

# 无源域自适应眼底图像的上下文感知伪标签校正分割

王少龙

## 摘要

在域迁移问题中，因为商业机密，隐私或知识产权问题，源域数据可能不会对目标域客户端进行开放。因此，产生出无源无监督域自适应方法，旨在不接触源域数据的情况下，仅微调目标域模型来对齐源模型和无标签的目标域分布。本文是对原论文的一次复现。原论文通过一种上下文感知的SF-UDA伪标签细化框架来提升模型性能，即提出了一个新的上下文相似性模块学习上下文关系，利用学到的上下文知识来校正伪标签。之后进一步校准修正后的伪标签，以补偿由不准确的上下文关系引起的错误修正。此外，采用一种像素级和类级去噪方案，用于选择可靠的伪标签进行域迁移的工作。我在原论文的基础上，针对简洁的上下文监督模块增加了时空注意力机制，以获得更好的效果。实验结果表明，改进后的算法有更好的域迁移效果。

**关键词：**域迁移；无源无监督学习；医学图像分割；伪标签校正

## 1 引言

医学图像处理是许多的临床诊断和医学研究的关键步骤，其中包括疾病的诊断，治疗方案的规划以及人群研究。现有的深度学习神经网络分割模型，在大量有标记的全监督数据上进行训练时，可以获得不错的分割性能。但是在医学图像领域，全监督训练往往不切实际，由于需要具有丰富的临床知识的医生进行标注，且会耗费大量的金钱和时间成本，获取大量全监督数据来进行网络训练很难进行。而且图像采集过程的差异也进一步加剧了这一问题，比如不同医院、不同设备，采集到的数据，会导致数据集间的分布出现偏移，最终使得网络模型应用在分布不同的数据集时性能大幅下降。

为了解决数据间的偏移和全监督数据的稀缺问题。无监督域迁移学习（UDA）的方法被提出研究。然而，在研究过程中，一些数据受制于医疗数据间病人的隐私或者医院间的知识产权问题，源域数据不会对研究者开放，仅提供使用源域数据训练好的分割模型。因此无源无监督领域自适应（SF-UDA）已成为一个重要的研究领域，其中源域数据无法访问,仅通过访问源域模型和目标数据，同时不需要在目标域中进行任何注释,来将知识传递到目标域。

眼底图像中视杯和视盘的准确分割对于测量视杯与视盘比率至关重要，而这对于青光眼筛查和检测至关重要。医院获取的眼底数据同样面临着上述的问题。而在SF-UDA研究中，现有的方法都忽略了图像的上下文关系，这可以在不需要访问源数据的情况下提高适应性。通过实验，我们注意到域间隙可能导致源模型做出上下文不一致的预测。对于具有相似视觉外观的图像的相邻块，源模型可以产生截然不同的预测。这是因为由于域偏移，目标数据在特征空间中移动，导致一些数据点移动到源域分割器的边界上。上下文不一致的问题促

使我们在提炼伪标签时利用上下文关系。此外，在我们还注意到（如图1所示），源模型产生的目标特征仍然形成聚类，这意味着具有相同类别的目标数据点的特征位置相近。这一发现使我们能够根据特征距离计算上下文关系。

