

基于类别平衡与伪标签去噪的无源域适应方法

王墨涵¹
WANG Mohan

摘要

医学眼底图像分割在眼科疾病诊断和治疗中至关重要，但面临域转移问题，导致预训练好的模型在不同临床中心的数据集上表现较差。传统的无监督领域适应（UDA）方法通过利用有标签的源域数据和无标签的目标域数据进行训练，但在医学影像领域存在严格的隐私保护问题，源数据的访问存在限制。为此，无源领域自适应（SFDA）方法应运而生，通过仅使用无标签的目标域数据适应预训练模型。然而，现有 SFDA 方法仍面临类别不平衡和伪标签存在噪声等挑战，限制了模型的适应性能。文章提出了一种类别平衡的教师-学生联合训练方案，通过动态加权不同类别的损失函数，用于解决类别不平衡问题。此外，通过估计伪标签的可靠性，降低伪标签中含有的噪声。以 REFUGE 数据集为源域、RIM-ONE-r3 和 Drishti-GS 数据集为目标域进行实验。实验结果显示，所提方法在 RIM-ONE-r3 数据集的 Dice 系数达到了 93.15%，在 Drishti-GS 数据集的 Dice 系数达到了 96.59%，显著提升了模型在医学图像分割任务中的适应性能。

关键词

无源领域自适应；眼底图像分割；教师-学生模型；类别平衡；伪标签去噪

doi: 10.3969/j.issn.1672-9528.2025.07.007

0 引言

医学眼底图像分割在眼科疾病的诊断与治疗中发挥着至关重要的作用，尤其是在视网膜疾病的早期发现和干预中。随着深度学习技术的快速发展，深度神经网络（DNN）在医学影像分割领域取得了显著的进展。这些模型能够自动处理大量的医学图像数据，从而提供快速、精确的分割结果，极大地提高了诊断效率和准确性。在实际应用中，深度学习模型常常面临着一个挑战——域转移问题^[1]，即在一个领域训练的模型，往往在另一个领域的数据集上表现较差。这一问题导致了模型在不同临床中心的数据集上无法实现良好的迁移。

为了解决域转移问题，传统的无监督领域适应方法（UDA）应运而生。UDA 通过利用源域的有标注数据和目标域是无标注数据，在没有目标域标签的情况下进行模型训练，从而实现跨域迁移。然而，大多数 UDA 方法假设能够访问源域数据，这在实际应用中存在隐私保护的挑战。特别是在医学影像领域，由于患者隐私的敏感性，源域数据的访问和共享常常受到严格的保护。因此，UDA 方法在临床环境中的应用受到很大限制，迫切需要一种不依赖于源数据的适应方法。

为了克服这些问题，无源领域自适应（SFDA）方法引起了广泛的关注。SFDA 方法通过自我训练机制，在没有源域数据的情况下实现域适应^[2]。首先，利用已预训练的源模型生成目标域数据的伪标签，之后用这些伪标签对目标模型进行训练。通过不断迭代更新，目标模型逐渐优化，使其适应目标域的分布。该过程减少了源域与目标域之间的分布差异，从而实现了源数据的知识迁移，提升了模型在目标域上的性能。SFDA 不仅避免了源数据的隐私问题，还能在没有源数据的情况下实现领域适应，满足了实际应用的需求。SFDA 方法和 UDA 方法的对比如图 1 所示。

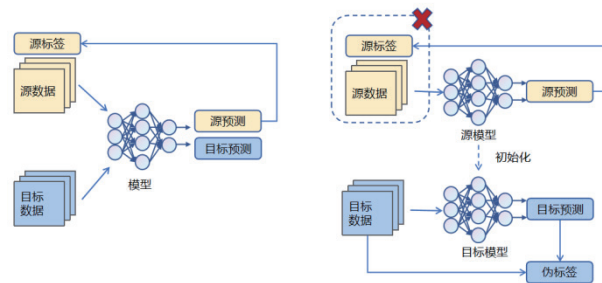


图 1 SFDA 方法和 UDA 方法的对比

当前 SFDA 方法中，Liu 等人^[3]利用源模型的批量归一化统计量来恢复类似源域的样本。SHOT^[4]则通过固定已训练的分类器，并最大化目标输出的互信息来实现分布对齐。

