

# 《半监督域泛化医学图像分割中伪标签质量的提高》复现

江思瑶

## 摘要

将医学图像分割算法推广到不可见域是计算机辅助诊断和外科手术的一个重要研究课题。大多数现有的方法都需要在每个源域中有一个完全标记的数据集。虽然有研究人员提出了半监督域泛化方法，但它仍然需要域标签。本文对原论文<sup>[1]</sup>进行复现，通过一种基于置信度感知的交叉伪监督算法来提高来自未知分布的未标记图像的伪标签质量，并使用傅里叶变换进行数据增强以合并跨域信息。我们在原论文的基础上，主要增加了对比损失模块和双层感知的交叉伪监督模块，实验结果表明，改进后的算法具有更高的分割精度。

**关键词：**半监督；域泛化；交叉伪监督

## 1 引言

语义分割是计算机视觉领域的一项基本任务，随着深度神经网络<sup>[2]</sup>的兴起，近年来得到了很大的发展，其中医学图像分割作为计算机辅助诊断和计算机辅助手术的基本任务之一也得到了大量关注。近年来，研究人员开发了许多用于医学图像分割的卷积神经网络，然而由于医学图像通常收集自不同的临床中心和不同的扫描仪，因为患者群体、扫描仪和扫描采集设置的变化，图像可能有明显的域偏移，导致在对未知区域的图像进行测试时，结果较差。因此，加强模型在不同领域位移下的泛化能力至关重要。

解决域偏移的一个简单解决方案是获取和注释尽可能多的数据，但是这些注释数据的成本是相当高的，既费时又耗力。为了缓解这个问题，人们对半监督语义分割进行了许多尝试，通过使用少量标记样本和大量未标记样本对模型进行训练。另一种解决方案是在源域上训练模型，并利用目标域的一些信息将其推广到目标域，即域适应<sup>[3]</sup>(DA)。而领域泛化 (DG) 则是一种更严格的设置，它与 DA 的区别在于模型不使用来自目标领域的任何信息。综合两种解决方案，可以通过半监督域泛化的方法来缓解医学图像注释困难的问题和域偏移的问题，但由于标注数据有限，半监督学习依然很难达到全监督学习的训练精度，所以基于半监督学习的图像语义分割研究仍有很大的研究意义。

## 2 相关工作

### 2.1 域泛化

领域泛化 (DG) 用多个源域训练模型，并将其推广到未见的目标域。现有的 DG 方法可以分为三类：表示学习<sup>[4]</sup>，学习策略设计<sup>[5]</sup>和数据操作<sup>[6]</sup>。表示学习方法主要遵循域自适应的思想，通过学习域不变特征或域间显式特征对齐。学习策略方法侧重于利用一般学习策略来提高泛化能力。数据操作旨在操作输入以帮助学习一般表示。最近，傅里叶变换方法也被证明是非常有效的 DG。Xu 等人<sup>[7]</sup>提出了一种基于傅里叶变换的域泛化框架，通过将源图像的振幅谱替换为来自不同源域的另一幅图像的振幅谱。该方法在分类任务中取得了显著的效果。Liu 等人<sup>[8]</sup>也在联邦学习中使用了傅里叶变换方法，并证明它是联邦学习下医学图像分割的一个有用的增强。

与上述方法不同，本文使用傅里叶变换作为数据增强来获得不同域之间的低级统计信息，并为未标记的图像合并跨域信息。利用这些扰动，可以测量伪标签的置信度，然后用可靠的伪标签来提高模型的性能。

## 2.2 半监督语义分割

与图像分类任务不同，为分割任务手动标记像素注释昂贵且耗时。现有的半监督分割方法可以大致分为两类：自我训练<sup>[9]</sup>和一致性学习<sup>[10]</sup>。自我训练方法主要采用伪标签方法来提高性能，最初是为在分类任务中使用未标记的数据而开发的，接着将其应用于半监督分割，该方法使用了在标记数据预训练的分割模型得到的未标记数据的伪分割图来重新训练分割模型。而一致性学习鼓励模型在不同的转换后，对相同的输入有相似的输出，它对不同增强图像的预测结果进行一致性约束，使决策函数位于低密度区域。最近，许多方法将这两种思想结合起来，以获得更好的分割任务性能。

因此，本文复现了两个创新的模块来提高伪标签的质量：一个是傅立叶变换，用于增加来自源域的单个输入的跨域信息；二是测量伪标签的置信度，得到可靠的监督信号。

## 3 本文方法

### 3.1 本文方法概述

本文对基于傅里叶变换的数据增强模块和基于置信度感知的交叉伪监督模块进行复现，通过傅里叶变换模块对图像进行数据增强，利用跨区域信息来增强训练样本，并通过置信度感知交叉伪监督模块，计算伪标签的方差作为置信度，对模型进行交叉伪监督，来提高伪标签的质量。整体框架如图 1 所示。