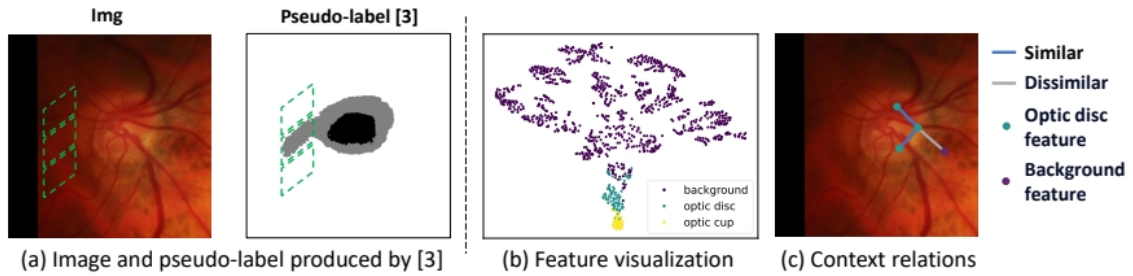


图 1. (a) 上下文不一致的伪标签示例。由于域间隙的原因，视盘伪标记有不规则的突起，这与相邻的预测不一致。(b) 源模型生成的目标像素特征的 t-SNE可视化。在域偏移下，尽管没有与源域分割器对齐，但同一类的目标特征仍然形成聚类。(c) 受(b)的启发，可以从特征距离计算上下文关系。

在本文中，我们提出了一种新的上下文感知伪标签精化（CPR）框架，用于无源无监督领域自适应。首先，我们提出了一个上下文相似性学习模块，通过上下文相似性头从特征的距离计算上下文关系。这利用了域偏移下的内在跨域聚类特征分布仍然相同的特性，其中源域编码器生成的目标特征对于同一类是接近的，而对于不同类是遥远的（图第1（b））。接着，利用上面过程学习到的上下文关系，利用相邻的伪标签来修正坏的伪标签，来进行上下文感知修正。此外，还提出了一种伪标签的校准策略，旨在缓解不准确的学习上下文关系带来的负面影响。最后，考虑到模型的特征分布，对细化的伪标签进行进一步去噪，以选择可靠的伪标签用于域自适应训练。通过将模型应用到跨域眼底图像分割实验，同最先进的SF-UDA方法进行比较，得出我们的方法有着更好的域迁移分割性能。

## 2 相关工作

### 2.1 无源无监督领域自适应

无源无监督领域自适应（SF-UDA）是一种新的研究方法，旨在在源领域没有标注数据的情况下，将模型从源领域迁移到目标领域。SF-UDA的基本思想是在源域和目标域之间建立一个共享的模型，并通过在目标域上进行特征对齐和领域适应来实现领域迁移。现有的SF-UDA解决方案可分为四大类：批量归一化（BN）统计自适应、近似源信息、熵最小化和伪标签。BN统计自适应方法旨在解决不同领域之间的统计差异。例如，[6, 7]使用不同的训练对象更新低阶和高阶BN统计，而[8]调整BN统计以最小化模型预测的熵。近似源图像旨在生成类源图像。例如，[9]首先通过冻结源模型和训练可学习图像来获得粗略的源图像，然后通过相互傅立叶变换来细化图像。精细的类源图像提供了源数据分布的表示，并有助于在自适应过程中进行域对准。例如，[3]学习域提示以添加到目标域图像，从而使总和模拟源图像。熵最小化方法的目的是产生更有信心的模型预测。例如，[1]使用类比率的正则化子来最小化输出熵。通过在源域上预先训练的辅助网络来估计类比率。对于伪标签，错误的伪标签

要么被丢弃，要么被纠正。例如，[2]在像素级别和类别上识别低置信度伪标签。另一方面，[25]通过估计类条件标签错误概率来执行不确定性加权伪标签校正。

## 2.2 医学图像分割

现有的医学图像分割网络大多基于UNet的网络架构，即拥有一个编码器和解码器，在编码器部分通过一系列卷积和下采样捕获图像的特征，在解码器过程通过不断的上采样和卷积逐步恢复图像信息。在同一层的编码器层和解码器层之间，通过跳跃连接将编码器部分提取的特征传递给解码器部分，以帮助恢复图像的细节信息。

Li等人[5]提出了一种基于UNet的IerNet，它采用了迷你UNet的多次迭代。在每次迭代中，使用短跳连接和长跳连接共享前面步骤中的功能。在注意力引导网络（AGNet）[11]中，提出了一种基于M-Net和注意力引导滤波器的方法，将结构信息从低级特征图转移到高级特征图。姜等人[4]提出了用于视网膜图像分割的多尺度多分支网络（MSMB-Net），利用逆卷积和跳跃连接进行更有效的特征提取。张等人[10]利用具有金字塔级聚合块的金字塔UNet进行各级特征聚合。本文所使用的骨架网络DeepLab v3+则是仿照U-Net的结构添加了一个向上采样的解码器模块，用来优化边缘的精度。

