arXiv:2305.05789v2 [cs.CV] 6 Jul 2023

基于特征空间密度匹配的医学图像分割无监督域适应

Tushar Kataria[1,2], Beatrice Knudsen[3], and Shireen Elhabian[1,2]

[1] 犹他大学卡勒特计算机学院[2] 犹他大学科学计算与成像研究所[3] 犹他大学病理学系 {tushar.kataria,shireen}@sci.utah.edu, beatrice.knudsen@path.utah.edu

摘要。 语义分割是自动化图像解读与分析中的关键步骤，其任务是将像素分类为一个或多个预先定义的、具有语义意义的类别。语义分割的深度学习方法依靠带标注的图像来学习能够表征这些语义类别的特征。然而，当训练数据（即 源域数据）与部署时遇到的数据集（即 目标域数据集）之间存在显著的域（即分布）偏移时，这些方法往往无法实现良好的泛化，因此需要对目标数据进行人工标注，才能获得可接受的性能。这一点在医学影像领域尤为重要，因为不同的成像模态会因成像方案和设备厂商的差异而产生显著的站点内和站点间变化。现有技术对超参数调整和目标数据集的规模较为敏感。本文提出一种用于语义分割的无监督域适应方法，以免对目标数据进行标注。我们采用核密度估计，在特征空间中将目标数据分布与源域数据分布进行匹配，尤其适用于目标样本数量有限的情况（占目标数据集规模的 3%）。我们在 2 个数据集上展示了所提方法的有效性：多中心前列腺磁共振成像数据集和组织病理图像数据集。

关键词： 域适应 · 语义分割 · 密度估计与匹配。

1 引言

人类视觉系统能够对环境中的每个物体进行分类和勾勒（分割）。医学影像中的分割尤为重要，因为勾勒相关对象（例如肿瘤、病变组织、癌症）需要高度专业的领域知识 [14]。准确识别这些对象的精确边界（或肿瘤的大小），对于实现可靠且可解释的疾病诊断、分析和治疗规划自动化至关重要[23]。

当使用具有代表性且数量充足的训练数据进行训练时，深度学习模型能够持续做出更准确的预测。然而，

这些模型可能会专注于学习虚假信号 [4] ，而非真实疾病病理特征。深度学习模型学习的低层次纹理特征多于高层次形状/形态特征 [8], ，这会影响所学习模型（在源数据集上训练）的性能，尤其是在推理过程中引入具有不同低层次数据统计特性的新数据（目标数据集）时。这被称为输入数据集中的分布偏移（或域偏移）。这种偏移会导致模型基于新数据做出的预测失去精确性和可信度。即使是输入图像仅为相同物体的素描这种轻微的分布偏移，也已被证明会导致模型性能显著下降 [22]。这种域偏移在医学影像 [12] 领域尤其棘手，因为获取大量训练数据受到限制，因此我们必须依赖经过训练的预训练模型。

域适应方法被提出，用于解决域偏移导致模型性能下降的问题 [20,21,13]。根据目标域中标注的可用性，人们提出了多种有监督和无监督域适应（UDA）技术[18,24,13]。无监督域适应技术更具优势，因为分割任务的逐像素标注，尤其是在医学图像领域，因需要专业知识而成本高昂，难以获取 [1]。

用于语义分割的无监督域适应方法大致可分为三类。第一类是对抗式域适应 [20,2,7],，其目标是利用源域和目标域特征，通过最大化域分类损失来学习域无关骨干特征，并通过梯度反转层向特征提取骨干网络传递负梯度。第二类是傅里叶域适应 [27,26], ，它基于域间相位信息不变的假设，利用傅里叶域变换调整频率幅度。第三类是密度匹配，即匹配输入空间 [3], 、输出空间 [19], 或特征空间 [15] 中的源域和目标域密度。 [3] 使用条件生成对抗网络（生成对抗网络）将源数据集中的图像转换为目标数据集的样貌，而 [19]和 [15]仅使用判别器来匹配源域与目标域特征之间的密度。也有人尝试结合其他惩罚项进行密度匹配，例如最大均值差异（MMD） [11,5,10] 或Wasserstein生成对抗网络 [5], 。基于对抗的方法对超参数选择非常敏感 [20,2]。傅里叶域适应框架对频率空间的选择和混合比例较为敏感。密度匹配方法对超参数[11,5], 非常敏感，并且由于涉及极小极大博弈而难以训练。它们还需要大量目标域数据才能收敛。

