	80.4	89.7
去噪 <sup>[177]</sup>	31.4 <sub>(+9.5)</sub>	9.3 <sub>(-0.7)</sub>	28.3 <sub>(-28.1)</sub>	90.3 <sub>(+2.5)</sub>	80.5 <sub>(+0.1)</sub>	89.7
抠图 <sup>[262]</sup>	32.3 <sub>(+10.4)</sub>	5.9 <sub>(-4.1)</sub>	25.0 <sub>(-31.4)</sub>	90.4 <sub>(+2.6)</sub>	80.3 <sub>(-0.1)</sub>	89.9 <sub>(+0.2)</sub>
Models Genesis <sup>[148]</sup>	47.4 <sub>(+25.5)</sub>	22.5 <sub>(+12.5)</sub>	44.9 <sub>(-11.5)</sub>	90.3 <sub>(+2.5)</sub>	79.9 <sub>(-0.5)</sub>	89.4 <sub>(-0.3)</sub>
旋转 <sup>[150]</sup>	56.1 <sub>(+34.2)</sub>	21.9 <sub>(+11.9)</sub>	54.1 <sub>(-2.3)</sub>	90.6 <sub>(+2.8)</sub>	81.1 <sub>(+0.7)</sub>	89.6 <sub>(-0.1)</sub>
深度聚类 <sup>[243]</sup>	55.9 <sub>(+34.0)</sub>	4.4 <sub>(-5.6)</sub>	67.5 <sub>(+11.1)</sub>	85.4 <sub>(-2.4)</sub>	80.5 <sub>(+0.1)</sub>	89.1 <sub>(-0.6)</sub>
SimSiam <sup>[153]</sup>	56.5 <sub>(+34.6)</sub>	9.7 <sub>(-0.3)</sub>	66.2 <sub>(+9.8)</sub>	87.5 <sub>(-0.3)</sub>	80.1 <sub>(-0.3)</sub>	89.8 <sub>(+0.1)</sub>
BYOL <sup>[151]</sup>	46.9 <sub>(+25.0)</sub>	8.6 <sub>(-1.4)</sub>	52.7 <sub>(-3.7)</sub>	88.6 <sub>(+0.8)</sub>	80.7 <sub>(+0.3)</sub>	89.5 <sub>(-0.2)</sub>
SimCLR <sup>[152]</sup>	48.7 <sub>(+26.8)</sub>	15.5 <sub>(+5.5)</sub>	58.7 <sub>(+2.3)</sub>	86.9 <sub>(-0.9)</sub>	79.9 <sub>(-0.5)</sub>	89.3 <sub>(-0.4)</sub>
去除“Z形匹配模块”	49.1 <sub>(+27.2)</sub>	21.1 <sub>(+11.1)</sub>	45.1 <sub>(-11.3)</sub>	88.3 <sub>(+0.5)</sub>	81.2 <sub>(+0.8)</sub>	89.7
去除“自修复”	45.3 <sub>(+23.4)</sub>	0.0 <sub>(-10.0)</sub>	48.5 <sub>(-7.9)</sub>	87.0 <sub>(-0.8)</sub>	79.5 <sub>(-0.9)</sub>	89.0 <sub>(-0.7)</sub>
去除“仿射配准模块”	57.7 <sub>(+35.8)</sub>	17.9 <sub>(+7.9)</sub>	53.4 <sub>(-3.0)</sub>	89.4 <sub>(+1.6)</sub>	82.3 <sub>(+1.9)</sub>	89.8 <sub>(+0.1)</sub>
完整的GVSL	<b>68.4</b> <sub>(+46.5)</sub>	<b>28.7</b> <sub>(+18.7)</sub>	<b>79.9</b> <sub>(+23.5)</sub>	<b>91.2</b> <sub>(+3.4)</sub>	81.3 <sub>(+0.9)</sub>	<b>90.0</b> <sub>(+0.3)</sub>

**同时适用大小结构的场景内迁移能力** 在场景内迁移学习任务中，GVSL强大的性能展现出在医学图像大数据量但少标签场景中的巨大应用潜力。在与预训练数据集的相同/近似的图像场景中，GVSL同时在大结构和小结构分割任务中均取得了最佳性能。**1)** 在分割大型心脏结构的MM-WHS任务中，GVSL在线性评估中取得了最高的DSC (68.4%)，比排名第二高的方法<sup>[153]</sup>高出11.9%。这是因为GVSL学习图像间相似性，能够促进语义区域学习一致的特征表征。特别是对于具有清晰语义区域的大解剖结构，如CTA图像上的心脏内部腔室，配准学习能够带来更加高效的表征学习。**2)** 在分割细小冠脉的ASOCA任务中，大量已有的预训练方法<sup>[262;153;151;152]</sup>都会导致模型忽略这些细小结构的特征，导致模型性能甚至低于随机初始化的模型。特别是为全局预测所设计的方法<sup>[153;151;152]</sup>，这些方法更关注全局大感受野下特征的学习，更容易丢失对细小结

构的表征。我们GVSL中的弹性形变配准学习和自修复学习训练网络学习精细的细小视觉语义特征表征，取得了线性评估中的最高DSC（28.7%）。有趣的是，在我们的Z形匹配模块中移除仿射配准模块后，GVSL进一步在ASOCA任务中取得了最佳的微调评估性能（82.3%），进一步证明了为细小物体学习密集表征的重要性。

**强大的场景间迁移能力** 在场景间迁移学习任务中，GVSL进一步展现出强大的性能，说明了我们的作为预训练参数初始化网络的有效性。CANDI任务使用了与预训练图像不同模态和图像上下文（身体范围）的脑MRI图像，构成了场景间迁移的评估。在线性评估中，大量的对比方法<sup>[148;151;150;151]</sup>由于预训练数据集（心脏CT）和下游任务数据集（脑MRI）之间巨大的差异而使得性能甚至弱于随机初始化。而我们的GVSL在线性评估中取得了最高的DSC（79.9%），与随机初始化相比提高了23.5%。这是因为GVSL学习图像间相似性，使得预训练网络对语义区域拥有了更加一致的表征，使得特征拥有更好的聚类特性，在该特性下模型的表征将更加容易整体地迁移至目标下游任务。但是值得注意的是，在微调评估中，所有方法的性能都与随机初始化性能相似甚至更差<sup>[152;151;243;150;148]</sup>。这是因为脑MRI和心脏CTA图像之间有着巨大的分布差异（不同的模态和身体范围），使得预训练模型的表征能力与下游任务之间相去甚远，无法获得有贡献的迁移。但是即便如此，GVSL仍然取得了0.3%的提升。

#### 6.4.3 消融实验和模型分析

**消融实验** 如表6.1所示，我们将GVSL与仅学习基础预训练任务（自修复任务）、仅用使用Z形匹配模块学习配准和基础预训练任务+仅实现弹性配准模块学习配准这三种方法进行比较。我们可以得到三个观察结论：1) 仅学习Z形匹配模块时，网路较弱的初始表征能力使得模型的优化非常低效，导致较差的表征能力。特别是在ASOCA任务的线性评估中，由于简单的配准学习导致的最终获得表征的低效性，在这类细小结构分割任务上无法豁达有效的性能。2) 当仅学习基础“任务时，模型整体性能较差，尤其是ASOCA任务上，无论是线性评估还是微调评估，该模型性能均比随机初始化更差。3) 当删除Z形匹配模块中的仿射模块时，由于缺乏全局表征学习，无论是大的心脏结构还是小的冠状动脉结构都降低了10.7%和10.84%。由此可见，我们的每个模块在整个框架中都具有其必要性和有效性。

**对学习效率的提升作用分析** 如图6-8所示，我们分析了随机初始化、GVSL预训练和Model Genesis预训练模型在ASOCA和CANDI任务中的学习性能。在线性评估中，由于GVSL对局部图像间相似性的有效学习，相比于随机初始化和Model Genesis，我们的方法取得了超过20%的DSC提升。在微调评估中，GVSL以其优秀的初始表征能力大大提高了训练的收敛速度，获得了超过30%的速度提升，表现出其强大的收敛能力和节省计算资源的巨大潜力。尽管在CANDI任务的微调评估中，随机初始化的收敛速度似乎