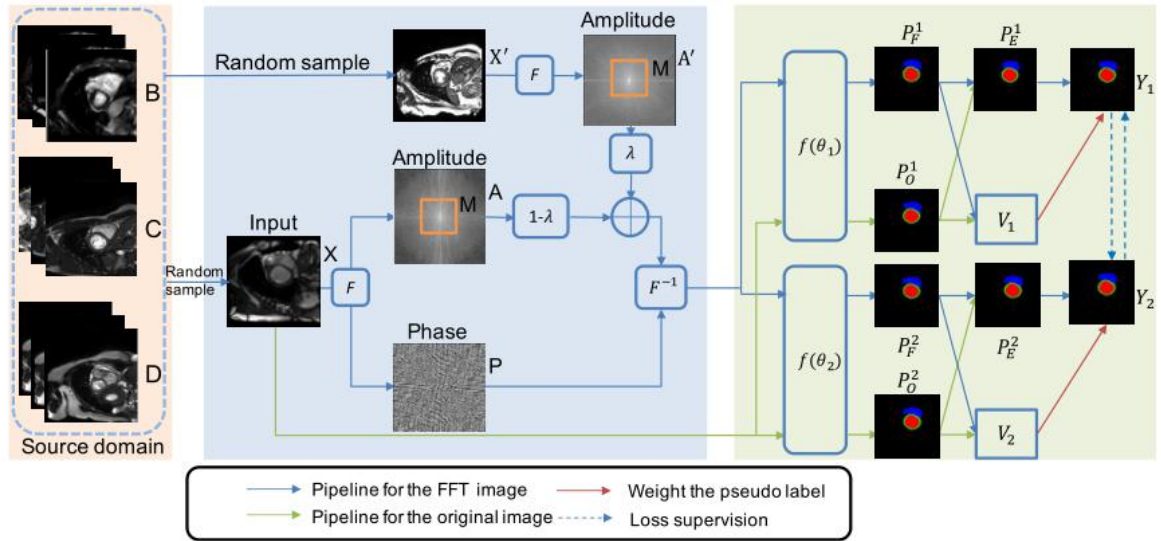


图 1: 整体框架

### 3.2 基于傅里叶变换的数据增强

在训练过程中，从源域随机取一张图像  $X$  作为输入图像，然后对图像进行傅里叶变换  $F$ ，将图像转移到频域，得到幅值谱  $A$  和相位图  $P$ ，其中幅值谱包含低级统计信息，相位图包含原始信号的高级语义。为了提高域泛化能力，从源域随机选择另一个样本  $X'$ ，进行傅里叶变换，得到另一个幅值谱  $A'$ 。然后，合并第二幅图像  $X'$  的幅值谱来增强第一张图像  $X$ ，对合并后的幅值谱  $A_{new}$  进行反傅里叶

变换得到增强样本  $Z$ 。幅值谱  $A_{new}$  和增强样本  $Z$  的计算公式如式 2.7 和式 2.8。

$$A_{new} = (1 - \lambda)A * (1 - M) + \lambda A' * M \quad (1)$$

$$Z = F^{-1}(A_{new}, P) \quad (2)$$

其中  $\lambda$  是用来调整  $X$  与  $X'$  相位信息比值的参数； $M$  为二进制掩码，用于控制要交换的振幅谱的空间范围，我们将  $M$  设为振幅谱中包含低频信息的中心区域。

### 3.3 基于置信度感知的交叉伪监督

将原始图像  $X$  和变换后的图像  $Z$  分别放入两个并行分割网络  $f(\theta_1)$  和  $f(\theta_2)$ ，对于每个分割网络，得到原始图像  $X$  和变换后图像  $Z$  的预测，分别计算两个模型预测结果的平均值作为总预测结果。为了选择性地使用伪标签作为监督，减少低质量标签，首先计算原始图像和变换图像的 KL 散度作为预测方差来衡量伪标签的质量。如果这两个预测的差异较大，则计算出的方差值也较大，说明这两个预测的质量相对较低，反之亦然。在得到这两个网络预测的方差后，利用交叉监督信号，通过置信度感知损失函数来优化网络：

$$L_{caccp} = L_a + L_b \quad (3)$$

$$L_a = E[e^{-V_1} L_{ce}(P_E^2, Y_1) + V_1] \quad (4)$$

$$L_b = E[e^{-V_2} L_{ce}(P_E^1, Y_2) + V_2] \quad (5)$$

其中  $Y_1$  和  $Y_2$  是由概率图  $P_E^1$  和  $P_E^2$  生成的 one-hot 向量， $L_{ce}$  表示交叉熵损失。