在此，我们提出一种新型的医学图像分割无监督域适应方法，通过（1）核密度估计（KDE）进行非参数密度估计，以及（2）通过詹森-香农散度（JSD）匹配密度，以适应分割网络中学习到的特征。事实证明，KDE [25,9]在较小数据集 [16]上的生成建模中表现更好。因此，与对抗学习相比，KDE 为匹配源域和目标域的特征密度提供了更稳定的解决方案，尤其是在医学影像中常见的低样本量场景下。KDE 的非参数特性

与仅使用矩来匹配密度的 MMD [11], 相比，它为域适应提供了丰富的训练信号。此外，核密度估计支持训练期间按批次进行密度匹配，通过批次样本匹配特征空间中的完整密度。核带宽通过随机抽取训练样本并将其映射到特征空间来估计。源域和目标数据集的估计密度通过 JSD 损失进行匹配。所提出的方法对特征空间进行正则化，使模型能够学习与域无关的通用特征。

我们将结果与其他密度匹配方法进行比较，例如采用固定带宽的 MMD [11,10], ，这与其他已提出的方法 [11,5]一致。我们还将结果与对抗训练 [20,2]、使用特征空间判别器进行密度匹配 [15]以及在输出空间进行密度匹配 [19]的方法进行比较。我们的方法与特征空间 [15]和输出空间 [19] 中的密度匹配密切相关，但我们不使用判别器进行密度匹配，而是使用 JSD 衡量散度，并使用核密度估计来估计底层特征概率分布。我们遵循相应论文中列出的方法，实现了自己的版本以供比较。本文的贡献如下：-

– 我们提出了一种用于语义分割无监督域适应的新方法，该方法基于对底层特征分布的丰富（非参数）表征。– 我们证明，当目标数据集规模较小时（占目标数据集大小的 3% 或 30%），所提出的方法在统计意义上优于其他密度匹配方法。– 我们通过消融实验评估特征空间选择、带宽估计频率、目标域数据样本数量以及 KDE样本数量的影响，并在不同模态（组织病理学 [6,17] 和多中心磁共振成像 [12]）上验证了所提出方法的有效性。

2 方法

2.1 问题设定

大多数用于语义分割的深度学习架构都采用如图1所示的编码器-解码器结构。设 $f_\theta(.)$ 为编码器， $g_\phi(.)$ 为解码器。对于输入图像 $\mathbf{I}$，模型执行以下运算：$\mathbf{x} = f_\theta(\mathbf{I})$ 和 $\mathbf{y} = g_\phi(\mathbf{x})$，其中 $\mathbf{x}$ 是学习到的特征空间中的编码特征。对于分割网络，我们可以有多个深层特征编码和解码空间，但为简单起见，我们假设最深层特征空间为 $\mathbf{x}$ （其空间分辨率最低、通道分辨率最高）。当输入空间发生域偏移时，深度学习模型无法泛化。我们假设这种域偏移会导致学习到的模型的特征空间发生密度偏移，从而使模型在未见过的数据上失效。我们提出，如果通过两个域的特征空间分布之间的密度匹配损失对模型进行正则化，那么模型在遇到新域时就不会受到同样的域偏移影响。这里有两个主要方面