## 3 本文方法

### 3.1 本文方法概述

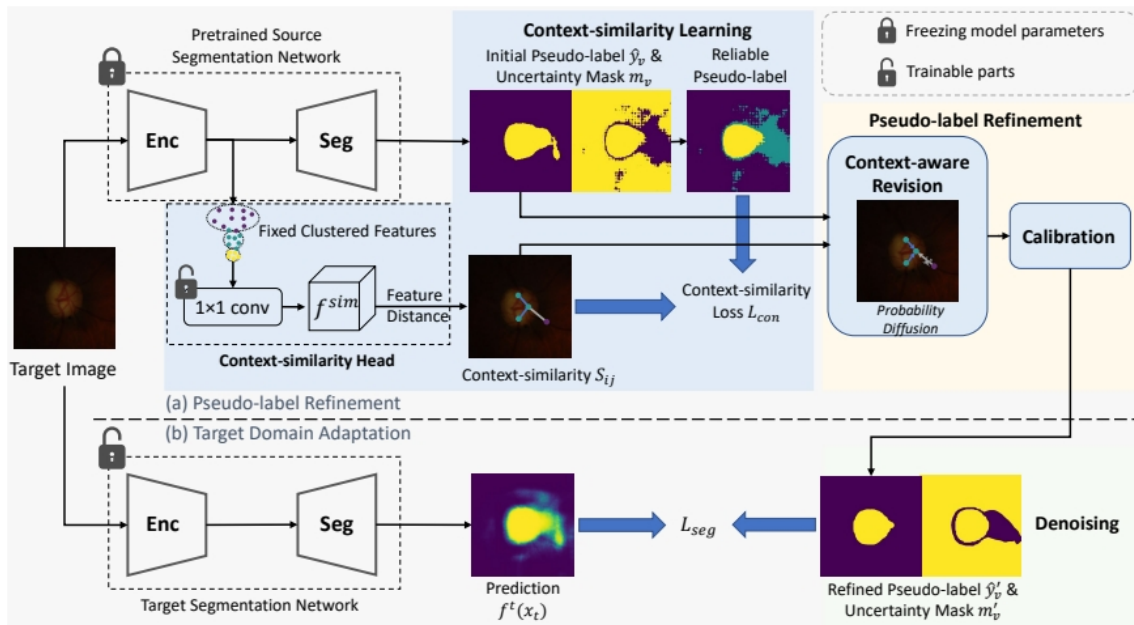


图 2. 网络架构

模型整体架构如图2所示。伪标签提纯阶段。将源域模型冻结，让目标域数据通过源域网络提取底层特征进行聚类，进行特征间相似性计算求出特征距离，通过聚类中心和特征距离的判断，校正伪标签，同时与直接使用源模型进行分割得到的伪标签进行伪标签的一致性正则化，进行伪标签修正。对于目标域迁移阶段，使用目标域数据训练分割网络，将目标域

提取到的特征同源域特征结合，进一步来提高伪标签的质量。在目标域迁移过程中根据目标域模型特征分布，通过像素级的设定上下阈值和根据类别相似度判断前景类和背景类结合的方法，对细化的伪标签进行进一步去噪，获得高质量的伪标签掩膜，用于域自适应训练。

