



云计算技术与大数据课程报告

研究主题：医疗大数据处理

论文题目： Towards Generic Semi-Supervised Framework for Volumetric Medical Image Segmentation

学号：1024041006

姓名：吴晟羽



论文详细信息

[1]Wang Haonan, Li Xiaomeng. Towards Generic Semi-Supervised Framework for Volumetric Medical Image Segmentation [C]// Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023). New Orleans, USA: NeurIPS, 2023: 1–13.



01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望



01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





一、引言

背景、挑战与目标

- 问题背景：3D医学图像标注在临床诊断中至关重要，但标注数据稀缺且耗时费力。因此，利用半监督学习（SSL）方法，通过有限标注数据来训练模型成为重要方向
- 挑战：现有SSL方法假设标注和未标注数据来自同一领域，忽略了实际应用中常见的领域偏移问题，泛化能力差，模型易过拟合
- 研究目标：为了解决现有方法存在的局限性，提出一个通用的SSL框架，能够处理SSL、无监督领域适应（UDA）和半监督领域泛化（SemiDG）三种场景。该框架通过提取分布不变特征和解耦标注与未标注数据的训练过程，提升模型的泛化能力，推动SSL方法在临床场景中的应用

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





目录

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





二、相关工作

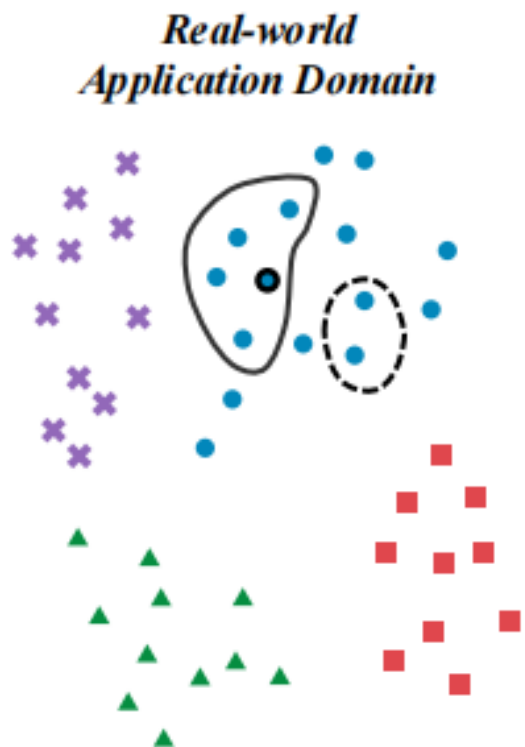
- 半监督分割：表现出色，但大多假设标注和未标注数据来自同一领域，无法处理**领域偏移**问题
- 无监督领域适应：通过联合训练源领域和目标领域的数据来解决领域偏移问题，依赖于目标领域的未标注数据，限制了其**泛化能力**
- 半监督领域泛化：训练时使用多个领域的数据，但不使用目标领域的标注数据，在实际应用中表现不佳，**缺乏通用性**
- 扩散模型具备强大的分布建模能力，已用于图像生成和分割。本文首次将其用于**提取跨域不变特征**