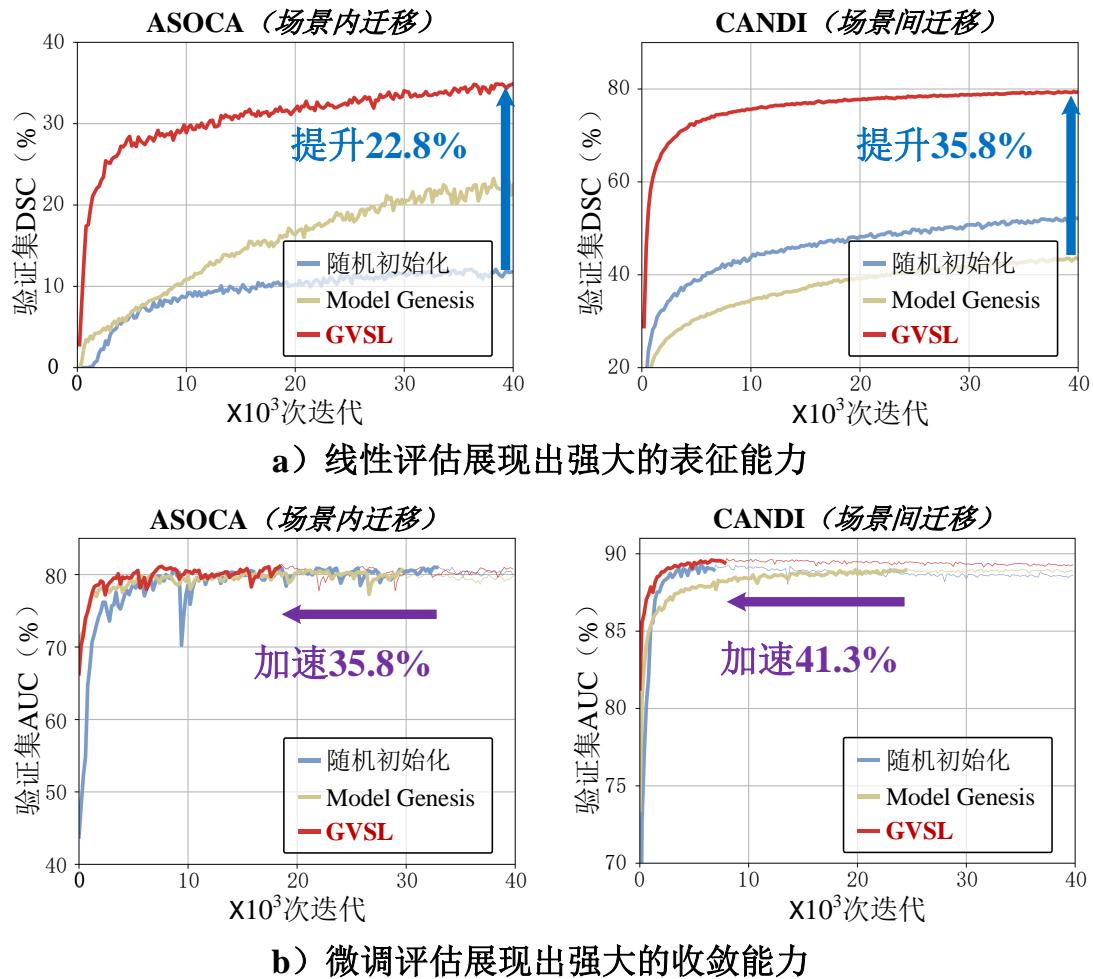


图 6-8: GVSL在线性评估中展现出了强大的表征能力，在微调评估中展现出了更快的收敛能力。在b) 中，粗线的部分表示验证集上未达到最佳模型的训练过程。

更快，但它很快陷入了过度拟合，最终性能弱于我们的方法。

**基础任务对于GVSL的必要性分析** 自修复学习为网络提供了基础的视觉特征表征，进而驱动配准学习过程中对图像间相似性的学习，具有必要性。我们验证了当去除自修复学习后，模型对于配准学习的收敛性。如图6-9所示，当仅学习配准代理任务时，网络由于较弱的初始表征能力，使得对于配准的学习效率非常低下，因此用于学习配准的相似性损失（NCC损失， $\mathcal{L}^{NCC}$ ）无法收敛，导致Z形匹配模块无法学习语义区域的对应关系。当添加我们基础的自修复代理任务时，由于网络从自修复学习中获取了对语义区域的基本表征能力，配准学习被有效驱动，NCC损失开始下降，从而我们的GVSL在整个学习过程中发挥作用，模型能够学习到更有效的语义特征聚类效应。如表6.1所示，加入自修复学习和仅进行Z形匹配模块的配准学习相比，加入自修复学习后性能获得了显著的提升。

**GVSL预训练模型的特征聚类效应分析** 如图6-10所示，我们对MM-WHS任务的图像

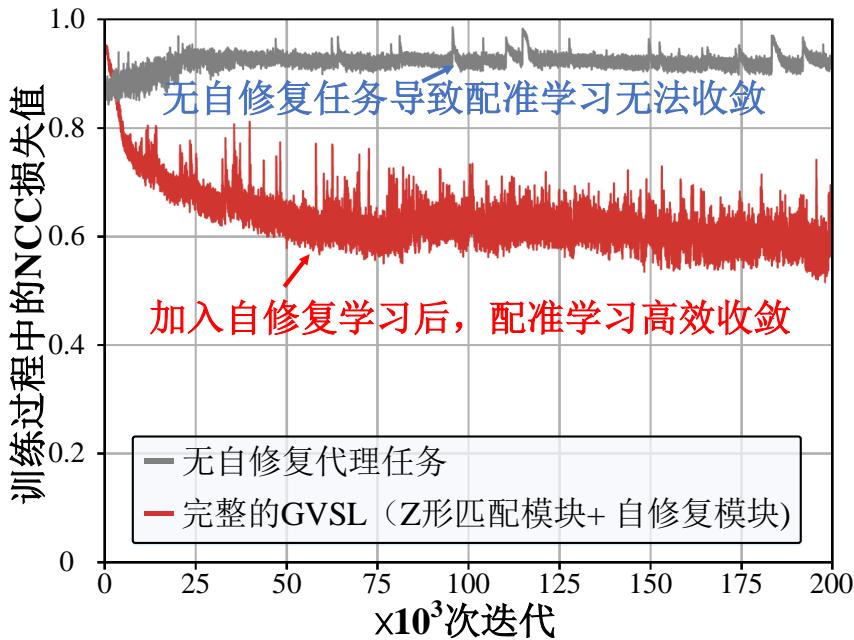


图 6-9: GVSL中基础任务的必要性分析图。a) 当仅训练GM任务时, NCC损失 $\mathcal{L}^{NCC}$ 无法收敛, 无法学习语义区域的对应关系。b) 当添加基础任务(自修复)时, 从语义区域的基本表示进行预热, 促进了GM的有效优化。

使用预训练模型提取局部特征 $f^l$ , 利用t-SNE压缩特征后绘制散点图以展示了模型对特征的聚类效应, 结果显示我们GVSL拥有更好的聚类能力, 表现出其强大的表征性能。随机初始化方法无法提取具有代表性的特征, 无法区分潜在的语义区域, 因此它将不同的语义特征混合, 表征能力较差。当只学习自修复任务时, 模型能够学习到特征的基本表征能力, 但整体模型由于无法学习图像间相似性, 表征能力仍然受限。因此, 其提取的特征 $f^l$ 仍然存在较严重的混合。当仅使用我们的Z形匹配模块学习配准任务时, 所提取的特征表现出明显更强的聚类效应, 说明了学习图像间相似性的重要性。然而, 某些语义特征仍然被混合到一起, 无法被分开。当我们融合学习配准和学习自修复一起来训练GVSL时, 来自基础任务的基本表征能力促进了配准的学习, 最终获得了良

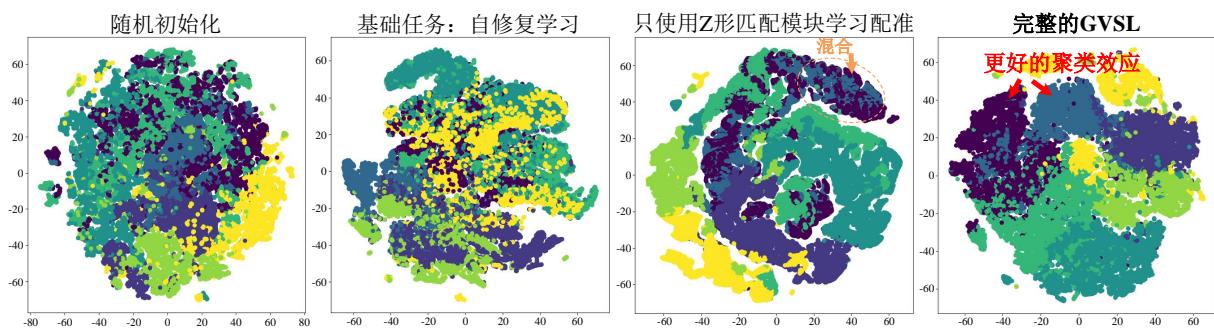


图 6-10: GVSL预训练模型的特征聚类效应散点图。通过t-SNE<sup>[263]</sup>将模型提取的特征压缩成2维向量后绘制散点图, 对预训练网络在MM-WHS任务中提取的七个语义区域(AO、RA、RV、LV、PA、LA和Myo)的局部特征 $f^l$ 进行的聚类可视化, 我们发现GVSL表现出了更好的特征聚类效应。

好的特征聚类效应，不同语义的特征具有较好的区分度。尽管黄色的点图中被划分为了三个部分，但是每个部分都是聚集的，这也说明了该区域内部语义表征的一致性。

## 6.5 本章小结

本章的研究构建了一种三维医学图像自监督预训练中的图像间相似性学习算法，几何视觉相似性学习（GVSL），为下游任务提供了具有强大表征能力的预训练初始化参数。该算法学习医学图像间相同语义区域的一致表征能力，在下游医学图像分割学习中表现出了显著的促进作用。虽然本研究的基础假说——医学图像间拓扑一致性，为本方法提供了强有力的动机，但是这也限制了该方法应用到更多不满足拓扑一致性的图像上。大量的实验给予了本研究很大的信心，学习图像间一致性的几何相似性学习将进一步促进没有标签情况下的医学图像学习研究，并且所获得的预训练参数为下游分割任务提供更有效的通用泛化能力。本章的研究成果被发表在CVPR 2023上，详见作者简介中文献[A7]，方法相关源代码已开源在<https://github.com/YutingHe-list/GVSL>。



## 第七章 总结与展望

### 7.1 全文总结

数据高效学习是应对深度学习在医学图像分割任务中数据效率问题的必研之路，提高医学图像分割学习模型在图像和标签上的学习效率，能够大规模构建可用数据库、大规模训练和大规模应用。在具有临床价值却需要大量标注和处理成本的医学图像分割中，数据高效学习的研究设计为其提供了可靠的算法基础，使这类方法迈向通用化大模型创造了条件。本文针对医学图像分割中的数据高效学习算法，研究了基于半监督学习的精细肾动脉分割算法（[第三章](#)）、基于元学习的肾脏多结构分割算法（[第四章](#)）、基于分割配准互补性的少标签通用多结构医学图像分割算法（[第五章](#)）、基于自监督学习的三维医学图像表征预训练算法（[第六章](#)），为解决医学图像分割学习中的效率问题提供了一条有效的研究思路、多个切实有效的分割新算法。

全文贡献总结如下：

- 1) 分析了医学图像分割学习中的效率问题，并介绍了涉及该问题的深度学习的四大元素（数据 $\mathbb{D}_{train}$ 、网络结构 $\mathcal{N}$ 、网络参数 $\theta$ 、学习目标 $\mathcal{L}$ ），进一步概览了分割学习以及数据高效学习中的关键技术和相关挑战，为高效医学图像分割学习研究领域指明了方向。（[第一章](#)和[第二章](#)）
- 2) 针对标签数量有限情况下的精细肾动脉分割任务，我们设计了一种基于半监督学习、密集偏置神经网络结构和难区域自适应损失的精细肾动脉分割算法。将更多的无标签图像的知识嵌入到学习过程中，并构建对肾动脉不同尺度适应能力强的有效学习神经网络，最后通过设计一种对肾动脉特征挖掘难度大问题更加有效的损失函数实现有效的肾动脉分割学习。（[第三章](#)）
- 3) 针对图像间分布变化大，图像内目标结构特征不显著的肾癌多结构分割任务，我们构建了一种自适应不同结构分布优势和图像分布优势的元学习算法。将自适应搜索对任务分布最有效的图像分布窗口，并设计了一种图像级元学习方法动态地融合多个窗口的特征，从而实现高效准确的肾癌多结构分割学习。（[第四章](#)）
- 4) 为了进一步降低标签需求，构建更加通用的医学图像分割算法框架，我们研究了医学图像分割任务与配准任务间的互补性，构建了一种可靠的基于配准的伪标签数据生成和鲁棒的分割学习算法。该算法面向更加通用的任务场景同时进一步减少了医学图像分割学习的标签数量需求，仅在1个标签和5个标签的两种设置下，该框架在心脏结构分割和脑组织分割在两个不同医学图像分割任务上均取得了非常显著的效果。（[第五章](#)）
- 5) 为了进一步研究无标签情况下的表征学习，提高任意潜在的医学图像分割任

务的学习效率，我们设计了一种基于图像间几何视觉相似性的医学图像自监督表征预训练算法。该算法建模由于人体结构一致性所带来的医学图像间的拓扑结构一致性的性质，并利用该性质驱动自监督学习中图像间语义一致的特征聚集，实现了高效的图像自监督表征预训练，从而服务于更广泛的下游医学图像分割学习。我们在3项不同的下游分割任务中进行了线性评估和微调评估，实验显示预训练模型在下游分割任务上展现出强大适应能力，能够有效提升初始表征能力，加速收敛，提升最终分割性能。（第六章）