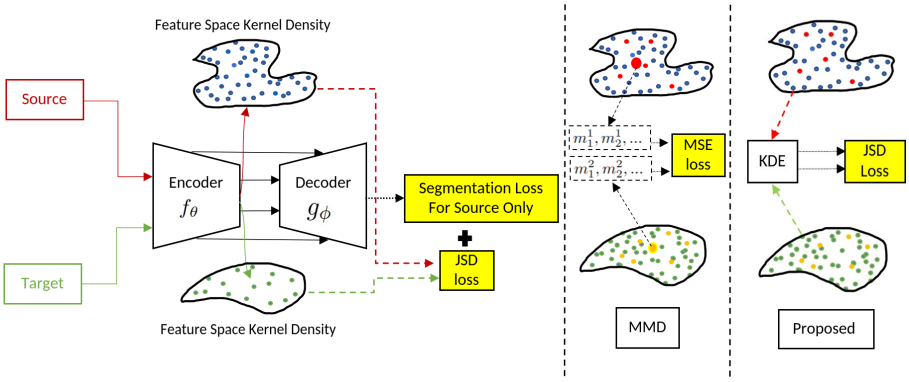

图1. 所提方法的框图。模型采用标准的编码器-解码器配置，并带有跳跃连接；训练时仅使用源数据集的标注。分别从源域和目标域提取深层特征，并使用核密度估计来估计其密度；最大均值差异仅匹配矩，而所提方法则使用 JSD 损失来匹配通过核密度估计得到的特征空间密度。

需要解决特征空间密度匹配问题：（1）密度的表示方式；（2）密度匹配损失。

密度的表示。特征空间中的密度可以用矩（均值、方差）表示，这种表示假设特征空间的默认分布为因子化高斯分布。然而，因子化高斯分布隐含着一种限制性假设，即特征空间中的分布是单峰且解耦的。我们也可以假设参数化密度遵循多元高斯分布或高斯混合分布的某些特征。然而，这两种方法都对样本点的分布作出了强假设，且并非由数据驱动。另一方面，核密度估计等非参数方法不会作出如此强的假设，因此更适合从数据中学习。因此，这些方法能够对特征空间密度进行更丰富、更灵活的描述。

密度匹配损失。KL散度具有非对称性，因此可能不适用于域适应任务，因为它是一种单向损失。另一方面，JSD是对称的，旨在约束源域特征接近目标域特征，反之亦然。

2.2 基于核密度估计的无监督域适应

我们所提方法的框图如图1所示。分割模型仅使用源数据集中的标注进行训练。在我们的设置中，不使用来自目标数据集的任何标注。密度匹配损失作为正则化项，确保源数据集与目标数据集的特征分布不会彼此偏离。因此，网络通过以下损失函数进行训练：

$$\mathcal{L} = \mathcal{L}_{(seg,source)} + \lambda JSD[p_s, p_t] \tag{1}$$

其中， $\mathcal{L}_{(seg,source)}$ 表示源数据集上的监督分割损失，λ 是一个用于定义密度匹配损失贡献程度的超参数，而 p_s 和 p_t 分别表示源数据集和目标数据集的密度估计。

核密度估计。 设 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$，……，$\mathbf{x}_N$为从编码特征空间中采样的点数。核密度估计值 $p_{est}(\mathbf{x})$ 可写为：

$$p_{est}(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} K\left(\frac{\|\mathbf{x}-\mathbf{x}_n\|_2}{\sigma}\right) \tag{2}$$

其中， K 在我们的实验中假定为高斯核。带宽参数（σ）估计为特征空间中最近邻之间距离的均值。

$$\sigma = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{x_n} - \gamma(\mathbf{x}_n)\|_2^2 \tag{3}$$

其中， $\gamma(\mathbf{x}_n)$ 返回$\mathbf{x}_n$的最近邻。由于带宽是根据数据估计的，因此用于估计特征分布的方法仍为非参数方法。使用公式 2，我们采用相同的核，但使用分别从各自特征空间中获得的不同带宽参数，来估计源域（p_s）和目标域（p_t）数据集的密度。JSD 损失通过以下方式计算：

$$JSD[p_s, p_t] = \frac{1}{2}\big\{KL[p_s, M] + KL[p_t, M]\big\}, M = \frac{p_s + p_t}{2} \tag{4}$$

其中， KL是两个分布之间的 KL 散度。

3 结果与讨论

3.1 实验设置

数据集。 我们使用了组织病理图像腺体分割数据集，以及用于前列腺分割的多中心 MRI 数据集。CRAG [6]和 GlaS [17] 这两个数据集用于结肠组织学数据集中的腺体分割。我们使用了一个来自六个不同中心的多中心 MRI 数据集 [12] ，这些中心的磁场强度（3 特斯拉和 1.5 特斯拉）和厂商各不相同，并采用了不同的源域和目标域配置。这使我们能够测试多源、多目标以及留出目标域设置。

训练设置和超参数。网络针对 5 种不同的训练/验证数据划分进行训练，并报告相应的性能（使用 Dice分数[17]）均值和标准差。训练从头开始，并采用高斯初始化，以考虑可能由用于源域训练和密度估计的训练数据及验证数据所引起的随机性。