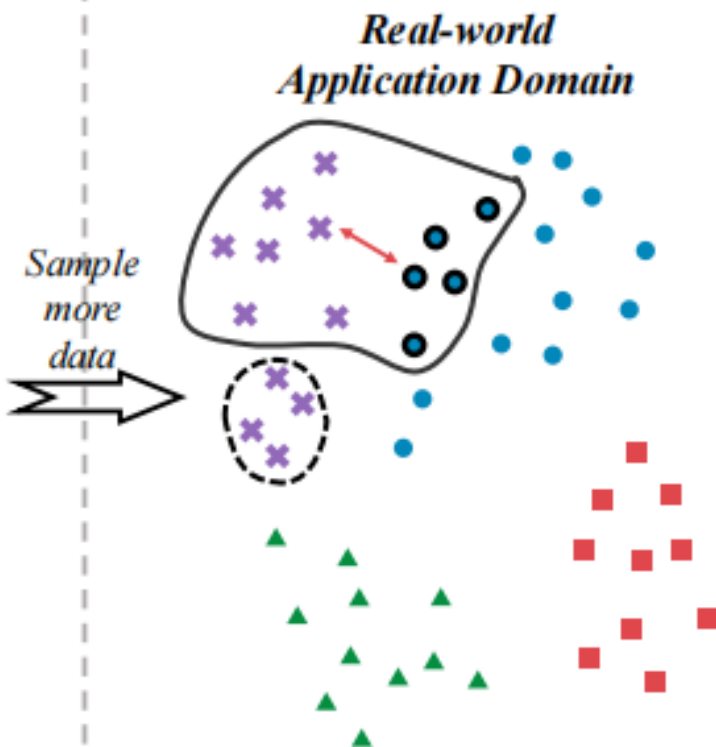
二、相关工作

✕ ● ▲ Labeled data from different domains ■ Data from any unseen domains ○ Training set ○ Testing set ↔ Domain shifts

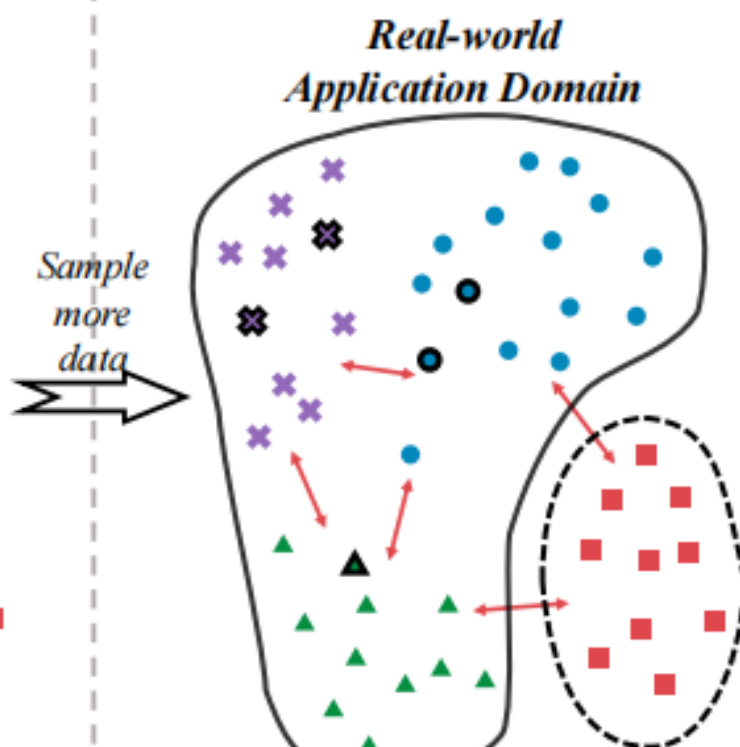
✕ ● ▲ Unlabeled data from different domains