### 3.2 上下文相似性学习

尽管源编码器生成的目标特征与源分割器不对齐，但同一类的特征往往在同一聚类中，而不同类的特征相距较远。这表示源特征编码器对于计算上下文关系是有用的。因此，我们冻结了源编码器，并在编码器中添加了一个额外的头，用于学习上下文语义关系。冻结源编码器的另一个作用是减少训练时间和所需内存。特征图上坐标*i*和坐标*j*之间的语义相似性定义为：

$$S_{ij} = \exp(-||f^{sim}(x_i, y_i) - f_{sim}(x_j, y_j)||_1) \quad (1)$$

计算特征图中每对坐标之间的相似性在计算上是昂贵的。因此，对于每个坐标*i*，在我们的实现中只考虑与位于半径*r*的圆内的坐标*j*的相似性。

给定目标图像*x*，可以从源模型中获得初始伪标签和不确定性掩码，如下所示：

$$d_{ij}^w = ||f_{l,v} - z^w||_2$$

$$m_v = 1[u_v < \eta](1[\hat{y}_v = 1][d_v^{f,g} < d_v^{b,g}] + 1[\hat{y}_v = 0][d_v^{f,g} > d_v^{b,g}])(2)$$

方程2首先从最后一次卷积前的层特征图*f<sub>liv</sub>*中提取类原型*z<sup>w</sup>*，然后通过结合到原型的距离和不确定性来计算不确定性掩码*m<sub>v</sub>*。如果*m<sub>v</sub>*=1，则第*v*个像素的伪标签是可靠的。

获得二进制相似性标签后。对于两个坐标*i*和*j*，如果伪标签为相同，则相似性标签为1，否则为0。只有可靠的伪标签才被认为可以提供较少的噪声监督。使用*S\**训练上下文相似度头。为了解决类不平衡问题，每种类型的相似性的损失（fg-fg，bg-bg，fg-bg）被计算并聚合为：

$$\mathcal{L}_{con} = -\frac{1}{4} \text{avg}_{\substack{\hat{y}_i=\hat{y}_j=1 \\ m_i=m_j=1}}^{\text{avg}} (\log S_{ij}) - \frac{1}{4} \text{avg}_{\substack{\hat{y}_i=\hat{y}_j=0 \\ m_i=m_j=1}} (\log S_{ij}) - \frac{1}{2} \text{avg}_{\substack{\hat{y}_i \neq \hat{y}_j \\ m_i=m_j=1}} (\log (1 - S_{ij})) \quad (3)$$

### 3.3 基于上下文相似性的伪标签提纯

训练后的上下文相似度头用于对初始粗伪标签进行重新细化。具体而言，通过使目标图像通过源编码器和训练的头部来计算上下文相似性*S<sub>ij</sub>*。然后，第*i*个坐标的细化概率被更新为第*i*个坐标周围的局部圆中的概率的加权平均值，如下

$$p_i^{re} = \sum_{d(i,j) \leq r} \frac{S_{ij}^\beta}{\sum_{d(i,j) \leq r} S_{ij}^\beta} \cdot p_j \quad (4)$$

其中*p<sub>i</sub><sup>re</sup>*是修正概率，*d*(·)是欧几里得距离。*β* > 1，以突出突出的相似性而忽略较小的相似性。通过结合基于上下文关系的相邻预测，修正后的概率更加稳健。上式对于*t*轮迭代执行，因为修正的概率可以用于进一步的修正。

### 3.4 基于去噪伪标签的模型自适应

精细化的伪标签可以通过 $\hat{y}_v=1[p_{\mathcal{L}}\gamma]$ 获得。然而，有噪声的伪标签不可避免地存在。模型知识和目标特征分布的组合显示了样本置信度的最佳估计。为此，在像素级和类级选择可靠的伪标签作为

$$m'_{v,p} = 1(p'_v < \gamma_{low} \text{ or } p'_v > \gamma_{high})$$

$$m'_{v,c} = 1(\hat{y}'_v = 1) 1(d_v^{fg} < d_v^{bg}) + 1(\hat{y}'_v = 0) 1(d_v^{fg} > d_v^{bg}), (5)$$