## 4 复现细节

### 4.1 与已有开源代码对比

由于本次实验复现的论文已开源，我们结合论文的基本框架和计算公式，参考开源代码进行复现。复现的代码整体框架与开源代码相似，经对比后，复现的代码与开源代码相比，整体减少了代码冗余部分，使得训练步骤更加容易实现。并对论文部分结构的代码进行改进，如下所示：

(1) 更改数据集加载代码：由于原代码的数据加载部分在划分有监督和无监督两个部分的数据时，在总数据集上划分出前面的部分数据作为标注数据，再将整个数据集作为无标注数据。而在半监督训练中，标注数据的选择是至关重要的，若只在数据集的前部分数据提取标注数据的话，一旦数据集的数据分布不均匀，监督训练后的网络将无法与无监督训练时的无标注数据相拟合，此时测试精度也无法反应出模型的好坏。因此，考虑数据集可能分布不均匀，而且标注数据对模型的训练相对重要，我们对数据集进行预处理，随机划分出训练集的标注数据和无标注数据，并保存其路径到 txt 文件中，在加载数据集时，只需读取 txt 文件中划分好的路径，而不需要重复遍历数据集，一定程度上减轻了模型训练时读取数据的负担。这样在调试和训练网络时，可以通过测试精度来直观地反应模型的好坏。

(2) 减少图像的交叉融合：由于原代码创建了两个无标注数据集，其中一个通过 cutmix 对图像的一部分区域进行裁剪并填充其它数据的区域像素值，然后在训练过程中将两个数据集的分割结果进行交叉融合获得最终的特征图。但由于我们选用的数据集噪声较多，且含有标注错误的数据，并且病灶的边界模糊，综合各种因素我们选择将 cutmix 模块去除，使得模型在训练过程中只需要对一个数据集进行分割训练。这样不仅使得代码更加简洁，也使得图像的分割结果更加精确。

(3) 增加双层感知的交叉伪监督模块：由于原论文中的置信度感知交叉伪监督模块使用的是模型输出的两个预测图的方差来作为置信度，再计算平均预测图与另一个模型的输出掩膜的交叉损失作为无监督的一致性损失。考虑到论文只在两个模型之间进行交叉伪监督，我们在单个模型的两个输出之间也再一次进行交叉伪监督，以此来拉近源图像和目标图像之间的预测距离，提高无监督训练中的伪标签的质量。

## 4.2 实验环境和数据集

我们在 Pytorch1.11 上复现了这个模型，并在 Ubuntu20.04 系统上使用两个 NVidia 3090 GPU 来训练。使用 DeepLabv3+ 和 ResNet50 来复现两个分割网络，并采用 ImageNet 训练的权值进行模型初始化。实验选用的数据集为乳腺超声数据集 (BUSI)<sup>[11]</sup>，该数据集于 2018 年从 Baheya 医院收集，其中包含 600 名患者的 780 张图像 (437 例良性，210 例恶性，133 例正常)，图像分辨率为 1280×1024。由于域信息集中在良性和恶性图像上，因此不采用正常图像。此外，由于该数据集中良性图像和恶性图像的比例为 2: 1，我们对所有恶性图像采用数据增强方法，通过对图像进行水平翻转，得到平衡的数据集。

## 4.3 创新点

在论文中，我们已知其具有基于置信度感知的交叉伪监督模块，该模块利用模型输出的两个预测图的方差作为置信度，计算平均预测图与另一个模型的输出掩膜的交叉损失值，来进行一致性学习。参考论文<sup>[12]</sup>中的双向一致性约束的思想，我们考虑在置信度感知交叉伪监督模块的基础上，增加一项交叉一致性感知模块，来拉近两个模型输出的距离。