(a) SSL



(b) UDA



(c) SemiDG

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





目录

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





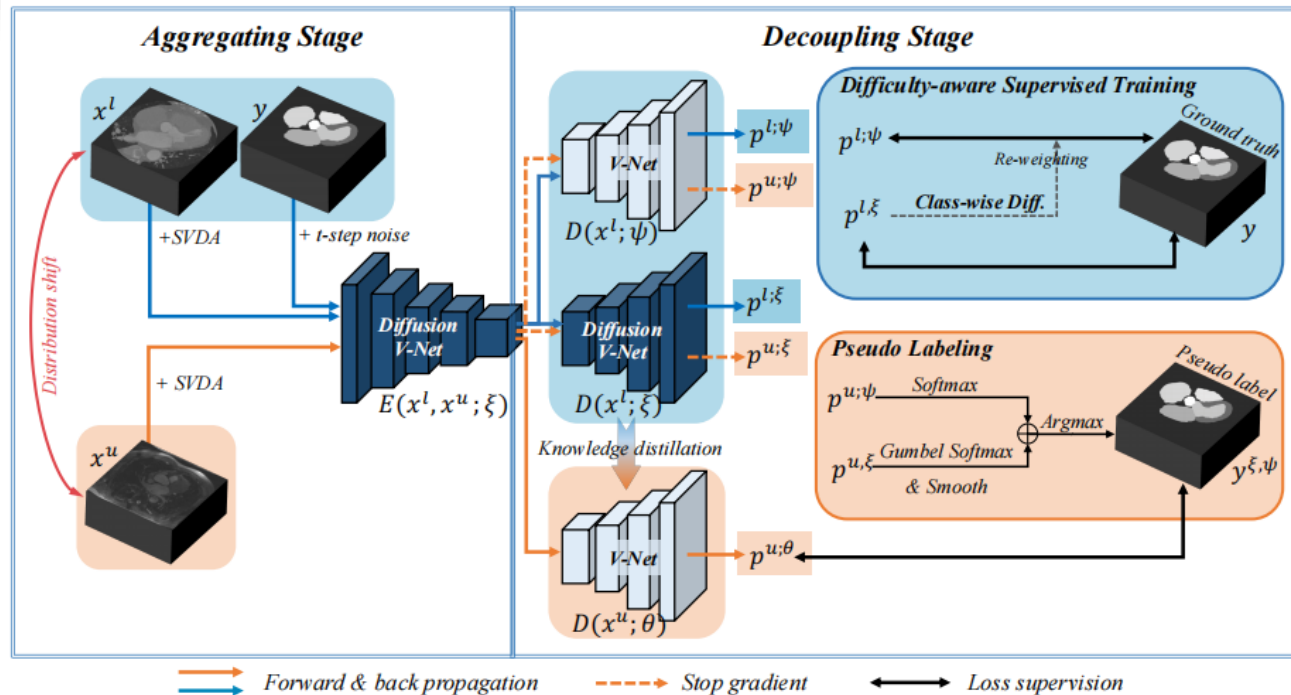
三、算法、模型

Aggregating & Decoupling 框架

A&D 框架由Aggregating（聚合）阶段和 Decoupling（解耦）阶段组成。

聚合：学习分布不变特征，构建一个跨域通用的知识集；

解耦：分离标注数据和未标注数据的训练过程，避免模型过度依赖标注数据，提高泛化能力。





三、算法、模型

Aggregating Stage

1) Sampling-based Volumetric Data Augmentation (SVDA)

目标：增加数据的多样性，使模型学习到更通用的特征

过程：

- a) 输入原始3D医学图像
- b) 随机选择多种数据增强操作（三维空间变换（随机裁剪、随机旋转、随机缩放）与体素值变换（高斯模糊、亮度调整、对比度调整、伽马校正））应用于标注和未标注数据
- c) 输出操作后的图像



三、算法、模型

Aggregating Stage

2) Diffusion for Capturing Invariant Features

目标：通过扩散模型提取分布不变特征

过程：

- 将标注数据的标签转换为one-hot format，逐步添加噪声，生成噪声标签
- 将噪声标签与标注数据拼接后输入到Diff-VNet
- 去噪流：编码器处理b)的输入和时间步长，生成时间步长嵌入的多尺度特征
- 困难感知训练流和未标注数据流：编码器仅处理图像，对应生成多尺度特征