综上所述，本文提出了一系列解决医学图像分割学习中的效率问题的方法，从一个临床相关的具体问题——肾癌影像分割出发，不断提高研究问题的通用性和算法的学习效率。部分本文中所提出的方法的相关源代码已开源，这些代码也将有望帮助医学图像分割中的数据高效学习领域进一步发展。

## 7.2 研究展望：小与大

数据高效学习似乎是一项着眼于“小”的研究，希望通过优化深度神经网络的各个环节，以降低获得医学图像分割模型所需的数据成本。而现如今，人工通用智能（AGI）在多个领域展现出了非凡的性能和应用潜力，越来越受到人们的关注。这是一个着眼于“大”的未来趋势，利用海量的数据训练大模型以获得具有通用表现的智能。也许读到这里，人们心中将产生一个疑惑：“数据高效学习研究似乎与如今的大模型与人工通用智能的发展趋势背道而驰，是否已经落后于潮流？”答案当然是否定的。数据高效学习恰是未来人工通用智能的基石，通过降低获得单一任务可靠性能所需的单位成本，才能够降低大模型时代的大规模通用任务的整体成本。解决了医学图像分割学习的数据效率问题的前提下，为构建大规模医学图像学习奠定基础。由此可见，“小”的数据高效学习是“大”的人工通用智能的保障，在未来，有着非常巨大的研究意义。基于当前数据高效学习、医学图像分割模型的发展现状，本文列举一些潜在的未来研究方向：

1) 医学图像的大容量神经网络结构。由于医学图像数据尺寸大、模态多、疾病类型多样、特征细小多变等特点，面向大模型如何构建满足医学图像特点的大容量神经网络结构，既能够高效地表征医学图像中各种对象语义特征，又能够容纳大规模特征表征的通用网络结构，是未来研究的一大重点。

2) 大规模通用医学图像预训练算法。以往的医学图像预训练算法（包括本文第六所研方法）往往目标是让模型能够有效地进行微调适应下游任务，而非直接无训练地将其应用下游任务。而近期的大模型研究中，已有许多别的领域的办法将大型预训练模型直接作为特征提取器应用于下游任务中。因此如何进一步提升预训练模型的表征能力，在大规模医学图像训练中获得有着较强泛化性能的预训练模型，也将是未来的一大研究方向。

3) 适用于医学图像的提示学习。提示学习能够通过对模型提供少量的提示使其快

速适应到目标任务上，是新一代被大量应用于大模型应用的高效迁移学习方法。在未来医学图像分割和配准的大模型应用中，构建能够应用于这些任务的提示学习有着非常大的意义。

4) 医学图像大规模联邦学习算法。由于医学图像的昂贵性，难以一次性收集到大规模数据进行大模型训练，同时由于人体数据的隐私性，原则上医学图像难以出医学中心，更进一步加剧大规模数据收集的难度。因此，需要构建一种大规模联邦学习方法，通过长期、多中心、分布式的训练大型医学图像模型，有着巨大的意义和挑战。

5) 面向大规模数据的学习驱动的标注算法。除了数据，当海量数据到来时，数据的标注必然会成为成本危机中的重要一环，如何面向通用大模型构建高效的标注算法，通过数据驱动的方式减低标注成本，是面向大模型时代高效学习的重要课题之一。

6) 在更广泛的医学图像数据模态、人体区域、临床场景进行模型验证，比如腹部、胰腺、头颈高危器官等，进一步验证本文所设计算法的通用性和有效性；同时将所研医学图像分割数据高效学习方法应用化、系统化，让所研方法真正服务于临床。



## 致谢

寒来暑往，春去秋来，五年时光匆匆而过，终于走到了博士生涯终点，落笔之时与收笔之刻，心境变化万千。东南大学美丽的校园风景与浓郁的学术氛围浸透了我的硕博时光，也一点一滴影响着我的梦想和追求。硕博连读的5年多时光里，艰难困苦，郁郁不解，又或是柳岸花明，豁然开朗，都成为了我人生中最重要的也是最美好的回忆时光。一路走来，感谢恩师、亲友、同窗相佐。

首先，我要感谢我的导师杨冠羽教授，从论文的选题、算法设计、结果分析到论文的撰写，感谢您始终给予我细心的指导和不懈的支持！您广博的专业知识，严谨的治学精神，精益求精的工作作风，深深感染并激励着我。师者传道授业，在此，向您表示最诚挚的敬意！感谢李硕教授对我课题设计和实验开展工作上的悉心帮助和指导，在加拿大西安大略大学联合培养期间，感谢您对学生的研究和发展上倾注的大量心血和精力，您治学严谨求实，具有敏锐的学术洞察力和丰富的实践经验并且善于引导和启发，是我成长的榜样，学生将永远铭记您的恩情！感谢王柏予教授在我加拿大联培期间对我机器学习理论方法的指导以及对我职业观的培养，感谢您悉心的照顾和帮助。

学生愚钝，虽对于学术仍不得要领，但恩师们的教诲将永远牢记心中。我曾沾沾自喜，也曾迷茫无措，感谢恩师们的时常的敲打与警醒，让我从迷茫中找到“学海”的方向。是老师们在我浮躁时告诉我，要上下求索，从迷蒙中寻找真理的辉光。“学海无涯”这四个字，我将深深地烙印在心中。走出校园，我将依旧带着虔诚的心态继续我的科研工作，在学术，在探索真理的道路上，继续不断前进。

感谢陈阳、伍家松、孔佑勇、杨淳汎等所有老师在论文进行过程中给予的关心、帮助和支持！

感谢葛荣骏、戚晓明、戚耀磊、张媛、王通、陈晓飞、石嘉诚、姜子玥、许立崭、董章福、许晗等所有同学，在我攻读博士期间给予我在研究开展上的帮助和生活中的关怀。

感谢在加拿大联培期间有幸结识的徐鹏程、曾秋皓、濮睿智等同学，在异国求学中给予的关心和帮助。

感谢魏敏娜、邢婉秋、张一凡等老师在我读博期间在评奖、评优和各类考核中给予的关心和帮助。

感谢我的家人和朋友们在读博期间对我的无条件的支持、信任和爱护，你们是我动力的源泉和精神的支柱！

最后，衷心地感谢孔金迪仙女对我的帮助！



## 参考文献

- [1] Panayides, Andreas S, Amini, Amir, Filipovic, Nenad D, et al. AI in medical imaging informatics: current challenges and future directions[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(7):1837–1857.
- [2] Brody, Herb. Medical imaging[J]. Nature, 2013, 502(7473):S81–S81.
- [3] Tajbakhsh, Nima, Jeyaseelan, Laura, Li, Qian, et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation[J]. Medical Image Analysis, 2020, 63:101693.
- [4] Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015. 234–241.
- [5] LeCun, Yann, Bengio, Yoshua, and Hinton, Geoffrey. Deep learning[J]. Nature, 2015, 521(7553):436–444.
- [6] He, Yuting, Yang, Guanyu, Yang, Jian, et al. Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation[J]. Medical Image Analysis, 2020, 63:101722.
- [7] He, Yuting, Yang, Guanyu, Yang, Jian, et al. Meta grayscale adaptive network for 3D integrated renal structures segmentation[J]. Medical Image Analysis, 2021, 71:102055.
- [8] Qi, Yaolei, Xu, Han, He, Yuting, et al. Examinee-Examiner Network: Weakly Supervised Accurate Coronary Lumen Segmentation Using Centerline Constraint[J]. IEEE Transactions on Image Processing, 2021, 30:9429–9441.
- [9] Xu, Chenchu, Xu, Lei, Gao, Zhifan, et al. Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture[J]. Medical Image Analysis, 2018, 50:82–94.
- [10] Jenkinson, Mark and Smith, Stephen. A global optimisation method for robust affine registration of brain images[J]. Medical Image Analysis, 2001, 5(2):143–156.
- [11] Rajpurkar, Pranav, Chen, Emma, Banerjee, Oishi, et al. AI in health and medicine[J]. Nature Medicine, 2022, 28(1):31–38.

- [12] Hesamian, Mohammad Hesam, Jia, Wenjing, He, Xiangjian, et al. Deep learning techniques for medical image segmentation: achievements and challenges[J]. *Journal of Digital Imaging*, 2019, 32(4):582–596.
- [13] Menghani, Gaurav. Efficient deep learning: A survey on making deep learning models smaller, faster, and better[J]. *ACM Computing Surveys*, 2023, 55(12):1–37.
- [14] Wang, Shanshan, Li, Cheng, Wang, Rongpin, et al. Annotation-efficient deep learning for automatic medical image segmentation[J]. *Nature communications*, 2021, 12(1):5915.
- [15] Cheplygina, Veronika, de Bruijne, Marleen, and Pluim, Josien PW. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis[J]. *Medical image analysis*, 2019, 54:280–296.
- [16] Litjens, Geert, Kooi, Thijs, Bejnordi, Babak Ehteshami, et al. A survey on deep learning in medical image analysis[J]. *Medical Image Analysis*, 2017, 42:60–88.
- [17] Isloor, Sreekaanth S and Marsland, T Anthony. The deadlock problem: An overview.[J]. *Computer*, 1980, 13(9):58–78.
- [18] van Dis, Eva AM, Bollen, Johan, Zuidema, Willem, et al. ChatGPT: five priorities for research[J]. *Nature*, 2023, 614(7947):224–226.
- [19] Willeminck, Martin J, Koszek, Wojciech A, Hardell, Cailin, et al. Preparing medical imaging data for machine learning[J]. *Radiology*, 2020, 295(1):4–15.
- [20] Kovashka, Adriana, Russakovsky, Olga, Fei-Fei, Li, et al. Crowdsourcing in computer vision[J]. *Foundations and Trends® in Computer Graphics and Vision*, 2016, 10(3):177–243.
- [21] Pal, Nikhil R and Pal, Sankar K. A review on image segmentation techniques[J]. *Pattern Recognition*, 1993, 26(9):1277–1294.
- [22] Minaee, Shervin, Boykov, Yuri Y, Porikli, Fatih, et al. Image segmentation using deep learning: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [23] Badue, Claudine, Guidolini, Rânik, Carneiro, Raphael Vivacqua, et al. Self-driving cars: A survey[J]. *Expert Systems with Applications*, 2021, 165:113816.
- [24] Horn, Berthold, Klaus, Berthold, and Horn, Paul. Robot vision[M]. MIT press, 1986.