其中 $\gamma_{low}$ 和 $\gamma_{high}$ 是在没有置信概率的情况下滤除伪标签的两个阈值。 $d_v^{fg}$ 和 $d_v^{bg}$ 是到特征原型的距离。最终的标签选择掩码是 $m'_{vfp}$ 和 $m'_{vfc}$ 的交集，即 $m'_v = m'_{vfp} * m'_{vfc}$ 。目标模型选择的伪标签的监督 $m'_v$ 进行交叉熵损失训练，：

$$\mathcal{L}_{seg} = - \sum_v m'_v \cdot [\hat{y}'_v \cdot \log(f^t(x_t)_v) + (1 - \hat{y}'_v) \cdot \log(1 - f^t(x_t)_v)] \quad (6)$$

## 4 复现细节

本次复现是基于作者原有代码来进行改进，作者的部分代码已经开源，但是部分代码并不完整，在结合论文的基本框架和计算公式，在源代码的基础上进行复现实验。同时修改了部分源代码。如修改了开源代码的数据集加载慢与处理各个域数据集生成模型文件复杂耦合等问题。对于原论文模型处理的上下文相似性头部分，并没有使用 $1 \times 1$ 卷积操作，对模型按照框架进行补全，同时尝试在相似性头中加入CBAM模块，来提高模型相似度计算准确度以提升模型分割效果。

### 4.1 实验环境设置

使用作者提供的三个主流数据集用于眼底图像分割，Drishti GS、RIM-ONE-r3和REFUGE挑战的验证集。这些数据集分别分为50/51、99/60和320/80用于训练合测试。

基本遵循原作者的实验设置，网络架构采用DeepLabv3+，上下文相似头分别包括视杯和视盘的两个分支。每个分支包括一个 $1 \times 1$ 的卷积和一个相似性特征图。用于伪标签的阈值 $\gamma$ 设置为0.75。半径 $r$ 、 $\beta$ 和迭代次数 $t$ 设置为4、2和4。根据经验，精细化伪标签的两个阈值分别设置为 $\gamma_{low}=0.4$ 和 $\gamma_{high}=0.85$ 。通过剪裁 $512 \times 512$ 的视盘区域对每张图像进行预处理。使用高斯噪声、对比度调整和随机擦除。采用Adam优化器，在上下文相似性学习阶段和目标域自适应阶段的学习率分别为 $3e-2$ 和 $3e-4$ 。Adam优化器的动量设置为0.9和0.99。批次大小设置为4。上下文相似性头部被训练16个epoch，目标模型被训练20个epoch。为了进行评估，我们采用了广泛使用的dice系数和平均表面距离（ASD）。

### 4.2 创新点

受制于SF-UDA无法对源域数据进行处理。因为使用源域模型预测目标域伪标签存在域间隙。模型中的一致相似性头仅仅通过了一个 $1 \times 1$ 的卷积核来提取相应特征进行像素间相似度的计算。我认为是不够充分的，尝试通过对源域模型所encoder的最后一层特征中引入注意力机制。可以更好地关注相似度计算的有力部分。基于此，通过在一致相似性头中的卷积注

注意力模块CBAM。让原始的高级特征经过注意力加权，得到更加有利于聚类间相似性计算的特征。来达到提高伪标签质量的目的。这里沿用CBAM论文中的设置，先使用通道注意力模块，再使用空间注意力模块进行维度整合，来达到抑制特征噪声，缓解域间隙的作用。

## 5 实验结果分析

我们仿照了原文对Drishti-GS到RIM-ONE-r3数据集进行迁移和REFUGE向Drishti-GS数据迁移。表1展示了模型在两个眼底图像域迁移任务的表现结果。受限于实验平台与设备不同，最终没有达到论文中的实验效果，但是总体的误差小于3%。可以说达到了论文复现的标准，同时比较了原文网络和在相似性计算头中引入CBAM模块的网络。

结果表明，在特征提取加入注意力机制后，较之原模型在各指标间均得到了小幅度的提升。在第二个迁移任务中，Optic disc性能提升幅度明显，表明改进后的模型能够更好的相似性聚类该类，且同cup存在明显的类边界。分析提升总体原因在伪标签细化方案利用域偏移下的特征分布特性来学习上下文关系的过程中，原文的模型因为域偏移，提取的特征不充分使得模型在聚类相似性的学习到的信息不够完整。因此我们的方法比原文提出的方法有了一定的改进。