三、算法、模型

Decoupling Stage

步骤	目标	过程	公式
监督去噪训练	生成域无偏的伪标签	输入标注数据 x_l 和噪声标签 y_l ，通过扩散模型的编码器和解码器生成无偏伪标签 p_u^ξ ，并计算损失函数 L_{deno} 。	$L_{\text{deno}} = \frac{1}{N_L} \sum_{i=0}^{N_L} L_{\text{DiceCE}}(p_l^\xi, y)$
监督困难感知训练	生成类别无偏的伪标签	输入标注数据 x_l 和标签 y ，通过困难感知解码器生成类别无偏伪标签 p_u^ψ ，并计算加权损失函数 L_{diff} 。	$L_{\text{diff}} = \frac{1}{N_L} \sum_{i=0}^{N_L} \sum_{k=0}^K w_{\text{diff}}^k L_{\text{DiceCE}}(p_l^\psi, y)$
伪标签生成	融合域无偏伪标签和类别无偏伪标签，生成高质量的伪标签。	使用 Reparameterize & Smooth (RS) Strategy 融合两种伪标签，生成最终的伪标签 $y_{\xi, \psi}$ 。	$y_{\xi, \psi} = \text{argmax}(\text{Gumbel-Softmax}(p_u^\xi) + \text{Softmax}(p_u^\psi))$
无监督训练	使用生成的伪标签监督未标注数据的训练，进一步优化模型。	输入未标注数据 x_u 和生成的伪标签 $y_{\xi, \psi}$ ，通过预测解码器生成预测标签 p_u^θ ，并计算损失函数 L_u 。	$L_u = \frac{1}{N_U} \sum_{i=0}^{N_U} L_{\text{DiceCE}}(p_u^\theta, y_{\xi, \psi})$

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





目录

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





四、实验与性能分析

数据集

数据集	数据类型	用途
LASeg	心房MRI (3D体数据)	半监督学习 (SSL)
Synapse	多器官CT (腹部区域)	类别不均衡半监督学习 (ImbSSL)
MMWHS	心脏MRI & CT (双模态)	无监督域适应 (UDA)
M&Ms	多中心心脏MRI (不同医院设备)	半监督领域泛化 (SemiDG)



四、实验与性能分析

评估指标

指标	作用
Dice分数和Jaccard指数	评估预测分割和真实分割之间的相似度，适用于评估分割的整体质量
ASD (平均表面距离)	评估分割结果与真实标签之间的平均距离，适用于评估分割边界的精度。
HD95 (95% Hausdorff距离)	评估分割结果与真实标签之间的最大距离，适用于评估分割边界的精度。



四、实验与性能分析

实验结果

Methods		Avg. Dice	Avg. ASD	Dice of Each Class													
				Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	PA	RAG	LAG	
	V-Net (fully)	62.09±1.2	10.28±3.9	84.6	77.2	73.8	73.3	38.2	94.6	68.4	72.1	71.2	58.2	48.5	17.9	29.0	
General	UA-MT [8]	20.26±2.2	71.67±7.4	48.2	31.7	22.2	0.0	0.0	81.2	29.1	23.3	27.5	0.0	0.0	0.0	0.0	
	URPC [10]	25.68±5.1	72.74±15.5	66.7	38.2	56.8	0.0	0.0	85.3	33.9	33.1	14.8	0.0	5.1	0.0	0.0	
	CPS [3]	33.55±3.7	41.21±9.1	62.8	55.2	45.4	35.9	0.0	91.1	31.3	41.9	49.2	8.8	14.5	0.0	0.0	
	SS-Net [17]	35.08±2.8	50.81±6.5	62.7	67.9	60.9	34.3	0.0	89.9	20.9	61.7	44.8	0.0	8.7	4.2	0.0	
	DST [6]	34.47±1.6	37.69±2.9	57.7	57.2	46.4	43.7	0.0	89.0	33.9	43.3	46.9	9.0	21.0	0.0	0.0	
	DePL [5]	36.27±0.9	36.02±0.8	62.8	61.0	48.2	54.8	0.0	90.2	36.0	42.5	48.2	10.7	17.0	0.0	0.0	
Imbalance	Adsh [71]	35.29±0.5	39.61±4.6	55.1	59.6	45.8	52.2	0.0	89.4	32.8	47.6	53.0	8.9	14.4	0.0	0.0	
	CReST [26]	38.33±3.4	22.85±9.0	62.1	64.7	53.8	43.8	8.1	85.9	27.2	54.4	47.7	14.4	13.0	18.7	4.6	
	SimiS [28]	40.07±0.6	32.98±0.5	62.3	69.4	50.7	61.4	0.0	87.0	33.0	59.0	57.2	29.2	11.8	0.0	0.0	
	Basak et al. [33]	33.24±0.6	43.78±2.5	57.4	53.8	48.5	46.9	0.0	87.8	28.7	42.3	45.4	6.3	15.0	0.0	0.0	
	CLD [15]	41.07±1.2	32.15±3.3	62.0	66.0	59.3	61.5	0.0	89.0	31.7	62.8	49.4	28.6	18.5	0.0	5.0	
	DHC [25]	48.61±0.9	10.71±2.6	62.8	69.5	59.2	66.0	13.2	85.2	36.9	67.9	61.5	37.0	30.9	31.4	10.6	
	A&D (ours)	60.88±0.7	2.52±0.4	85.2	66.9	67.0	52.7	62.9	89.6	52.1	83.0	74.9	41.8	43.4	44.8	27.2	