- [25] Bhargavi, K and Jyothi, S. A survey on threshold based segmentation technique in image processing[J]. International Journal of Innovative Research and Development, 2014, 3(12):234–239.
- [26] Adams, Rolf and Bischof, Leanne. Seeded region growing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6):641–647.
- [27] Muthukrishnan, Ranjan and Radha, Miyilsamy. Edge detection techniques for image segmentation[J]. International Journal of Computer Science & Information Technology, 2011, 3(6):259.
- [28] Long, Jonathan, Shelhamer, Evan, and Darrell, Trevor. Fully convolutional networks for semantic segmentation[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. 3431–3440.
- [29] Lesage, David, Angelini, Elsa D, Bloch, Isabelle, et al. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes[J]. Medical Image Analysis, 2009, 13(6):819–845.
- [30] Yang, Guanyu, Kitslaar, Pieter, Frenay, Michel, et al. Automatic centerline extraction of coronary arteries in coronary computed tomographic angiography[J]. The International Journal of Cardiovascular Imaging, 2012, 28:921–933.
- [31] Martínez-Pérez, M Elena, Hughes, Alun D, Stanton, Alice V, et al. Retinal blood vessel segmentation by means of scale-space analysis and region growing[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 1999. 90–97.
- [32] Zhou, Zongwei, Siddiquee, Md Mahfuzur Rahman, Tajbakhsh, Nima, et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 39(6):1856–1867.
- [33] Chlap, Phillip, Min, Hang, Vandenberg, Nym, et al. A review of medical image data augmentation techniques for deep learning applications[J]. Journal of Medical Imaging and Radiation Oncology, 2021, 65(5):545–563.
- [34] Creswell, Antonia, White, Tom, Dumoulin, Vincent, et al. Generative adversarial networks: An overview[J]. IEEE Signal Processing Magazine, 2018, 35(1):53–65.
- [35] Tran, Ngoc-Trung, Tran, Viet-Hung, Nguyen, Ngoc-Bao, et al. On data augmentation for gan training[J]. IEEE Transactions on Image Processing, 2021, 30:1882–1897.

- [36] Pan, Sinno Jialin and Yang, Qiang. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345–1359.
- [37] Moradi, Reza, Berangi, Reza, and Minaei, Behrouz. A survey of regularization strategies for deep models[J]. Artificial Intelligence Review, 2020, 53:3947–3986.
- [38] He, Yuting, Ge, Rongjun, Wu, Jiasong, et al. Thin semantics enhancement via high-frequency priori rule for thin structures segmentation[C]. In: IEEE International Conference on Data Mining. 2021. 1096–1101.
- [39] Johnson, Justin M and Khoshgoftaar, Taghi M. Survey on deep learning with class imbalance[J]. Journal of Big Data, 2019, 6(1):1–54.
- [40] Ma, Jun, Chen, Jianan, Ng, Matthew, et al. Loss odyssey in medical image segmentation[J]. Medical Image Analysis, 2021, 71:102035.
- [41] Drummond, Chris, Holte, Robert C, et al. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling[C]. In: Workshop on Learning from Imbalanced Datasets II. 2003. 1–8.
- [42] Chai, Lu, Wang, Zidong, Chen, Jianqing, et al. Synthetic augmentation for semantic segmentation of class imbalanced biomedical images: A data pair generative adversarial network approach[J]. Computers in Biology and Medicine, 2022, 150:105985.
- [43] Wang, Jiaqi, Zhang, Wenwei, Zang, Yuhang, et al. Seesaw loss for long-tailed instance segmentation[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. 9695–9704.
- [44] Lin, Tsung-Yi, Goyal, Priya, Girshick, Ross, et al. Focal loss for dense object detection[C]. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2017. 2980–2988.
- [45] Puyol-Antón, Esther, Ruijsink, Bram, Piechnik, Stefan K, et al. Fairness in cardiac MR image analysis: an investigation of bias due to data imbalance in deep learning based segmentation[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2021. 413–423.
- [46] Wang, Hongyi, Lin, Lanfen, Hu, Hongjie, et al. Patch-free 3D medical image segmentation driven by super-resolution technique and self-supervised guidance[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2021. 131–141.

- [47] Kollem, Sreedhar, Reddy, Katta Rama Linga, and Rao, Duggirala Srinivasa. A review of image denoising and segmentation methods based on medical images[J]. International Journal of Machine Learning and Computing, 2019, 9(3):288–295.
- [48] Xu, Chenchu, Zhang, Dong, Chong, Jaron, et al. Synthesis of gadolinium-enhanced liver tumors on nonenhanced liver MR images using pixel-level graph reinforcement learning[J]. Medical Image Analysis, 2021, 69:101976.
- [49] Bengio, Yoshua, Courville, Aaron, and Vincent, Pascal. Representation learning: A review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8):1798–1828.
- [50] Luo, Xiangde, Chen, Jieneng, Song, Tao, et al. Semi-supervised medical image segmentation through dual-task consistency[C]. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2021. 8801–8809.
- [51] Xia, Yingda, Yang, Dong, Yu, Zhiding, et al. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation[J]. Medical Image Analysis, 2020, 65:101766.
- [52] Chen, Shuai, Bortsova, Gerda, García-Uceda Juárez, Antonio, et al. Multi-task attention-based semi-supervised learning for medical image segmentation[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2019. 457–465.
- [53] Budd, Samuel, Robinson, Emma C, and Kainz, Bernhard. A survey on active learning and human-in-the-loop deep learning for medical image analysis[J]. Medical Image Analysis, 2021, 71:102062.
- [54] Luo, Xiangde, Wang, Guotai, Song, Tao, et al. MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning[J]. Medical Image Analysis, 2021, 72:102102.
- [55] Zhang, Man, Zhou, Yong, Zhao, Jiaqi, et al. A survey of semi-and weakly supervised semantic segmentation of images[J]. Artificial Intelligence Review, 2020, 53:4259–4288.
- [56] Chen, Zhang, Tian, Zhiqiang, Zhu, Jihua, et al. C-CAM: Causal CAM for weakly supervised semantic segmentation on medical image[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 11676–11685.

- [57] Karimi, Davood, Dou, Haoran, Warfield, Simon K, et al. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis[J]. *Medical Image Analysis*, 2020, 65:101759.
- [58] Li, Jundong, Cheng, Kewei, Wang, Suhang, et al. Feature selection: A data perspective[J]. *ACM Computing Surveys (CSUR)*, 2017, 50(6):1–45.
- [59] Liu, Junbo, Huang, Yaping, Zou, Qi, et al. Learning visual similarity for inspecting defective railway fasteners[J]. *IEEE Sensors Journal*, 2019, 19(16):6844–6857.
- [60] Viola, Paul and Wells III, William M. Alignment by maximization of mutual information[J]. *International Journal of Computer Vision*, 1997, 24(2):137–154.
- [61] Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron. *Deep learning*[M]. MIT press, 2016.
- [62] Shen, Dinggang, Wu, Guorong, and Suk, Heung-II. Deep learning in medical image analysis[J]. *Annual Review of Biomedical Engineering*, 2017, 19:221.
- [63] Henriques, Joao F, Carreira, Joao, Caseiro, Rui, et al. Beyond hard negative mining: Efficient detector learning via block-circulant decomposition[C]. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2013. 2760–2767.
- [64] Wolterink, Jelmer M, Leiner, Tim, de Vos, Bob D, et al. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks[J]. *Medical Image Analysis*, 2016, 34:123–136.
- [65] Milletari, Fausto, Navab, Nassir, and Ahmadi, Seyed-Ahmad. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]. In: *International Conference on 3D Vision (3DV)*. 2016. 565–571.
- [66] Liu, Tian-Yu. Easyensemble and feature selection for imbalance data sets[C]. In: *International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*. 2009. 517–520.
- [67] Thung, Kim-Han and Wee, Chong-Yaw. A brief review on multi-task learning[J]. *Multimedia Tools and Applications*, 2018, 77:29705–29725.
- [68] Shao, Wei, Wang, Tongxin, Sun, Liang, et al. Multi-task multi-modal learning for joint diagnosis and prognosis of human cancers[J]. *Medical Image Analysis*, 2020, 65:101795.

- [69] Zhou, Yue, Chen, Houjin, Li, Yanfeng, et al. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images[J]. *Medical Image Analysis*, 2021, 70:101918.
- [70] Silla, Carlos N and Freitas, Alex A. A survey of hierarchical classification across different application domains[J]. *Data Mining and Knowledge Discovery*, 2011, 22:31–72.
- [71] Barto, Andrew G and Mahadevan, Sridhar. Recent advances in hierarchical reinforcement learning[J]. *Discrete Event Dynamic Systems*, 2003, 13(1-2):41–77.
- [72] Jiang, Yiding, Gu, Shixiang Shane, Murphy, Kevin P, et al. Language as an abstraction for hierarchical deep reinforcement learning[J]. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, 32.
- [73] Gordon, Allan D. A review of hierarchical classification[J]. *Journal of the Royal Statistical Society: Series A (General)*, 1987, 150(2):119–137.
- [74] Ayyar, Meghna P, Benois-Pineau, Jenny, and Zemmari, Akka. A hierarchical classification system for the detection of covid-19 from chest X-Ray images[C]. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021. 519–528.
- [75] Kowsari, Kamran, Sali, Rasoul, Ehsan, Lubaina, et al. Hmic: Hierarchical medical image classification, a deep learning approach[J]. *Information*, 2020, 11(6):318.
- [76] Hermessi, Haithem, Mourali, Olfa, and Zagrouba, Ezzeddine. Multimodal medical image fusion review: Theoretical background and recent advances[J]. *Signal Processing*, 2021, 183:108036.
- [77] Jin, Zhiwei, Cao, Juan, Guo, Han, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]. In: *Proceedings of the ACM International Conference on Multimedia*. 2017. 795–816.
- [78] Xu, Xiaodong, Li, Wei, Ran, Qiong, et al. Multisource remote sensing data classification based on convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 56(2):937–949.
- [79] Qi, Xiaoming, He, Yuting, Yang, Guanyu, et al. MVSGAN: Spatial-aware multi-view CMR fusion for accurate 3D left ventricular myocardium segmentation[J]. *IEEE Journal of Biomedical and Health Informatics*, 2021, 26(5):2264–2275.
- [80] Goldman, Lee W. Principles of CT and CT technology[J]. *Journal of Nuclear Medicine Technology*, 2007, 35(3):115–128.

- [81] Ge, Rongjun, Shi, Fanqi, Chen, Yang, et al. Improving anisotropy resolution of computed tomography and annotation using 3D super-resolution network[J]. Biomedical Signal Processing and Control, 2023, 82:104590.
- [82] Yang, Guanyu, He, Yuting, Lv, Yang, et al. Multi-task learning for pulmonary arterial hypertension prognosis prediction via memory drift and prior prompt learning on 3D chest CT[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(4):1967–1978.
- [83] Li, Yuhong, Zhang, Xiaofan, and Chen, Deming. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. 1091–1100.
- [84] Dai, Jifeng, Qi, Haozhi, Xiong, Yuwen, et al. Deformable convolutional networks[C]. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2017. 764–773.
- [85] Li, Ruoyu, Wang, Sheng, Zhu, Feiyun, et al. Adaptive graph convolutional neural networks[C]. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2018.
- [86] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, et al. Deep residual learning for image recognition[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2016. 770–778.
- [87] Gao, Huang, Zhuang, Liu, and Weinberger, Kilian Q. Densely connected convolutional networks[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2017.
- [88] Wang, Jingdong, Sun, Ke, Cheng, Tianheng, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
- [89] Lin, Tsung-Yi, Dollár, Piotr, Girshick, Ross, et al. Feature pyramid networks for object detection[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2017. 2117–2125.
- [90] Çiçek, Özgün, Abdulkadir, Ahmed, Lienkamp, Soeren S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]. In: International conference on medical image computing and computer-assisted intervention. 2016. 424–432.