1. 南京理工大学计算机科学与工程学院 江苏南京 210000

然而，当前的 SFDA 方法仍然面临一些挑战，尤其是在类别平衡和伪标签去噪方面。由于域偏移的存在，伪标签往往带有噪声，严重影响了模型的适应效果。此外，眼底图像存在严重的类别不平衡问题，视杯区域明显小于视盘区域，导致在适应过程中少数类别遗忘。

为了解决这些问题，本文提出了一个类别平衡的教师-学生联合训练方案，结合去噪策略，提高了模型的准确性和鲁棒性。具体来说，首先通过教师模型生成伪标签，利用弱增强数据进行训练，同时学生模型利用强增强数据与这些伪标签进行优化。动态加权不同类别的损失函数，有效解决了类别不平衡问题，确保了背景和前景类别在训练中的公平性。接着，采用像素级去噪策略和原型去噪策略，通过对伪标签进行优化，减少了噪声对训练过程的干扰。像素级去噪通过不确定性估计来加权每个像素的置信度，而原型去噪则通过计算像素特征与类别原型的距离来调整噪声的影响。实验结果表明，本文提出的方法在多个医学影像数据集上的效果良好，表现出更好的鲁棒性和稳定性。

1 方法

1.1 综述

本方法提出了一种结合类别平衡的教师-学生联合训练方案和去噪策略的无源域自适应 (SFDA) 框架。首先，对目标域数据进行弱增强和强增强处理，分别将其输入教师模型和学生模型。教师模型通过弱增强数据进行训练，并利用不确定性估计生成伪标签，学生模型则在强增强数据上利用这些伪标签进行优化。通过引入类别平衡策略，以保证前景和背景类别的公平学习，避免背景类别对训练过程的主导作用。为提升模型在目标域的适应性，通过像素级去噪策略，利用不确定性估计计算每个像素的置信度，并降低高不确定性像素在损失计算中的权重，以减少错误伪标签的影响。此外，还利用原型去噪策略通过计算像素特征与类别原型的距离，降低那些特征与原型距离较远的像素在损失计算中的权重，从而进一步减少噪声伪标签的影响。最终，本文方法有效地提高了无源域自适应任务中的模型性能。如图 2 所示。

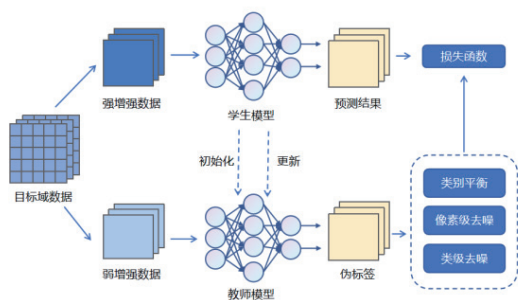


图 2 本文提出的 SFDA 方法框架

1.2 类别平衡的教师-学生联合训练方案

1.2.1 教师-学生联合训练方案

为了避免源域和目标域之间的差异并实现训练过程的稳定性，引入了一个教师-学生联合训练方案。在该方案中，教师模型 h_t 和学生模型 h_s 被共同使用。并进一步提出了弱增强策略，将增强过的图像分别输入到教师模型和学生模型中。

具体来说，对于每个目标图像 X_i ，生成其对应的弱增强版本 X_i^w ，并将其输入教师模型进行处理。对于每个目标图像 X_i ，生成强增强版本 X_i^s ，并将其输入到学生模型。教师模型 $h_t(x)$ 生成伪标签 $\hat{y}^w$ ，并使用公式 $\hat{y}^w = h_t(x)$ 进行计算。后使用学生模型 h_s 进行训练，并通过计算损失来优化模型。通过反向传播来更新学生模型的权重：

$$\hat{\theta} \leftarrow \lambda \hat{\theta} + (1 - \lambda) \theta \quad (1)$$

式中： $\hat{\theta}$ 和 θ 分别表示教师和学生模型的权重。并非直接使用梯度更新教师模型，而是采用教师模型的指数移动平均 (EMA) 来更新模型，从而使得教师模型在适应过程中更加稳定。

1.2.2 类别平衡

在处理眼底图像时，属于前景区域的像素通常较少，大部分像素属于背景区域，因此背景类别往往会主导损失，从而导致模型过度关注背景，忽略前景信号。为解决这一问题，提出了类别平衡的损失校准方法，分析所有图像的类别不平衡，并用这些全局知识来校准每个图像的损失。