➤ 20%标注数据下，针对类别不平衡SSL任务的Synapse数据集结果

Method	2% Labeled data					5% Labeled data				
	Domain A	Domain B	Domain C	Domain D	Average	Domain A	Domain B	Domain C	Domain D	Average
nnUNet [69]	52.87	64.63	72.97	73.27	65.94	65.30	79.73	78.06	81.25	76.09
SDNet+Aug [53]	54.48	67.81	76.46	74.35	68.28	71.21	77.31	81.40	79.95	77.47
LDDG [78]	59.47	56.16	68.21	68.56	63.16	66.22	69.49	73.40	75.66	71.29
SAML [79]	56.31	56.32	75.70	69.94	64.57	67.11	76.35	77.43	78.64	74.88
BCP [76]*	71.57	76.20	76.87	77.94	75.65	73.66	79.04	77.01	78.49	77.05
DGNet [52]	66.01	72.72	77.54	75.14	72.85	72.40	80.30	82.51	83.77	79.75
vMFNet [55]	73.13	77.01	81.57	82.02	78.43	77.06	82.29	84.01	85.13	82.12
A&D (ours)	79.62	82.26	80.03	83.31	81.31	81.71	85.44	82.18	83.9	83.31

➤ 针对半监督领域泛化任务的M&Ms数据集结果



四、实验与性能分析

实验结果

Table 2: Results on two settings of LASEg dataset for **SSL** task.

5% labeled data (labeled:unlabeled=4:76)				
Method	Dice	Metrics		
		Jaccard	95HD	ASD
V-Net (fully)	91.47	84.36	5.48	1.51
V-Net (5%)	52.55	39.60	47.05	9.87
UA-MT [8]	82.26	70.98	13.71	3.82
SASSNet [9]	81.60	69.63	16.16	3.58
DTC [11]	81.25	69.33	14.90	3.99
URPC [10]	82.48	71.35	14.65	3.65
MC-Net [12]	83.59	72.36	14.07	2.70
SS-Net [17] [†]	86.33	76.15	9.97	2.31
BCP [76] [†]	88.02	78.72	7.90	2.15
A&D (ours)	89.93	81.82	5.25	1.86

10% labeled data (labeled:unlabeled=8:72)				
Method	Dice	Metrics		
		Jaccard	95HD	ASD
V-Net (10%)	82.74	71.72	13.35	3.26
UA-MT [8]	87.79	78.39	8.68	2.12
SASSNet [9]	87.54	78.05	9.84	2.59
DTC [11]	87.51	78.17	8.23	2.36
URPC [10]	86.92	77.03	11.13	2.28
LMISA-3D [77] [*]	86.06	76.53	12.99	2.41
vMFNet [55] [*]	73.88	62.56	16.81	5.04
MC-Net [12]	87.62	78.25	10.03	1.82
SS-Net [17] [†]	88.55	79.62	7.49	1.90
BCP [76] [†]	89.62	81.31	6.81	1.76
A&D (ours)	90.31	82.40	5.55	1.64

[†] use test set for validation, we use labeled data instead.