- [91] Xia, Yingda, Xie, Lingxi, Liu, Fengze, et al. Bridging the gap between 2D and 3D organ segmentation with volumetric fusion net[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2018. 445–453.
- [92] Isensee, Fabian, Jaeger, Paul F, Kohl, Simon AA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2021, 18(2):203–211.
- [93] Dong, Zhangfu, He, Yuting, Qi, Xiaoming, et al. MNet: Rethinking 2D/3D networks for anisotropic medical image segmentation[C]. In: Proceedings of the International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2022. 870–876. Main Track.
- [94] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30.
- [95] Ma, Ningning, Zhang, Xiangyu, Huang, Jiawei, et al. WeightNet: Revisiting the design space of weight networks[C]. In: Proceedings of the European Conference on Computer Vision. Cham: Springer International Publishing, 2020. 776–792.
- [96] Bommasani, Rishi, Hudson, Drew A, Adeli, Ehsan, et al. On the opportunities and risks of foundation models[J]. arXiv preprint arXiv:2108.07258, 2021.
- [97] Wiegreffe, Sarah and Pinter, Yuval. Attention is not explanation[C]. In: 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019. 2020. 11–20.
- [98] Qiu, Xipeng, Sun, Tianxiang, Xu, Yige, et al. Pre-trained models for natural language processing: A survey[J]. Science China Technological Sciences, 2020, 63(10):1872–1897.
- [99] Floridi, Luciano and Chiriatti, Massimo. GPT-3: Its nature, scope, limits, and consequences[J]. Minds and Machines, 2020, 30:681–694.
- [100] Dosovitskiy, Alexey, Beyer, Lucas, Kolesnikov, Alexander, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[C]. In: International Conference on Learning Representations. 2020.
- [101] Han, Kai, Wang, Yunhe, Chen, Hanting, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1):87–110.

- [102] Guo, Meng-Hao, Xu, Tian-Xing, Liu, Jiang-Jiang, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022, 8(3):331–368.
- [103] He, Kaiming, Chen, Xinlei, Xie, Saining, et al. Masked autoencoders are scalable vision learners[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 16000–16009.
- [104] Strudel, Robin, Garcia, Ricardo, Laptev, Ivan, et al. Segmenter: Transformer for semantic segmentation[C]. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. 7262–7272.
- [105] Li, Xiaomeng, Chen, Hao, Qi, Xiaojuan, et al. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes[J]. IEEE Transactions on Medical Imaging, 2018, 37(12):2663–2674.
- [106] Li, Junning, Lo, Pechin, Taha, Ahmed, et al. Segmentation of renal structures for image-guided surgery[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2018. 454–462.
- [107] Shi, Jiacheng, He, Yuting, Kong, Youyong, et al. Xmorpher: Full transformer for deformable medical image registration via cross attention[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2022. 217–226.
- [108] Xie, Enze, Wang, Wenhai, Yu, Zhiding, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34:12077–12090.
- [109] Hatamizadeh, Ali, Tang, Yucheng, Nath, Vishwesh, et al. Unetr: Transformers for 3d medical image segmentation[C]. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022. 574–584.
- [110] Hady, Mohamed Farouk Abdel and Schwenker, Friedhelm. Semi-supervised learning[EB/OL]. 2013.
- [111] Zhou, Zhi-Hua. A brief introduction to weakly supervised learning[J]. National Science Review, 2018, 5(1):44–53.
- [112] Jing, Longlong and Tian, Yingli. Self-supervised visual feature learning with deep neural networks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

- [113] Wang, Yaqing, Yao, Quanming, Kwok, James T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys, 2019, 53(3):63.
- [114] Qiao, Siyuan, Shen, Wei, Zhang, Zhishuai, et al. Deep co-training for semi-supervised image recognition[C]. In: Proceedings of the European Conference on Computer Vision. 2018. 135–152.
- [115] Gu, Xiaowei and Angelov, Plamen P. Semi-supervised deep rule-based approach for image classification[J]. Applied Soft Computing, 2018, 68:53–68.
- [116] Hung, Wei Chih, Tsai, Yi Hsuan, Liou, Yan Ting, et al. Adversarial learning for semi-supervised semantic segmentation[C]. In: 29th British Machine Vision Conference, BMVC 2018. 2019.
- [117] Tang, Yuxing, Wang, Josiah, Wang, Xiaofang, et al. Visual and semantic knowledge transfer for large scale semi-supervised object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12):3045–3058.
- [118] Rasmus, Antti, Berglund, Mathias, Honkala, Mikko, et al. Semi-supervised learning with ladder networks[C]. In: Advances in Neural Information Processing Systems. 2015. 3546–3554.
- [119] Wu, Chuhan, Wu, Fangzhao, Wu, Sixing, et al. Semi-supervised dimensional sentiment analysis with variational autoencoder[J]. Knowledge-Based Systems, 2019, 165:30–39.
- [120] Sedai, Suman, Mahapatra, Dwarikanath, Hewavitharanage, Sajini, et al. Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2017. 75–82.
- [121] Nie, Dong, Gao, Yaozong, Wang, Li, et al. ASDNet: Attention based semi-supervised deep networks for medical image segmentation[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2018. 370–378.
- [122] Bai, Wenjia, Oktay, Ozan, Sinclair, Matthew, et al. Semi-supervised learning for network-based cardiac MR image segmentation[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2017. 253–260.

- [123] You, Chenyu, Zhou, Yuan, Zhao, Ruihan, et al. Simcvd: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2022, 41(9):2228–2237.
- [124] Qu, Liangqiong, Wang, Shuai, Yap, Pew-Thian, et al. Wavelet-based semi-supervised adversarial learning for synthesizing realistic 7T from 3T MRI[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2019. 786–794.
- [125] He, Yuting, Li, Tiantian, Ge, Rongjun, et al. Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching[J]. IEEE Journal of Biomedical and Health Informatics, 2021.
- [126] He, Yuting, Li, Tiantian, Yang, Guanyu, et al. Deep complementary joint model for complex scene registration and few-shot segmentation on medical images[C]. In: Proceedings of the European Conference on Computer Vision. 2020. 770–786.
- [127] Zhao, Amy, Balakrishnan, Guha, Durand, Fredo, et al. Data augmentation using learned transformations for one-shot medical image segmentation[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. 8543–8553.
- [128] Chang, Yu-Ting, Wang, Qiaosong, Hung, Wei-Chih, et al. Weakly-supervised semantic segmentation via sub-category exploration[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [129] Liu, Xiaoming, Yuan, Quan, Gao, Yaozong, et al. Weakly supervised segmentation of COVID-19 infection with scribble annotation on CT images[J]. Pattern Recognition, 2022, 122:108341.
- [130] Lee, Hyeonsoo and Jeong, Won-Ki. Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2020. 14–23.
- [131] Wang, Juan and Xia, Bin. Bounding box tightness prior for weakly supervised image segmentation[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2021. 526–536.
- [132] Mahapatra, Dwarikanath, Bozorgtabar, Behzad, Thiran, Jean-Philippe, et al. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2018. 580–588.

- [133] Li, Haohan and Yin, Zhaozheng. Attention, suggestion and annotation: a deep active learning framework for biomedical image segmentation[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2020. 3–13.
- [134] Zhu, Linchao and Yang, Yi. Label independent memory for semi-supervised few-shot video classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1):273–285.
- [135] Zhang, Miao, Li, Huiqi, Pan, Shirui, et al. One-shot neural architecture search: Maximising diversity to overcome catastrophic forgetting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(9):2921–2935.
- [136] Snell, Jake, Swersky, Kevin, and Zemel, Richard. Prototypical networks for few-shot learning[C]. In: Advances in Neural Information Processing Systems. 2017. 4077–4087.
- [137] Ding, Yuhang, Yu, Xin, and Yang, Yi. Modeling the probabilistic distribution of unlabeled data for one-shot medical image segmentation[C]. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2021. 1246–1254.
- [138] Puch, Santi, Sánchez, Irina, and Rowe, Matt. Few-shot learning with deep triplet networks for brain imaging modality recognition[EB/OL]. 2019.
- [139] Narkhede, Meenal V, Bartakke, Prashant P, and Sutaone, Mukul S. A review on weight initialization strategies for neural networks[J]. Artificial Intelligence Review, 2022, 55(1):291–322.
- [140] Taher, Mohammad Reza Hosseinzadeh, Haghghi, Fatemeh, Feng, Ruibin, et al. A systematic benchmarking analysis of transfer learning for medical image analysis[C]. In: Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health workshop in MICCAI. 2021.
- [141] Tang, Yucheng, Yang, Dong, Li, Wenqi, et al. Self-supervised pre-training of swin transformers for 3d medical image analysis[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 20730–20740.
- [142] Wang, Xinlong, Zhang, Rufeng, Shen, Chunhua, et al. DenseCL: A simple framework for self-supervised dense visual pre-training[J]. Visual Informatics, 2022.
- [143] Erhan, Dumitru, Courville, Aaron, Bengio, Yoshua, et al. Why does unsupervised pre-training help deep learning?[C]. In: Proceedings of the International Conference on Artificial Intelligence and Statistics. 2010. 201–208.

- [144] Wu, Yue, Chen, Yinpeng, Wang, Lijuan, et al. Large scale incremental learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. 374–382.
- [145] Ade, RR and Deshmukh, PR. Methods for incremental learning: a survey[J]. International Journal of Data Mining & Knowledge Management Process, 2013, 3(4):119.
- [146] Kornblith, Simon, Shlens, Jonathon, and Le, Quoc V. Do better imagenet models transfer better?[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. 2661–2671.
- [147] Liu, Xiao, Zhang, Fanjin, Hou, Zhenyu, et al. Self-supervised learning: Generative or contrastive[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(1):857–876.
- [148] Zhou, Zongwei, Sodha, Vatsal, Pang, Jiaxuan, et al. Models genesis[C]. Elsevier, 2021. 101840.
- [149] Chen, Pengguang, Liu, Shu, and Jia, Jiaya. Jigsaw clustering for unsupervised visual representation learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. 11526–11535.
- [150] Komodakis, Nikos and Gidaris, Spyros. Unsupervised representation learning by predicting image rotations[C]. In: International Conference on Learning Representations. 2018.
- [151] Grill, Jean-Bastien, Strub, Florian, Altché, Florent, et al. Bootstrap your own latent-a new approach to self-supervised learning[C]. In: Advances in Neural Information Processing Systems. 2020. 21271–21284.
- [152] Chen, Ting, Kornblith, Simon, Norouzi, Mohammad, et al. A simple framework for contrastive learning of visual representations[C]. In: International Conference on Machine Learning. 2020. 1597–1607.
- [153] Chen, Xinlei and He, Kaiming. Exploring simple siamese representation learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. 15750–15758.
- [154] Kirkpatrick, James, Pascanu, Razvan, Rabinowitz, Neil, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences, 2017, 114(13):3521–3526.