对于密度估计，KDE 使用的样本数量设为 20。每隔 5 个训练周期对 KDE 点进行采样，以估计带宽。 λ 设置为 0.01，用于

组织病理学数据集；对于 MRI 数据集，则根据验证集上的性能设置为 0.001。本文所述的所有实验中，学习率、权重衰减、批量大小和训练轮次分别设为 1e-4、1e-4、10 和 1000。所有模型均使用 PyTorch 框架，在 Nvidia A30（24 GB）GPU 上训练。根据验证集上的性能选择分割模型的最优超参数。将训练集中的 20 % 固定为验证集。保存验证损失最低时的最佳模型。仅在超参数优化完成后的训练流程末尾使用测试集。我们将在之后发布相关代码和模型，供开源使用。

用于密度估计的潜在空间我们的实验采用了具有跳跃连接的 U-Net，最多包含 5 次分解。权重使用 PyTorch 默认初始化进行初始化。密度匹配在最深的编码层进行，该层的潜在维度为 1024x8x8，并展开为 65536x1 向量。补充表 5 还报告了采用不同潜在空间的消融实验。

3.2 结果

腺体分割结果。图 2 显示，未进行任何域适应时，性能会显著下降。无论目标数据集的大小如何，我们提出的方法都不会影响源模型的准确率；而其他方法的源数据集性能会随目标数据集大小而变化。此外，与其他方法相比，我们提出的方法在源数据集上取得了更高的准确率。

在目标数据集上，仅使用 3% 或 30% 的目标数据集时，我们提出的方法优于所有其他对比方法，其中采用所提出带宽的最大均值差异表现最为接近。仅当源数据集为 GlaS、目标数据集为 CRAG，且使用 100% 的目标数据集时，采用固定带宽的最大均值差异（MMD-C）优于我们提出的方法。在所有其他情况下，我们提出的方法均优于其他方法。与 GlaS 相比，源数据集为 CRAG 时性能提升更大，这可能归因于样本数量的差异。CRAG 数据集包含更多样本，这可能导致模型更加偏向源数据集。不过，我们提出的方法成功克服了这种偏差，从而带来了更大的性能提升。图 3 展示了定性结果。其他方法难以准确分割腺体轮廓，会将腺体内部的像素误判为背景。相比之下，我们提出的方法能够将腺体内部的所有像素正确分割为相应类别。所有消融研究均在这些数据集上进行（结果见下文）。

多中心前列腺分割结果。对于多中心磁共振成像数据，有 6 个站点提供前列腺分割数据。因此，我们修改了测试方法，设置了一个分布外（留出）域，该域在训练或无监督域适应过程中不会向任何网络展示。这种设置有助于评估所提出的无监督域适应方法是否能够改善模型对未见数据集的预测。我们将这 6 个数据集组合成多个源域、目标域，