具体来说，通过存储所有图像的像素预测，并维护前景平均损失 r_{fg}^{nk} 和背景平均损失 r_{bg}^{nk} ，用公式表示为：

$$r_{fg}^{nk} = \frac{\sum_i L_i, k \cdot \mathbb{1}[\hat{y}_i, k = 1]}{\sum_i \mathbb{1}[\hat{y}_i, k = 1]} \quad (2)$$

$$r_{bg}^{nk} = \frac{\sum_i L_i, k \cdot \mathbb{1}[\hat{y}_i, k = 0]}{\sum_i \mathbb{1}[\hat{y}_i, k = 0]} \quad (3)$$

式中： L 表示损失函数；fg 和 bg 分别表示前景和背景类别。通过前景损失和背景损失的比率来调整损失函数：

$$\widehat{L}_{bce} = E_{x \sim X_i, \hat{y}_i \sim \hat{y}_i} [\widehat{y}_k \log(p_k) + r_{fg}^{nk} \cdot \widehat{y}_k \log(1 - p_k)] \quad (4)$$

这种方法通过对前景和背景类别进行公平加权，能够有效地减少背景类别对模型训练过程的主导作用，确保模型在处理具有较强背景噪声的图像时仍然能够专注于前景区域，解决了类别不平衡问题。

1.3 伪标签去噪

1.3.1 不确定性估计的像素级去噪策略

通过蒙特卡洛丢弃来估算不确定性^[5]，以便进行贝叶斯近似。具体来说，给定目标图像 x_t^i ，在推理过程中启用丢弃，并通过源模型执行 K 次随机前向传播，获取不确定性图 $u \in \mathbb{R}^{H \times W \times C}$ 。

为利用不确定性图来过滤伪标签，定义了一个标签选择掩码 $m_v \in \{0,1\}^{H \times W \times C}$ ，其在像素 v 处的值由不确定性度量 u_v 决定，具体计算公式为：

$$m_v = 1[u_v < \eta] \quad (5)$$

式中： η 为不确定性阈值。

如不确定性较高，则认为该伪标签为噪声，并将其排除在损失计算之外。通过这种方式，可以在训练过程中减少部分不确定性较高的标签对模型的负面影响，筛选出更加可靠的伪标签，减少伪标签的噪声。

1.3.2 原型估计的类级去噪策略

本文进一步提出了一种新的类别级去噪方法，利用原型为识别潜在的伪标签提供指导^[6]。该方法假设对于一个可靠的伪标签，其特征应该接近相应类别的原型。背景原型 b^{bg} 和前景原型 b^{obj} 可以从 $\hat{y}^t$ 中获得，其中 $b^{obj} = 1[\hat{y}^t = 1] \cdot 1[u < \eta]$ 和 $b^{bg} = 1[\hat{y}^t = 0] \cdot 1[u < \eta]$ 。然后，类别级原型 z^{obj} 和 z^{bg} 通过计算每个像素的特征与原型之间的相对特征距离 d_v^{obj} 和 d_v^{bg} 得出，公式为：

$$z^{obj} = \frac{\sum_v e_v \cdot b_v^{obj} \cdot p_v}{\sum_v b_v^{obj} \cdot p_v} \quad (6)$$

$$z^{bg} = \frac{\sum_v e_v \cdot b_v^{bg} \cdot (1 - p_v)}{\sum_v b_v^{bg} \cdot (1 - p_v)} \quad (7)$$

式中： e_v 表示每个像素的特征； d_v^{obj} 和 d_v^{bg} 分别表示前景和背景类别的标签。通过计算每个像素与类别原型的相对距离，来决定该像素是否为该类别的伪标签。

通过结合不确定性估计和原型估计，定义了一个类别标签选择掩码 $m \in \{0,1\}^{H \times W \times C}$ ，其值根据不确定性度量 u_v 来决定：

$$m_v = 1[u_v < \eta] \cdot 1[\hat{y}_v^t = 1] \cdot [d_v^{obj} < d_v^{bg}] + 1[u_v < \eta] \cdot 1[\hat{y}_v^t = 0] \cdot [d_v^{bg} < d_v^{obj}] \quad (8)$$