- [155] Ljungberg, Borje, Bensalah, Karim, Canfield, Steven, et al. EAU guidelines on renal cell carcinoma: 2014 update[J]. European Urology, 2015, 67(5):913–924.
- [156] Shao, Pengfei, Qin, Chao, Yin, Changjun, et al. Laparoscopic partial nephrectomy with segmental renal artery clamping: technique and clinical outcomes[J]. European Urology, 2011, 59(5):849–855.
- [157] Shao, Pengfei, Tang, Lijun, Li, Pu, et al. Precise segmental renal artery clamping under the guidance of dual-source computed tomography angiography during laparoscopic partial nephrectomy[J]. European Urology, 2012, 62(6):1001–1008.
- [158] Porpiglia, Francesco, Fiori, Cristian, Checcucci, Enrico, et al. Hyperaccuracy three-dimensional reconstruction is able to maximize the efficacy of selective clamping during robot-assisted partial nephrectomy for complex renal masses[J]. European Urology, 2018, 74(5):651–660.
- [159] Zhang, Shaobo, Yang, Guanyu, Tang, Lijun, et al. Application of a Functional 3-dimensional Perfusion Model in Laparoscopic Partial Nephrectomy With Precise Segmental Renal Artery Clamping[J]. Urology, 2019, 125:98–103.
- [160] Petru, Bordei, Elena, Sapte, Dan, Iliescu, et al. Morphological assessments on the arteries of the superior renal segment[J]. Surgical and Radiologic Anatomy, 2012, 34(2):137–144.
- [161] Weston, Jason, Ratle, Frédéric, Mobahi, Hossein, et al. Deep learning via semi-supervised embedding[EB/OL]. 2012.
- [162] Lee, Chen-Yu, Xie, Saining, Gallagher, Patrick, et al. Deeply-supervised nets[C]. In: Proceedings of the International Conference on Artificial Intelligence and Statistics. 2015. 562–570.
- [163] Metz, Coert, Schaap, Michiel, Van Der Giessen, Alina, et al. Semi-automatic coronary artery centerline extraction in computed tomography angiography data[C]. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2007. 856–859.
- [164] Tscherren, Juerg, Hoffman, Eric A, McLennan, Geoffrey, et al. Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans[J]. IEEE Transactions on Medical Imaging, 2005, 24(12):1529.

- [165] Toledo, Ricardo, Orriols, Xavier, Binefa, Xavier, et al. Tracking elongated structures using statistical snakes[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2000. 157–162.
- [166] Mille, Julien and Cohen, Laurent D. Deformable tree models for 2D and 3D branching structures extraction[C]. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2009. 149–156.
- [167] Moccia, Sara, Momi, Elena De, Hadji, Sara El, et al. Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics[J]. Computer Methods and Programs in Biomedicine, 2018, 158:71–91.
- [168] Wolterink, Jelmer M, van Hamersveld, Robbert W, Viergever, Max A, et al. Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier[J]. Medical Image Analysis, 2019, 51:46–60.
- [169] Bai, Wenjia, Suzuki, Hideaki, Qin, Chen, et al. Recurrent neural networks for aortic image sequence segmentation with sparse annotations[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2018. 586–594.
- [170] Noothout, Julia MH, de Vos, Bob D, Wolterink, Jelmer M, et al. Automatic segmentation of thoracic aorta segments in low-dose chest CT[C]. In: Medical Imaging 2018: Image Processing. 2018. 105741S.
- [171] Taha, Ahmed, Lo, Pechin, Li, Junning, et al. Kid-Net: convolution networks for kidney vessels segmentation from CT-volumes[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2018. 463–471.
- [172] Pelt, Daniël M and Sethian, James A. A mixed-scale dense convolutional neural network for image analysis[J]. Proceedings of the National Academy of Sciences, 2018, 115(2):254–259.
- [173] Yang, Maoke, Yu, Kun, Zhang, Chi, et al. Denseaspp for semantic segmentation in street scenes[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. 3684–3692.
- [174] Masci, Jonathan, Meier, Ueli, Cireşan, Dan, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]. In: International Conference on Artificial Neural Networks. 2011. 52–59.

- [175] Kingma, Durk P, Mohamed, Shakir, Rezende, Danilo Jimenez, et al. Semi-supervised learning with deep generative models[C]. In: Advances in Neural Information Processing Systems. 2014. 3581–3589.
- [176] Xiao, Yawen, Wu, Jun, Lin, Zongli, et al. A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data[J]. Computer Methods and Programs in Biomedicine, 2018, 166:99–105.
- [177] Vincent, Pascal, Larochelle, Hugo, Lajoie, Isabelle, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.[J]. Journal of Machine Learning Research, 2010, 11(12).
- [178] Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1):1929–1958.
- [179] Wu, Yuxin and He, Kaiming. Group normalization[C]. In: Proceedings of the European Conference on Computer Vision. 2018. 3–19.
- [180] Yang, Guanyu, Gu, Jinjin, Chen, Yang, et al. Automatic kidney segmentation in CT images based on multi-atlas image registration[C]. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014. 5538–5541.
- [181] Baur, Christoph, Albarqouni, Shadi, and Navab, Nassir. Semi-supervised deep learning for fully convolutional networks[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2017. 311–319.
- [182] Chen, Xiaokang, Yuan, Yuhui, Zeng, Gang, et al. Semi-Supervised Semantic Segmentation With Cross Pseudo Supervision[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. 2613–2622.
- [183] Kingma, Diederik P. and Ba, Jimmy Lei. Adam: A method for stochastic optimization[C]. In: International Conference on Learning Representations. 2015.
- [184] Taha, Abdel Aziz and Hanbury, Allan. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool[J]. BMC Medical Imaging, 2015, 15(1):1–28.
- [185] Boskamp, Tobias, HAHN, HORST K, Hindennach, Milo, et al. Geometrical and structural analysis of vessel systems in 3D medical image datasets[EB/OL]. 2005.
- [186] Simone, Giuseppe, Gill, Inderbir S, Mottrie, Alexandre, et al. Indications, techniques, outcomes, and limitations for minimally ischemic and off-clamp partial nephrectomy: a systematic review of the literature[J]. European Urology, 2015, 68(4):632–640.

- [187] Ljungberg, Börje, Albiges, Laurance, Abu-Ghanem, Yasmin, et al. European association of urology guidelines on renal cell carcinoma: the 2019 update[J]. European Urology, 2019, 75(5):799–810.
- [188] Bianchi, Lorenzo, Barbaresi, Umberto, Cercenelli, Laura, et al. The impact of 3D digital reconstruction on the surgical planning of partial nephrectomy: a case-control study. Still time for a novel surgical trend?[J]. Clinical Genitourinary Cancer, 2020.
- [189] Nicolau, Stéphane, Soler, Luc, Mutter, Didier, et al. Augmented reality in laparoscopic surgical oncology[J]. Surgical Oncology, 2011, 20(3):189–201.
- [190] Herrell, S Duke, Galloway, Robert L, and Miga, Michael I. Image guidance in robotic-assisted renal surgery[EB/OL]. 2015.
- [191] Hamada, Akihiro, Sawada, Atsuro, Kono, Jin, et al. The current status and challenges in augmented-reality navigation system for robot-assisted laparoscopic partial nephrectomy[C]. In: International Conference on Human-Computer Interaction. 2020. 620–629.
- [192] Wang, Chenglong, Oda, Masahiro, Hayashi, Yuichiro, et al. Tensor-cut: A tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation[J]. Medical Image Analysis, 2020, 60:101623.
- [193] Jin, Chao, Shi, Fei, Xiang, Dehui, et al. 3D fast automatic segmentation of kidney based on modified AAM and random forest[J]. IEEE Transactions on Medical Imaging, 2016, 35(6):1395–1407.
- [194] He, Yuting, Yang, Guanyu, Chen, Yang, et al. DPA-densebiasnet: Semi-supervised 3d fine renal artery segmentation with dense biased network and deep priori anatomy[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2019. 139–147.
- [195] Young, Jonathan R, Margolis, Daniel, Sauk, Steven, et al. Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT[J]. Radiology, 2013, 267(2):444–453.
- [196] Hospedales, Timothy, Antoniou, Antreas, Micaelli, Paul, et al. Meta-learning in neural networks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9):5149–5169.
- [197] Rother, Carsten, Kolmogorov, Vladimir, and Blake, Andrew. "GrabCut" interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics, 2004, 23(3):309–314.

- [198] Dietterich, Thomas G et al. Ensemble learning[J]. *The Handbook of Brain Theory and Neural Networks*, 2002, 2:110–125.
- [199] Polikar, Robi. Ensemble learning[EB/OL]. 2012.
- [200] Wang, Shuangling, Yin, Yilong, Cao, Guibao, et al. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning[J]. *Neurocomputing*, 2015, 149:708–717.
- [201] Zheng, Hao, Zhang, Yizhe, Yang, Lin, et al. A new ensemble learning framework for 3D biomedical image segmentation[C]. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019. 5909–5916.
- [202] Bergstra, James and Bengio, Yoshua. Random search for hyper-parameter optimization[J]. *Journal of Machine Learning Research*, 2012, 13(1):281–305.
- [203] Tan, Mingxing and Le, Quoc V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[J]. *International Conference on Machine Learning*, 2019.
- [204] Real, Esteban, Aggarwal, Alok, Huang, Yanping, et al. Regularized evolution for image classifier architecture search[C]. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019. 4780–4789.
- [205] Huang, Gao, Liu, Zhuang, Van Der Maaten, Laurens, et al. Densely connected convolutional networks[C]. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2017. 4700–4708.
- [206] Wang, Guotai, Song, Tao, Dong, Qiang, et al. Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks[J]. *Medical Image Analysis*, 2020, 65:101787.
- [207] Ha, David, Dai, Andrew M, and Le, Quoc V. HyperNetworks[C]. In: *International Conference on Learning Representations*. 2017.
- [208] Brock, Andrew, Lim, Theo, Ritchie, JM, et al. SMASH: One-Shot Model Architecture Search through HyperNetworks[C]. In: *International Conference on Learning Representations*. 2018.
- [209] Munkhdalai, Tsendsuren and Yu, Hong. Meta networks[C]. In: *International Conference on Machine Learning*. 2017. 2554–2563.