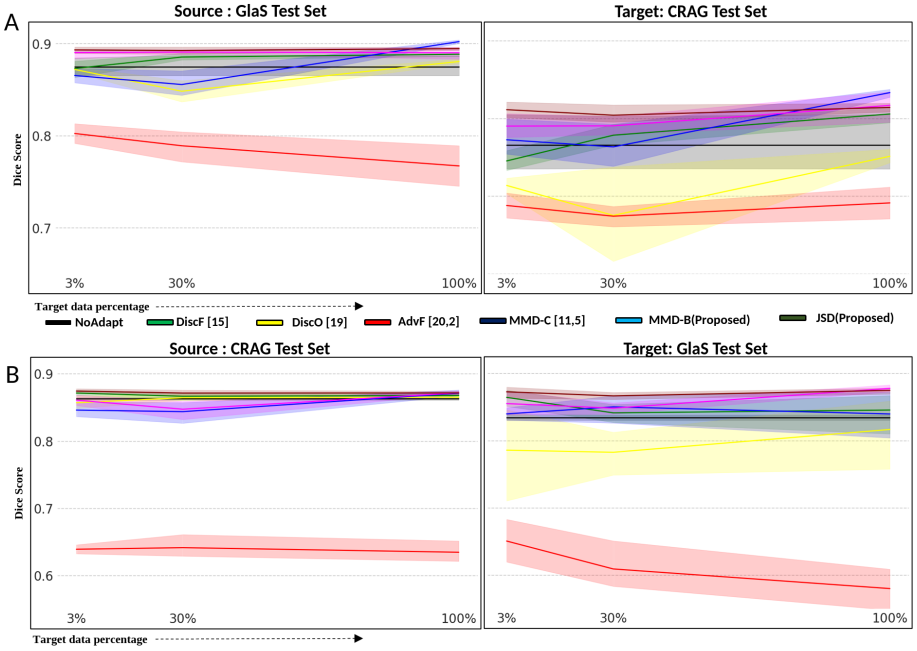

图 2. 腺体分割的无监督域适应结果。A）GlaS 数据集为源域，CRAG 为目标域。B）CRAG 数据集为源域，GlaS 为目标域。使用训练集中的样本进行域适应，测试集仅用于评估训练好的模型。目标数据集规模较小时，无论使用 3% 还是 30% 的数据，所提出的方法都优于其他对比方法。**无适应**指一种常见情形：模型在源数据集上训练，并在目标数据集上测试，期间不进行任何微调。我们将此情形视为基线，因为我们的目标是实现无监督域适应，在这种情况下无法获取目标数据集标注。对于其他基线，此处**Adver** 指 [20,2], **DiscF** 指特征空间判别器 [15], , 而**DiscO** 指输出空间判别器 [19]。**MMD-C** 采用此前提出的方法中使用的固定带宽 [11,5]。**MMD-B** 采用上文第 2.2 节提出的带宽，而**JSD** 则是本文提出的方法。我们可以清楚地看到，在目标数据集规模较小时，所提出方法在目标数据上的平均表现优于其他对比方法；使用全部数据时，其表现也具有竞争力。

以及留出数据集。表 1（补充材料中的表 2、3 和 4）展示了上述不同配置的结果。表 1 展示了单源多目标配置。可以看到，我们提出的方法（MMD-B 和 JSD）不仅在源域和多目标数据集上表现更优，在留出数据集上也是如此。补充材料中的表 2 展示的多源单目标配置也呈现出类似结果。我们提出的方法不仅在源域和目标数据集上保持了准确率，而且在留出数据集上的准确率也始终更高。定性结果见补充材料中的图 4。

特征空间消融。我们尝试使用分割模型的不同特征空间进行特征密度估计。我们观察到

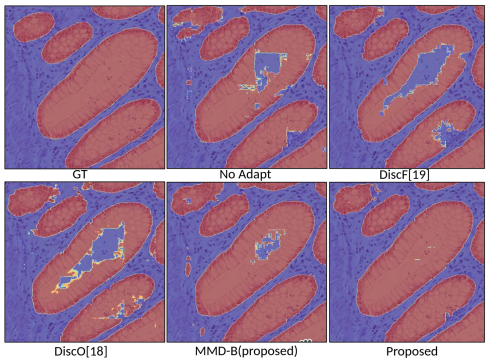

图 3. 腺体分割的定性结果。使用 GlaS 数据集作为源域训练模型时，在目标数据集（CRAG）上取得的结果。与其他技术相比，所提方法在仅使用少量目标数据集时能取得更好的结果。

表 1. 单源多目标配置：源域和目标数据集上的 Dice 分数均值。仅使用目标数据集 3% 的数据进行分布估计时，我们发现，所提出的方法（MMD-B 和 JSD）不仅在源域和多个目标数据集上表现更优，在留出数据集上也同样如此。

	源域	**目标域**				**留出集**
	RUNMC	**BMC**	**I2CVB**	**UCL**	**BIDMC**	**HK**
无适应	0.87	0.79	0.63	0.74	0.53	0.649
Adver [2]	0.717	0.623	0.61	0.564	0.488	0.496
DiscF [15]	0.902	0.795	0.622	0.812	0.547	0.667
DiscO [19]	0.878	0.765	0.604	0.768	0.518	0.626
MMD-C	0.917	0.835	0.63	0.822	0.568	0.692
MMD-B	0.911	0.824	**0.646**	0.822	0.568	0.69
JSD	**0.918**	**0.83**	0.615	**0.829**	**0.58**	**0.711**