式中： η 是不确定性阈值。如果不确定性较低，将选择相应的伪标签 $\hat{y}_v^t$ 。否则，该伪标签将被视为噪声，降低在损失计算中的权重。

最终，通过将类别平衡结合不确定性估计和原型估计来筛选伪标签，并进行更新，以使得伪标签的可靠性更高。模型的损失函数则通过公式进行更新：

$$L_{seg} = \sum_v m_v \cdot \bar{L}_{bce} \quad (9)$$

2 应用实验

2.1 数据集和评价指标

本文在视网膜眼底图像的视盘和视杯分割任务中验证方法的可用性，所使用的数据集具有分布偏移的特征。具体地，选择 REFUGE 挑战赛的训练数据集作为源域，选择 RIM-ONE-r3 和 Drishti-GS 数据集作为目标域^[7]。评价指标有

2 个，分别是 Dice 系数和平均表面距离（ASD）。Dice 系数用于评估像素级准确度，ASD 用于衡量边界一致性。较高的 Dice 系数和较低的 ASD 值表明模型的表现更为优越。

2.2 对比实验

在对比实验中，分别在两个目标数据集上评估了视盘和视杯分割任务的性能。实验结果如表 1 和表 2 所示。

表 1 在 REFUGE $\rightarrow$ RIM-ONE_r3 设定下的实验结果

Method	Optic disc segmentation		Optic cup segmentation	
	Dice/%	ASD/px	Dice/%	ASD/px
w/o adaptation	86.76	10.91	64.31	10.08
Tent ^[8]	88.36	10.90	72.65	12.29
TTSFDA ^[9]	88.29	11.00	77.52	7.91
DPL ^[10]	88.94	10.52	78.09	8.32
Ours	93.15	6.18	79.61	7.75

表 2 在 REFUGE $\rightarrow$ Drishti-GS 设定下的实验结果

Method	Optic disc segmentation		Optic cup segmentation	
	Dice/%	ASD/px	Dice/%	ASD/px
w/o adaptation	96.24	4.28	77.80	14.76
Tent ^[8]	95.83	4.76	76.02	16.21
TTSFDA ^[9]	96.09	4.48	78.88	14.11
DPL ^[10]	96.07	4.47	79.82	13.68
Ours	96.59	3.85	80.99	12.82

实验结果表明，本文方法在多个方面表现出色，在没有依赖源域数据的设置下，展现了优异的适应能力和分割精度。将本文方法与近期的领域适应方法进行比较，包括 Tent^[8]：通过最大化源域和目标域之间的不确定性差异来实现自适应；TTSFDA^[9]：通过学习目标域的表征并引导源域知识迁移；DPL^[10]：通过动态伪标签生成来进行领域适应的方法。在相同的网络结构和数据划分设置下，参考这些方法在其论文中报告的结果进行比较。如图 3 所示，本文方法在分割效果上优于其他对比方法，表现更为精确。

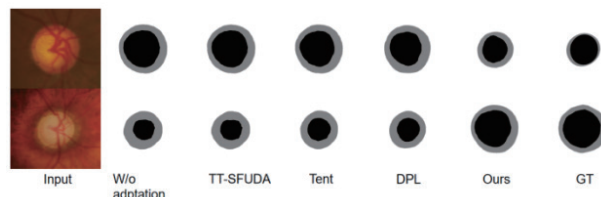


图 3 眼底图像分割任务中的表现对比

尽管 Tent、TTSFDA 和 DPL 方法在视盘分割任务上表现出不同程度的改进,但本文方法在视盘分割的 Dice 系数和 ASD 上均优于其他方法,显示了明显的优势。例如,本文方法在 REFUGE->RIM-ONE_r3 和 REFUGE->Drishti-GS 两种实验设定下,较没有适应的情况 Dice 系数分别提高了 6.39 和 0.35 个百分点,ASD 分别降低了 4.72 和 0.43 px。本文方法比 Tent、TTSFDA 和 DPL 方法都展现了一定的优势。此外,在视杯分割任务中也取得了良好结果,均领先于其他领域适应方法,说明本文方法减少了类别不平衡。

总的来说,本文方法在两个目标域的数据集上都展现了最优的性能,能够在视盘和视杯分割任务中提供准确的分割结果。与其他依赖源域数据的领域适应方法相比,不仅提高了分割精度,还增强了跨域适应的能力,表明其在医学图像分割任务中的广泛应用潜力。