- [210] Hu, Jie, Shen, Li, and Sun, Gang. Squeeze-and-excitation networks[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. 7132–7141.
- [211] Yang, Brandon, Bender, Gabriel, Le, Quoc V, et al. Condconv: Conditionally parameterized convolutions for efficient inference[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [212] Lin, Min, Chen, Qiang, and Yan, Shuicheng. Network in network[C]. In: International Conference on Learning Representations. 2014.
- [213] Paszke, Adam, Gross, Sam, Massa, Francisco, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [214] Xu, Zhenlin and Niethammer, Marc. DeepAtlas: Joint semi-supervised learning of image registration and segmentation[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2019. 420–429.
- [215] Wang, Shuxin, Cao, Shilei, Wei, Dong, et al. LT-Net: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [216] Avants, Brian B, Epstein, Charles L, Grossman, Murray, et al. Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain[J]. Medical Image Analysis, 2008, 12(1):26–41.
- [217] Liu, Fengze, Cai, Jinzheng, Huo, Yuankai, et al. Jssr: A joint synthesis, segmentation, and registration system for 3d multi-modal image alignment of large-scale pathological ct scans[C]. In: Proceedings of the European Conference on Computer Vision Workshops. 2020. 257–274.
- [218] Holden, Mark. A review of geometric transformations for nonrigid body registration[J]. IEEE Transactions on Medical Imaging, 2007, 27(1):111–128.
- [219] Ashburner, John. A fast diffeomorphic image registration algorithm[J]. Neuroimage, 2007, 38(1):95–113.
- [220] Balakrishnan, Guha, Zhao, Amy, Sabuncu, Mert R, et al. An unsupervised learning model for deformable medical image registration[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. 9252–9260.

- [221] Song, Hwanjun, Kim, Minseok, Park, Dongmin, et al. Learning from noisy labels with deep neural networks: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022.
- [222] Badrinarayanan, Vijay, Kendall, Alex, and Cipolla, Roberto. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481–2495.
- [223] Haskins, Grant, Kruger, Uwe, and Yan, Pingkun. Deep learning in medical image registration: a survey[J]. Machine Vision and Applications, 2020, 31(1):1–18.
- [224] Balakrishnan, Guha, Zhao, Amy, Sabuncu, Mert R., et al. VoxelMorph: A learning framework for deformable medical image registration[J]. IEEE Transactions on Medical Imaging, 2019, 38:1788–1800.
- [225] Elmahdy, Mohamed S, Wolterink, Jelmer M, Sokooti, Hessam, et al. Adversarial optimization for joint registration and segmentation in prostate CT radiotherapy[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2019. 366–374.
- [226] Dinsdale, Nicola K, Jenkinson, Mark, and Namburete, Ana IL. Spatial warping network for 3D segmentation of the hippocampus in MR images[C]. In: International Conference on Medical Image Computing and Computer Assisted Intervention. 2019. 284–291.
- [227] Yang, Heran, Sun, Jian, Li, Huibin, et al. Neural multi-atlas label fusion: Application to cardiac MR images[J]. Medical Image Analysis, 2018, 49:60–75.
- [228] Iglesias, Juan Eugenio and Sabuncu, Mert R. Multi-atlas segmentation of biomedical images: a survey[J]. Medical Image Analysis, 2015, 24(1):205–219.
- [229] Isgum, Ivana, Staring, Marius, Rutten, Annemarieke, et al. Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans[J]. IEEE Transactions on Medical Imaging, 2009, 28(7):1000–1010.
- [230] Zhang, Jun. Inverse-consistent deep networks for unsupervised deformable image registration[J]. arXiv preprint arXiv:1809.03443, 2018.
- [231] Zhuang, Xiahai, Li, Lei, Payer, Christian, et al. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge[J]. Medical Image Analysis, 2019, 58:101537.

- [232] Gharleghi, Ramtin, Samarasinghe, Dr. Gihan, Sowmya, Professor Arcot, et al. Automated segmentation of coronary arteries[EB/OL]. <https://doi.org/10.5281/zenodo.3819799>. 2020-March.
- [233] Schaap, Michiel, Metz, Coert T, van Walsum, Theo, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms[J]. Medical Image Analysis, 2009, 13(5):701–714.
- [234] Payette, Kelly, de Dumast, Priscille, Kebiri, Hamza, et al. An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset[J]. Scientific Data, 2021, 8(1):167.
- [235] Kennedy, David N, Haselgrove, Christian, Hodge, Steven M, et al. CANDIShare: A Resource for Pediatric Neuroimaging Data[J]. Neuroinformatics, 2012, 10(3):319.
- [236] Yu, Lequan, Wang, Shujun, Li, Xiaomeng, et al. Uncertainty-aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation[C]. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2019. 605–613.
- [237] Rueckert, Daniel, Sonoda, Luke I, Hayes, Carmel, et al. Nonrigid registration using free-form deformations: application to breast MR images[J]. IEEE Transactions on Medical Imaging, 1999, 18(8):712–721.
- [238] Milbich, Timo. Visual Similarity and Representation Learning[D]: [PhD thesis]. 2021.
- [239] Zhang, Borui, Zheng, Wenzhao, Zhou, Jie, et al. Attributable visual similarity learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 7532–7541.
- [240] Roth, Karsten, Milbich, Timo, and Ommer, Bjorn. Pads: Policy-adapted sampling for visual similarity learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. 6568–6577.
- [241] Netter, Frank H. Atlas of human anatomy, Professional Edition E-Book: including NetterReference. com Access with full downloadable image Bank[M]. Elsevier health sciences, 2014.
- [242] Sabharwal, Chaman L and Leopold, Jennifer L. A Completeness of metrics for topological relations in 3D qualitative spatial reasoning[J]. Polibits, 2015, 52:5–15.

- [243] Caron, Mathilde, Bojanowski, Piotr, Joulin, Armand, et al. Deep clustering for unsupervised learning of visual features[C]. In: Proceedings of the European Conference on Computer Vision. 2018. 132–149.
- [244] Li, Xiaoni, Zhou, Yu, Zhang, Yifei, et al. Dense semantic contrast for self-supervised visual representation learning[C]. In: Proceedings of the ACM International Conference on Multimedia. 2021. 1368–1376.
- [245] Bronstein, Michael M, Bruna, Joan, LeCun, Yann, et al. Geometric deep learning: going beyond euclidean data[J]. IEEE Signal Processing Magazine, 2017, 34(4):18–42.
- [246] Wang, Zhaoqing, Li, Qiang, Zhang, Guoxin, et al. Exploring set similarity for dense self-supervised representation learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 16590–16599.
- [247] Miller, Michael I. and Younes, Laurent. Group actions, homeomorphisms, and matching: A general framework[J]. International Journal of Computer Vision, 2001, 41(1):61–84.
- [248] Han, Joon Hee and Park, Jong Seung. Contour matching using epipolar geometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(4):358–370.
- [249] Rocco, Ignacio, Arandjelovic, Relja, and Sivic, Josef. Convolutional neural network architecture for geometric matching[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2017. 6148–6157.
- [250] Han, Kai, Rezende, Rafael S, Ham, Bumsub, et al. Scnet: Learning semantic correspondence[C]. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2017. 1831–1840.
- [251] He, Yuting, Ge, Rongjun, Qi, Xiaoming, et al. Learning better registration to learn better few-shot medical image segmentation: authenticity, diversity, and robustness[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022.
- [252] Yan, Ke, Cai, Jinzheng, Jin, Dakai, et al. SAM: Self-supervised learning of pixel-wise anatomical embeddings in radiological images[J]. IEEE Transactions on Medical Imaging, 2022, 41(10):2658–2669.
- [253] He, Kaiming, Fan, Haoqi, Wu, Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. 9729–9738.

- [254] Van Gansbeke, Wouter, Vandenhende, Simon, Georgoulis, Stamatios, et al. Revisiting contrastive methods for unsupervised learning of visual representations[C]. In: Advances in Neural Information Processing Systems. 2021. 16238–16250.
- [255] Haghghi, Fatemeh, Hosseinzadeh Taher, Mohammad Reza, Zhou, Zongwei, et al. Learning semantics-enriched representation via self-discovery, self-classification, and self-restoration[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2020. 137–147.
- [256] Haghghi, Fatemeh, Taher, Mohammad Reza Hosseinzadeh, Zhou, Zongwei, et al. Transferable visual words: Exploiting the semantics of anatomical patterns for self-supervised learning[J]. IEEE Transactions on Medical Imaging, 2021.
- [257] Weisstein, Eric W. Affine transformation[J]. <https://mathworld.wolfram.com/>, 2004.
- [258] Jaderberg, Max, Simonyan, Karen, Zisserman, Andrew, et al. Spatial transformer networks[J]. Advances in Neural Information Processing Systems, 2015, 28.
- [259] Bellet, Aurélien, Habrard, Amaury, and Sebban, Marc. Metric learning[J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2015, 9(1):1–151.
- [260] Saad, Walid, Han, Zhu, Debbah, Mérouane, et al. Coalitional game theory for communication networks[J]. IEEE Signal Processing Magazine, 2009, 26(5):77–97.
- [261] Gharleghi, Ramtin, Adikari, Dona, Ellenberger, Katy, et al. Automated segmentation of normal and diseased coronary arteries-the ASOCA challenge[J]. Computerized Medical Imaging and Graphics, 2022, page 102049.
- [262] Pathak, Deepak, Krahenbuhl, Philipp, Donahue, Jeff, et al. Context encoders: Feature learning by inpainting[C]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2016. 2536–2544.
- [263] Van der Maaten, Laurens and Hinton, Geoffrey. Visualizing data using t-SNE.[J]. Journal of Machine Learning Research, 2008, 9(11).

## 附录：本文使用的数据集介绍

**肾癌CTA影像数据集（私有）。** 本文第三章所使用的数据集来自江苏省人民医院放射科，回顾性地收集196名肾癌患者的腹部CTA图像，这些患者的一侧肾脏带有肾肿瘤，而另一侧肾脏健康。肾脏大小在85.76ml到262.78 ml之间，肿瘤大小在1.75ml到144.82 ml之间。肾脏肿瘤类型包括透明肾细胞癌、乳头状肾细胞癌、嫌色肾细胞癌、血管平滑肌脂肪瘤和嗜酸性腺瘤。所有图像都在西门子双源64层CT扫描仪上采集，并在采集过程中注射了造影剂。图像的体素大小在0.59mm/体素到0.74mm/体素之间。z方向上的切片厚度和间距固定为0.75mm/像素和0.5mm/像素。本文第四章所使用的数据集为第三章数据集的一个子集，根据图像中所涉及的肾脏、肾动脉、肾静脉、肾肿瘤四个结构的成像质量进行筛选，最终获得了123名仅有一侧肾脏肿瘤的患者的腹部CTA图像。

**心脏CCTA影像数据集（私有）。** 本文第六章所使用的数据集从东部战区总医院收集302位心脏疾病患者CCTA图像。这些图像通过SOMATOM Definition Flash CT扫描仪上采集获得，在扫描过程中病人被注射造影剂以获得清晰的心脏内部结构和冠状动脉结构。图像的x轴和y轴分辨率在0.25mm/体素至0.57mm/体素之间，大小为 $512 \times 512$ 个体素，z轴层厚在0.75mm/体素至3mm/体素之间，大小在128至994个体素之间。