源域和目标数据集的测试指标之间的差异在统计学上并不显著。特征空间密度估计采用不同维度时的结果见补充表 5。

带宽估计的频率。将带宽估计频率设为每 1、5、25 和 125 个训练周期一次，目标性能指标均未出现显著变化。每 5 个训练周期估计一次时效果最佳；结果见补充材料中的表6。

使用的核密度估计样本数量。使用不同数量的核密度估计样本进行密度估计的消融实验结果见补充表7。可以清楚地看到，对于组织病理学数据集，使用 20 个核密度估计样本时效果最佳。

4 结论与未来工作

我们提出了一种基于密度匹配和非参数密度估计的无监督域适应技术。我们展示了

所提出方法在组织病理学和多中心磁共振成像这两种模态上的有效性。该技术不仅改善了目标数据集上的结果，还持续提升了源域和留出集结果。评估在多个特征空间中执行密度匹配是否能够帮助模型获得更准确的表征，是未来研究的一个课题。尽管所提出的方法对超参数不敏感，但仍需根据数据集恰当地选择核密度估计点的数量和核带宽。未来工作的另一个方向是使这些超参数从根本上取决于特征空间多样性。

参考文献

1. Aljabri, M., AlAmir, M., AlGhamdi, M., Abdel-Mottaleb, M., Collado-Mesa, F.：深入理解医学影像标注工具：一项综述。Multimedia Tools and Applications **81**(18), 25877–25911 (2022) 2. Bolte, J.A., Kamp, M., Breuer, A., Homoceanu, S., Schlicht, P., Huger, F., Lipinski, D., Fingscheidt, T.：通过无监督域适应提高源域和目标域中的图像分割质量。载于：IEEE/CVF计算机视觉与模式识别会议研讨会论文集。第0–0页 (2019) 3. Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D.：利用生成对抗网络进行无监督像素级域适应。载于：IEEE计算机视觉与模式识别会议论文集。第3722–3731页 (2017) 4. DeGrave, A.J., Janizek, J.D., Lee, S.I.：用于放射影像新冠肺炎检测的人工智能选择捷径特征而非信号。Nature Machine Intelligence **3**(7), 610–619 (2021) 5. Erkent, Ö., Laugier, C.：面向自动驾驶车辆的变化天气条件下基于无监督域适应的语义分割。IEEE Robotics and Automation Letters **5**(2), 3580–3587 (2020) 6. Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., Snead, D., Tsang, Y.W., Rajpoot, N.：Mild-net：用于结肠组织学图像中腺体实例分割的最小信息损失空洞网络。Medical Image Analysis **52**, 199–211 (2019) 7. Haq, M.M., Huang, J.：用于细胞分割的对抗式域适应。载于：基于深度学习的医学影像。第277–287页。PMLR (2020) 8. Hermann, K., Chen, T., Kornblith, S.：卷积神经网络中纹理偏差的起源与普遍性。神经信息处理系统进展**33**, 19000–19015 (2020) 9. Kim, J., Scott, C.D.：鲁棒核密度估计。《机器学习研究杂志》**13**(1), 2529–2565 (2012) 10. Kumagai, A., Iwata, T.：通过单边变换，基于最大均值差异匹配分布实现无监督域适应。载于：AAAI人工智能会议论文集。第33卷，第4106–4113页 (2019) 11. Li, C.L., Chang, W.C., Cheng, Y., Yang, Y., Póczos, B.：MMD GAN：深入理解矩匹配网络。神经信息处理系统进展**30** (2017) 12. Liu, Q., Dou, Q., Yu, L., Heng, P.A.：MS-Net：利用异质磁共振成像数据改进前列腺分割的多站点网络。IEEE医学影像汇刊**39**(9), 2713–2724 (2020)