2.3 消融实验

在消融实验中,评估了不同设置对模型性能的影响,重点考察了去噪技术和教师-学生架构对视盘和视杯分割任务的贡献。实验结果如表 3 所示。

表 3 消融实验视盘和视杯分割任务实验结果

Setup	Optic disc segmentation		Optic cup segmentation	
	Dice/%	ASD/px	Dice/%	ASD/px
baseline	89.22	10.03	76.43	7.99
w/o Denoise	89.97	9.30	77.81	7.79
w/o Teacher-student	91.53	7.75	78.93	7.82
Ours	93.15	6.18	79.61	7.75

实验结果表明,基线方法提供了基础性能,加入去噪技术后,视盘分割的 Dice 系数提升了 2.31 个百分点,ASD 降低了 2.28 px,表明去噪技术在改进模型的分割精度和边界一致性方面起到了重要作用。加入教师-学生架构,模型表现出更好的性能,尤其是在视盘分割任务中,Dice 系数提高到了 89.97%,表明教师-学生架构也具有重要贡献。最终,结合去噪技术和教师-学生架构的方法在两个分割任务中均展现出最佳性能,视盘分割的 Dice 系数为 93.15%,ASD 降至 6.18 px,明显优于其他设置。实验结果表明这两项技术的有效结合显著提升了模型在无源领域适应任务中的适应性和精度,验证了本文方法的优势,并提供了未来无源域自适应研究的重要参考。

3 结语

本文针对医学眼底图像分割中的无源域自适应(SFDA)问题提出了一种创新的方法,仅通过目标域无标签数据进行模型的适应,解决领域转移带来的挑战。提出的类别平衡的

教师-学生联合训练方案有效解决了类别不平衡的问题,同时结合像素级去噪和原型去噪策略,进一步降低了伪标签中的噪声,确保了更可靠的伪标签在训练中的使用。实验结果表明,所提方法在多个医学影像数据集上取得了显著的性能提升,模型在不使用源域数据的情况下依然能够达到较高的准确度,充分展示了该方法在实际应用中的潜力。

参考文献:

[1] 刘二虎. 基于跨域知识多样性的无监督域自适应方法研究[D]. 淮南:安徽理工大学,2024.

[2] 孟凡安. 源域受限条件下的无监督领域自适应方法研究[D]. 济南:山东财经大学,2024.

[3] LIU Y, ZHANG W, WANG J. Source-free domain adaptation for semantic segmentation[EB/OL]. (2021-03-30)[2024-12-09].<https://doi.org/10.48550/arXiv.2103.16372>.

[4] 赵凯琳,靳小龙,王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2):349-369.

[5] 赵东旭. 基于不确定性学习的表情识别算法研究[D]. 长春: 长春工业大学,2024.

[6] 邓皓文,王恒升. 基于对比学习与去噪扩散模型的薄膜表面瑕疵图像分类[J/OL]. 计算机工程与应用, 1-13[2025-02-20].<http://kns.cnki.net/kcms/detail/11.2127.tp.20240924.1755.014.html>.

[7] 李豪杰,郭子洋,刘伟志,等. 基于 ReSE-AG-Net 的青光眼视盘视杯分割方法[J]. 长春理工大学学报(自然科学版), 2024, 47(4):115-121.

[8] WANG D Q, SHELHAMER E, LIU S T, et al. Tent: fully test-time adaptation by entropy minimization[EB/OL].(2021-03-18)[2024-12-23].<https://doi.org/10.48550/arXiv.2006.10726>.

[9] VS V, VALANARASU J M J, PATEL V M. Target and task specific source-free domain adaptive image segmentation[EB/OL].(2023-03-10)[2025-01-12].<https://doi.org/10.48550/arXiv.2203.15792>.

[10] ZHANG Y, MIAO S, MANSI T, et al.Task driven generative modeling for unsupervised domain adaptation: application to X-ray image segmentation[C]//Medical Image Computing and Computer Assisted Interventio-MICCAI 2018. Berlin: Springer, 2018:599-607.

【作者简介】

王墨涵(2004—),女,山东青岛人,本科在读,研究方向:计算机科学与技术。

(收稿日期: 2025-02-23 修回日期: 2025-07-01)