表1：域迁移分割结果

Methods	Dice[%] ↑			ASD[pixel] ↓		
	Optic cup	Optic disc	Avg	Optic cup	Optic disc	Avg
Source: Drishti-GS ; Target :RIM-ONE-r3						
CPR(exp)	73.43	94.50	83.97	11.83	5.04	8.43
CPR(modi)	<b>73.84</b>	<b>94.94</b>	<b>84.39</b>	<b>9.96</b>	<b>4.63</b>	<b>7.30</b>
Source: REFUGE ; Target : Drishti-GS						
CPR(exp)	80.10	93.36	86.73	13.41	7.27	10.34
CPR(modi)	<b>83.76</b>	<b>96.55</b>	<b>90.15</b>	<b>11.00</b>	<b>3.86</b>	<b>7.43</b>

## 6 总结与展望

这次的前沿论文复现对我来说得到了很好的能力提升。在这项工作中，我们提出了一个SF-UDA的医学图像分割框架。所提出的方法在不需要源域数据的情况下，成功地实现了对目标域的域迁移任务，且性能基本和SOTA方法持平。在对模型相似度计算头的改进下，实验效果有所提高。可见注意力机制对特征的提取仍然存在帮助。文章仅仅是对CNN模型进行了修改，并没有清晰的与Transformer等模型进行定量比较。且图像的上下文相似性判别存在图像内差距过大误判情况。因此未来进一步考虑模型引入Transformer模块以及加入预训练的大模型SAM，来更好的提升域迁移分割的性能。

## 参考文献

- [1] Mathilde Bateson, Hoel Kervadec, Jose Dolz, Hervé Lombaert, and Ismail Ben Ayed. Source-relaxed domain adaptation for image segmentation. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 490–499, Cham, 2020. Springer International Publishing.
- [2] Cheng Chen, Quande Liu, Yueming Jin, Qi Dou, and Pheng-Ann Heng. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021.
- [3] Shishuai Hu, Zehui Liao, and Yong Xia. Proslda: Prompt learning based source-free domain adaptation for medical image segmentation. *ArXiv*, abs/2211.11514, 2022.
- [4] Yun Jiang, Wenhuan Liu, Chao Wu, and Huixia Yao. Multi-scale and multi-branch convolutional neural network for retinal image segmentation. *Symmetry*, 13:365, 2021.
- [5] Liangzhi Li, Manisha Verma, Yuta Nakashima, Hajime Nagahara, and Ryo Kawasaki. Iter-net: Retinal image segmentation utilizing structural redundancy in vessel networks. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3645–3654, 2019.
- [6] Xiaofeng Liu, Fangxu Xing, Georges El Fakhri, and Jonghye Woo. Memory consistent unsupervised off-the-shelf model adaptation for source-relaxed medical image segmentation. *Medical Image Analysis*, 83:102641, 2023.
- [7] Xiaofeng Liu, Fangxu Xing, Chao Yang, Georges El Fakhri, and Jonghye Woo. Adapting off-the-shelf source segmenter for target medical image segmentation. 2021.
- [8] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021.



- [9] Chen. Yang, Xiaoqing. Guo, Zhen. Chen, and Yixuan. Yuan. Source free domain adaptation for medical image segmentation with fourier style mining. *Medical image analysis*, 79:102457, 2022.
- [10] Jiawei Zhang, Yanchun Zhang, and Xiaowei Xu. Pyramid u-net for retinal vessel segmentation. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1125–1129, 2021.
- [11] S. Zhang, H. Fu, Yuguang Yan, Yubing Zhang, Qingyao Wu, Ming Yang, Mingkui Tan, and Yanwu Xu. Attention guided network for retinal image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019.