13. Liu, X., Yoo, C., Xing, F., Oh, H., El Fakhri, G., Kang, J.W., Woo, J., 等：深度无监督域适应：近期进展与展望综述。APSIPA 信号与信息处理汇刊 **11**(1) (2022) 14. Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., Zhang, X., Jin, Y., Zhou, H.：基于深度学习的脑肿瘤分割：综述。复杂与智能系统 **9**(1)，1001–1026 (2023) 15. Long, M., Cao, Y., Wang, J., Jordan, M.：利用深度适应网络学习可迁移特征。载于：国际机器学习会议。第 97–105 页。PMLR (2015) 16. Saha, S., Elhabian, S., Whitaker, R.：Gens：生成式编码网络。机器学习 **111**(11)，4003–4038 (2022) 17. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., 等：结肠组织学图像中的腺体分割：GlaS 挑战赛。医学图像分析 **35**，489–502 (2017) 18. Toldo, M., Maracani, A., Michieli, U., Zanuttigh, P.：语义分割中的无监督域适应：综述。技术 **8**(2)，35 (2020) 19. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.：学习适应语义分割的结构化输出空间。载于：IEEE 计算机视觉与模式识别会议论文集。第 7472–7481 页 (2018) 20. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.：对抗性判别域适应。载于：IEEE 计算机视觉与模式识别会议论文集。第 7167–7176 页 (2017) 21. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.：Advent：语义分割中用于域适应的对抗性熵最小化。载于：IEEE/CVF 计算机视觉与模式识别会议论文集。第 2517–2526 页 (2019) 22. Wang, H., Ge, S., Lipton, Z., Xing, E.P.：通过惩罚局部预测能力学习鲁棒的全局表示。神经信息处理系统进展**32** (2019) 23. Wang, Y.C., Hsieh, T.C., Yu, C.Y., Yen, K.Y., Chen, S.W., Yang, S.N., Chien, C.R., Hsu, S.M., Pan, T., Kao, C.H., 等：4d 18f-fdg pet/ct 在食管鳞状细胞癌放疗计划大体肿瘤体积勾画中的临床应用。放射研究杂志 **53**(4)，594–600 (2012) 24. Wang, Z., Wei, Y., Feris, R., Xiong, J., Hwu, W.M., Huang, T.S., Shi, H.：缓解语义层级偏移：一种用于语义分割的半监督域适应方法。载于：IEEE/CVF 计算机视觉与模式识别会议研讨会论文集。第 936–937 页 (2020) 25. Węglarczyk, S.：核密度估计及其应用。载于：ITM 会议网络论文集。第 23 卷，第 00037 页。EDP Sciences (2018) 26. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.：基于傅里叶的域泛化框架。载于：IEEE/CVF 计算机视觉与模式识别会议论文集。第 14383–14392 页 (2021) 27. Yang, Y., Soatto, S.：Fda：用于语义分割的傅里叶域适应。载于：IEEE/CVF 计算机视觉与模式识别会议论文集。第 4085–4095 页 (2020)

5 补充材料

表 2. 多中心源域和单目标域配置。 仅使用目标数据集的 3% 时报告平均 Dice 分数。在多源单目标配置中，我们提出的方法在目标数据集和留出数据集上的表现优于其他方法（尤其是在*HK*上）。

	源域			**目标域**	**留出集**	
	RUNMC	**I2CVB**	**BIDMC**	**UCL**	**BMC**	**HK**
无适应	0.86	0.87	0.70	0.79	0.83	0.70
Adver[2]	0.64	0.73	0.57	0.545	0.611	0.516
DiscF [15]	0.9	0.91	0.71	0.82	0.86	0.69
DiscO[19]	0.87	0.90	0.71	0.80	0.84	0.66
MMD-C	0.91	**0.927**	0.747	0.854	0.876	0.722
MMD-B	0.912	0.925	**0.778**	**0.873**	0.876	0.737
JSD	**0.916**	0.926	0.751	0.867	**0.88**	**0.751**

表3. 多源多目标配置。仅使用目标数据集的3%时，报告平均Dice分数。对于此配置，采用所提出的带宽的最大均值差异（MMD）与所提出的密度匹配方法的性能相当，但两者均优于固定带宽和其他方法。

	源域		**目标域**		**留出集**	
	RUNMC	**I2CVB**	**BMC**	**BIDMC**	**UCL**	**HK**
无适应	0.85	0.89	0.79	0.50	0.66	0.51
Adver[2]	0.645	0.757	0.604	0.49	0.551	0.496
DiscF[15]	0.903	0.909	0.846	0.549	0.841	0.601
DiscO [19]	0.847	0.889	0.763	0.514	0.741	0.512
MMD-C	0.909	0.908	0.85	**0.564**	**0.85**	**0.625**
MMD-B	**0.913**	0.91	**0.859**	0.561	0.838	0.597
JSD	**0.913**	**0.913**	0.858	0.559	0.836	0.607