^{*} SOTA methods tailored for UDA and SemiDG, respectively.

Table 3: Results on two settings of MMWHS dataset for **UDA** task.

MR to CT						
Method	Dice				ASD	Average
	AA	LAC	LVC	MYO		
Supervised Training	92.7	91.1	91.9	87.8	90.9	2.2
PnP-AdaNet [38]	74.0	68.9	61.9	50.8	63.9	12.8
AdaOutput [37]	65.2	76.6	54.4	43.6	59.9	9.6
CycleGAN [34]	73.8	75.7	52.3	28.7	57.6	10.8
CyCADA [36]	72.9	77.0	62.4	45.3	64.4	9.4
SIFA [41]	81.3	79.5	73.8	61.6	74.1	7.0
DSFN [42]	84.7	76.9	79.1	62.4	75.8	N/A
DSAN [43]	79.9	84.8	82.8	66.5	78.5	5.9
LMISA-3D [77]	84.5	82.8	88.6	70.1	81.5	2.3
A&D (ours)	93.2	89.5	91.7	86.2	90.1	1.7

CT to MR						
Method	Dice				ASD	Average
	AA	LAC	LVC	MYO		
Supervised Training	82.8	80.5	92.4	78.8	83.6	2.9
PnP-AdaNet [38]	43.7	68.9	61.9	50.8	63.9	8.9
AdaOutput [37]	60.8	39.8	71.5	35.5	51.9	5.7
CycleGAN [34]	64.3	30.7	65.0	43.0	50.7	6.6
CyCADA [36]	60.5	44.0	77.6	47.9	57.5	7.9
SIFA [41]	65.3	62.3	78.9	47.3	63.4	5.7
DSAN [43]	71.3	66.2	76.2	52.1	66.5	5.4
LMISA-3D [77]	60.7	72.4	86.2	64.1	70.8	3.6
SS-Net [17] [*]	62.1	58.4	68.9	51.4	60.2	5.9
BCP [76] [*]	63.6	63.7	70.9	58.0	64.1	4.5
A&D (ours)	62.8	87.4	61.3	74.1	71.4	7.9

^{*} SOTA methods on semi-supervised segmentation.

磁共振成像 (MRI) 到计算机断层扫描 (CT) 的领域适应

模型使用MR图像的标注数据进行训练，在CT图像上进行测试



四、实验与性能分析

消融研究

Methods	A&D	w/o SVDA	w/o Diffusion	w/o DRS	w/o RS
IBSSL	60.9	55.3	56.7	52.0	58.7
UDA	90.1	84.6	79.2	85.8	87.0
SemiDG	80.6	77.9	74.9	78.2	78.6

- 在20%标注数据的Synapse (IBSSL)、MMWHS的MR到CT迁移 (UDA)、2%标注数据的M&Ms (SemiDG) 中，对框架组件的消融研究。验证了每个组件对整体性能的贡献。

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





目录

01

引言

02

相关工作

03

算法、模型

04

实验与性能分析

05

总结与展望





五、总结与展望

- 总结：提出一种半监督学习（SSL）框架——Aggregating & Decoupling (A&D)，旨在多种实际应用场景中解决医学图像分割问题，包括半监督学习、类别不平衡SSL、无监督领域适应和半监督领域泛化。通过聚合和解耦，框架能够有效地提取分布不变特征，避免过拟合，并处理类别不平衡问题。在四个基准数据集上进行了实验，结果表明A&D框架均能取得显著的性能提升
- 展望：该方法在计算效率和深度信息处理方面仍有改进空间。未来将集中在提高计算效率、改进深度信息处理、扩展多模态数据融合以及进一步提升模型的泛化能力，以更好地应对实际应用中的复杂场景。

The End