**ASOCA数据集。** 该数据集为2020年ASOCA挑战赛所公开的数据集<sup>[232]</sup>，总共涉及60例CCTA图像，来自于Medrano-Gracia等人所构建的“冠状动脉图谱”数据库<sup>1</sup>。这60例数据根据数据库中所包含的医疗报告，对数据库中的CCTA数据进行筛选，最终获得了30名患有冠状动脉疾病的患者和30名没有心脏疾病的患者。所有图像都使用GE Light-Speed 64层CT扫描仪，根据心电门控在心脏舒张末期的时间点进行重建，并在采集过程中注射造影剂。图像的体素大小在0.3mm/体素到0.4mm/体素之间，z轴方向分辨率为0.625mm/体素。该数据集被应用于本文第五章和第六章的研究。

**CAT08数据集。** 该数据集为2008年CAT挑战赛所公开的数据集<sup>[233]</sup>，总共涉及32例CCTA图像，来自于荷兰鹿特丹大学医学中心。这32例数据从2005年6月至2006年6月期间接受心脏CTA检查的患者中随机选择获得，其中的20例数据使用Sensation 64层CT扫描仪采集，其他12例数据使用Somatom Definition 双源CT扫描仪采集，根据心电门控在心脏舒张末期的时间点进行重建，并在采集过程中注射造影剂。两种扫描仪的探测器宽度均为0.6毫米。数据集的平均体素分辨率为 $0.32 \times 0.32 \times 0.4$ 。该数据集被应用于本文第五章的研究。

<sup>1</sup><https://www.coronaryatlas.org/>

**MM-WHS数据集。** 该数据集为2017年MM-WHS挑战赛所公开的数据集<sup>[231]</sup>，总共涉及60例心脏CTA图像和60例心脏MR图像。60例心脏CTA数据来自中国上海的两个医疗中心，使用标准冠状动脉CT血管造影协议从飞利浦64层CT扫描仪上获得。所有数据覆盖了从上腹部到主动脉弓的整个心脏区域，轴向切片的面内分辨率为0.78mm/体素，平均切片厚度为1.60mm/体素。60例心脏MRI数据来自英国伦敦的两家医院，一组数据是从圣托马斯医院采用1.5T飞利浦扫描仪获得，另一组数据从Royal Brompton医院采用Siemens Magnetom Avanto 1.5T扫描仪获得。这些MR数据的采集分辨率为 $(1.6 > 2) \times (1.6 > 2) \times (2 > 3.2)$ mm/体素，并被重建为其采集分辨率的一半，即 $(0.8 > 1) \times (0.8 > 1) \times (1 > 1.6)$ mm/体素。该数据集中的60个CTA图像被用于本文的第五章和第六章的研究。

**CANDI数据集。** 该数据集来自马萨诸塞大学医学院从103名儿童和青少年上采集的T1加权MRI扫描图像，其中包括57名男性和46名女性，年龄为4-17岁。这些数据来自四个诊断组，包括29个健康对照组、20个精神分裂症谱系障碍、19个双相情感障碍伴精神病和35个双相情感障碍不伴精神病。这103例图像数据从1996年至2006年在麦克林医院脑成像中心用1.5特斯拉GE信号扫描仪获得，层厚为1.5mm/体素。该数据集被用于本文的第五章和第六章的研究。

## 作者简介

### 一、攻读博士学位期间发表的学术论文

#### 1) 与本文相关的一作学术论文:

- [A1] **He, Yuting**, Yang, Guanyu, Yang, Jian, Chen, Yang, Kong, Youyong, Wu, Jiasong, Tang, Lijun, Zhu, Xiaomei, Dillenseger, Jean-Louis, Shao, Pengfei, Zhang, Shaobo, Shu, Huazhong, Coatrieux, Jean-Louis and Li, Shuo. “Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation”, *Medical image analysis (Media)*, 63, 101722, 2020. (中科院一区, IF: 10.9, 医学图像分析顶刊) (对应论文第三章)
- [A2] **He, Yuting**, Yang, Guanyu, Yang, Jian, Ge, Rongjun, Kong, Youyong, Zhu, Xiaomei, Shaobo, Zhang, Shao, Pengfei, Shu, Huazhong, Dillenseger, Jean-Louis, Coatrieux, Jean-Louis, and Li, Shuo. “Meta grayscale adaptive network for 3D integrated renal structures segmentation”, *Medical Image Analysis (Media)*, 71, 102055, 2021. (中科院1区, IF=10.9, 医学图像分析顶刊) (对应论文第四章)
- [A3] **He, Yuting**, Ge, Rongjun, Qi, Xiaoming, Chen, Yang, Wu, Jiasong, Coatrieux, Jean-Louis, Yang, Guanyu, and Li, Shuo. “Learning better registration to learn better few-shot medical image segmentation: Authenticity, Diversity, and Robustness”, *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, (Early Access), 2022. (中科院一区, IF: 10.4, 人工智能顶刊) (对应论文第五章)
- [A4] **He, Yuting**, Yang, Guanyu, Chen, Yang, Kong, Youyong, Wu, Jiasong, Tang Lijun, Zhu, Xiaomei, Dillenseger, Jean-Louis, Shao, Pengfei, Zhang, Shaobo, Shu, Huazhong, Coatrieux, Jean-Louis, and Li, Shuo. “Dpa-densebiasnet: Semi-supervised 3d fine renal artery segmentation with dense biased network and deep priori anatomy”, *In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 139-147, 2019. (医学图像分析顶会) (对应论文第三章)
- [A5] **He, Yuting**, Li, Tiantian, Yang, Guanyu, Kong, Youyong, Chen, Yang, Shu, Huazhong, Coatrieux, Jean-Louis, Dillenseger, Jean-Louis, and Li, Shuo. “Deep complementary joint model for complex scene registration and few-shot segmentation on medical images”, *In Proceedings of European Conference on Computer Vision (ECCV)*, 770-786, 2020. (计算机视觉顶会) (对应论文第五章)

- [A6] He, Yuting, Ge, Rongjun, Qi, Xiaoming, Yang, Guanyu, Chen, Yang, Kong, Youyong, Shu, Huazhong, Coatrieux, Jean-Louis, and Li, Shuo. “EnMcGAN: Adversarial Ensemble Learning for 3D Complete Renal Structures Segmentation”, *In Proceedings of International Conference on Information Processing in Medical Imaging (IPMI)*, 465-477, 2021. (医学图像分析顶会) (对应论文第四章)
- [A7] He, Yuting, Yang, Guanyu, Ge, Rongjun, Chen, Yang, Coatrieux, Jean-Louis, Wang, Boyu, and Li, Shuo. “Geometric visual similarity learning in 3d medical image self-supervised pre-training”, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9538-9547, 2023. (CCF A类, 计算机视觉顶会) (对应论文第六章)

2) 其他一作/共同一作学术论文:

- [B1] He, Yuting, Li, Tiantian, Ge, Rongjun, Yang, Jian, Kong, Youyong, Zhu, Jian, Shu, Huazhong, Yang, Guanyu, and Li, Shuo. “Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching”, *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, 26(3), 1177-1187, 2021. (中科院一区, IF=7.7)
- [B2] Yang, Guanyu\*, He, Yuting\*, Lv, Yang, Chen, Yang, Coatrieux, Jean-Louis, Sun, Xiaoxuan, Wang, Qiang, Wei, Yongyue, Li, Shuo, and Zhu, Yinsu. “Multi-Task Learning for Pulmonary Arterial Hypertension Prognosis Prediction Via Memory Drift and Prior Prompt Learning on 3D Chest CT”, *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, 27(4), 1967-1978, 2023. (中科院一区, IF=7.7)
- [B3] He, Yuting, Ge, Rongjun, Wu, Jiasong, Coatrieux, Jean-Louis, Shu, Huazhong, Chen, Yang, Yang, Guanyu, and Li, Shuo. “Thin Semantics Enhancement via High-Frequency Prior Rule for Thin Structures Segmentation”, *In Proceedings of IEEE International Conference on Data Mining (ICDM)*, 1096-1101, 2021. (数据挖掘顶会)
- [B4] Wang, Song\*, He, Yuting\*, Kong, Youyong, Zhu, Xiaomei, Zhang, Shaobo, Shao, Pengfei, Dillenseger, Jean-Louis, Coatrieux, Jean-Louis, Li, Shuo, and Yang, Guanyu. “CPNet: Cycle Prototype Network for Weakly-Supervised 3D Renal Compartments Segmentation on CT Images”, *In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 592-602, 2021. (医学图像分析顶会)

3) 作为合作者参与的学术论文:

- [C1] Qi, Xiaoming, He, Yuting, Yang, Guanyu, Chen, Yang, Yang, Jian, Liu, Wangyan, Zhu, Yinsu, Xu, Yi, Shu, Huazhong, and Li, Shuo. “MVSGAN: Spatial-Aware Multi-View CMR Fusion for Accurate 3D Left Ventricular Myocardium Segmentation”, *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, 26(5), 2264-2275, 2021. (中科院一区, IF: 7.7)
- [C2] Qi, Yaolei, Xu, Han, He, Yuting, Li, Guanyu, Li, Zehang, Kong, Youyong, Coatrieux, Jean-Louis, Shu, Huazhong, Yang, Guanyu, and Tu, Shenxian. “Examinee-Examiner Network: Weakly Supervised Accurate Coronary Lumen Segmentation Using Centerline Constraint”, *IEEE Transactions on Image Processing (TIP)*, 30, 9429-9441, 2021. (CCF A类, IF: 10.6, 计算机视觉顶刊)
- [C3] Dong, Zhangfu, He, Yuting, Qi, Xiaoming, Chen, Yang, Shu, Huazhong, Coatrieux, Jean-Louis, Yang, Guanyu, and Li, Shuo. “MNet: Rethinking 2D/3D Networks for Anisotropic Medical Image Segmentation”, *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, 870–876, 2022. (CCF A类, 人工智能顶会)
- [C4] Qi, Yaolei., He, Yuting, Qi, Xiaoming, Zhang, Yuan, and Yang, Guanyu. “Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation”, *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6070-6079, 2023. (CCF A类, 计算机视觉顶会)

## 二、攻读学位期间授权的专利

- 杨冠羽, 何宇霆, 戚耀磊, 朱晓梅, 张少波, 陈阳, 孔佑勇, 舒华忠, “基于密集偏置网络和自编码器的半监督肾动脉分割方法”, 已授权, 授权公告号: CN 110428427 B
- 杨冠羽, 何宇霆, 李甜甜, 赵子腾, 吴显政, 伍家松, 孔佑勇, 舒华忠, “一种基于三维卷积神经网络的CT造影图像肾动脉分割方法”, 已授权, 授权公告号: CN 109801268 B

## 三、攻读博士学位期间主持的课题

- 东南大学优秀博士学位论文培育基金, 基于深度学习的肾癌影像智能处理与分析算法研究 (基金号: YBYP2139)

## 四、攻读学位期间的学术服务

- 在MICCAI2022上组织 MICCAI 2022 Challenge on Kidney PArsing Challenge 2022: Multi-Structure Segmentation for Renal Cancer Treatment (KiPA)





心於至善

---

---