表4. 第二种多源多目标配置。 对于此配置，使用所提出带宽的最大均值差异优于所有其他方法。该结果表明，选择数据依赖带宽（即最近邻距离的平均值）有助于域适应。

	源域		**目标域**	**留出集**		
	HK	**UCL**	**RUNMC**	**BMC**	**BIDMC**	**I2CVB**
无适应	0.87	0.84	0.78	0.77	0.56	0.64
Adver[2]	0.749	0.675	0.681	0.647	0.617	0.514
DiscF [15]	0.876	0.869	0.803	0.787	0.66	0.554
DiscO[19]	0.81	0.77	0.74	0.71	0.53	0.63
MMD-C	0.842	0.841	0.772	0.748	0.65	0.552
MMD-B	0.887	**0.889**	**0.814**	**0.817**	**0.678**	**0.579**
JSD	**0.888**	0.886	0.813	0.813	0.667	0.561

表 5. 特征空间消融。 不同编码器和解码器特征空间下密度匹配性能的均值和标准差。可以注意到，不同特征空间下的性能指标在统计学上并无显著差异。

	CRAG 作为源数据集		
特征空间	**特征维度**	**源域**	**目标域**
最深层	1024x8x8	0.87 ±0.003	0.876 ±0.012
ENC1	128x128x128	0.871 ±0.005	0.877 ±0.015
ENC2	256x64x64	0.874 ±0.001	0.876 ±0.009
ENC3	512x32x32	0.872 ±0.003	0.874 ±0.009
ENC4	1024x16x16	0.871 ±0.001	**0.877 ±0.004**
DEC4	1024x16x16	0.869 ±0.014	0.874 ±0.009
DEC3	512x32x32	0.870 ±0.007	0.876 ±0.011
DEC2	256x64x64	**0.875 ±0.003**	0.875 ±0.013
DEC1	128x128x128	0.866 ±0.013	0.868 ±0.016

表6. 带宽估计频率消融实验。不同带宽估计频率/训练轮次下测试指标的平均Dice分数和标准差。我们发现，对于两个数据集，如果每5个训练周期估计一次带宽，结果最优。

	GlaS 作为源域		CRAG 作为源域	
带宽频率	**源域**	**目标域**	**源域**	**目标域**
5 个训练轮次	**0.895 ±0.003**	0.815 ±0.012	0.87 ±0.004	**0.879 ±0.004**
1 个训练周期	0.894 ±0.007	0.816 ±0.017	**0.874 ±0.001**	0.877 ±0.003
25 个训练周期	0.894 ±0.004	0.805 ±0.023	**0.874 ±0.004**	0.875 ±0.004
125	0.894 ± 0.002	**0.821 ± 0.013**	0.866 ± 0.003	0.863 ± 0.043

表7. KDE样本数量消融实验。报告了用于估计特征空间密度的不同KDE样本数量对应的平均Dice分数和标准差。20个KDE样本最适合用于腺体分割域适应。

	以GlaS为源域		以CRAG为源域	
核密度估计样本数	**源域**	**目标域**	**源域**	**目标域**
20	**0.895 ±0.003**	**0.81 ±0.012**	**0.87 ±0.004**	**0.879 ±0.005**
10	0.874 ±0.014	0.784 ±0.036	0.865 ±0.008	0.828 ±0.05
20	0.875 ±0.018	0.78 ±0.036	0.851 ±0.015	0.843 ±0.026
80	0.868 ±0.018	0.019 ±0.057	0.86 ±0.008	0.861 ±0.001

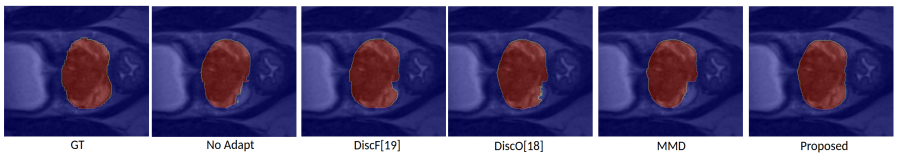

图 4. 磁共振成像队列的定性结果。与其他方法相比，我们提出的方法生成了高质量的结果，且轮廓准确。.