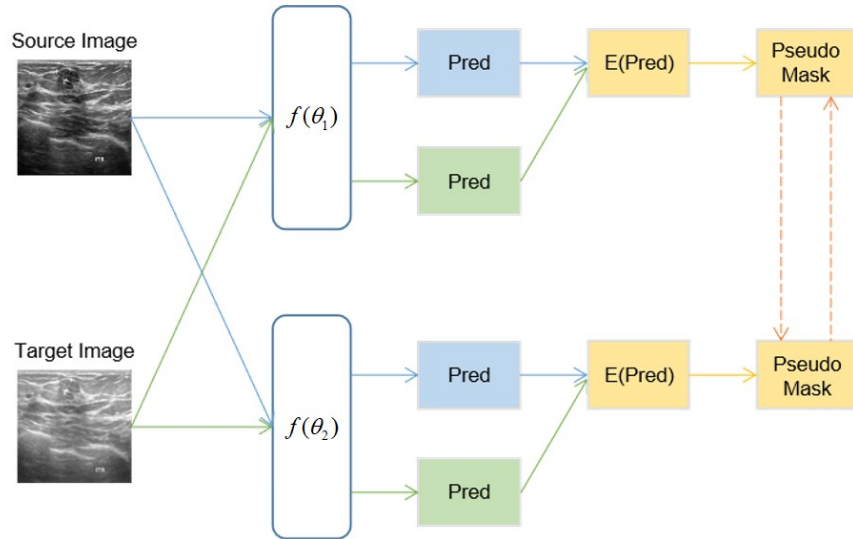


图 2: 操作界面示意

具体操作如图 2 所示。首先对原始图像进行增强，得到目标图像，将两张图像都输入两个模型中，两个模型结构完全相同，但初始化参数不同，然后模型对输入进行预测得到两张概率图，我们对原始图像的概率图和目标图像的概率图求平均，得到一张平均概率图：

$$P_E^1 = (P_O^1 + P_F^1)/2 \quad (6)$$

$$P_E^2 = (P_O^2 + P_F^2)/2 \quad (7)$$

再对两个模型的平均概率图进行 KL 散度的计算，作为损失值：

$$V_1 = E[P_E^1 \log(\frac{P_E^1}{P_E^2})] \quad (8)$$

$$V_2 = E[P_E^2 \log(\frac{P_E^2}{P_E^1})] \quad (9)$$

我们在原论文的基于置信度感知的交叉伪监督模块中融入交叉一致性感知模块，构成一个双层感知的交叉伪监督模型，使得网络可以更好地进行一致性学习，约束两个模型的输出越来越接近。

## 5 实验结果分析

为了验证本实验创新点的有效性，我们在 BUSI 数据集中对不同百分比（25% 和 50%）的标记图像进行了实验。实验结果如表 1 所示，在复现论文原有方法（EPL）和加上创新点的方法上分别使用不同百分比的带标签数据进行训练。可以看到，相对于 EPL 方法，我们改进后的方法取得了显著的改进。我们使用 Dice 精度指标作为评估实验效果的方法，我们的方法在使用 25% 标记数据时比原有方法提高了 1.68%，在使用 50% 标记数据时比原有方法提高了 1.04%。

为了更直观地表示实验结果的分割情况，我们对分割结果进行可视化，图 3 显示了我们的方法在数据集图像上产生的视觉结果，可以看到，我们的方法成功地在数据集图像上生成了与真实分割图十分接近的预测结果。

表 1: 在不同百分比的标记数据上的 Dice 结果

Method	Percentage		Dice(%)
	Labeled	Unlabeled	
EPL	25%	75%	67.25
Ours	25%	75%	68.93
EPL	50%	50%	70.52
Ours	50%	50%	71.56

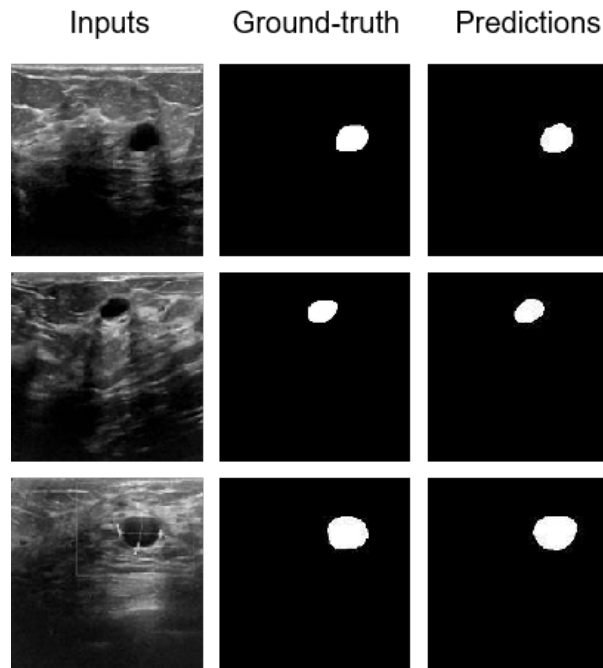


图 3: 实验结果示意

## 6 总结与展望

本文对论文“半监督域泛化医学图像分割中伪标签质量的提高”进行复现，并在原代码的基础上进行改进，论文的主要模块为基于傅里叶变换的数据增强和基于置信度感知的交叉伪监督，我们对交叉伪监督算法进行改进，实现双层感知的交叉伪监督，并对原代码的数据加载部分和数据集处理部分进行修改，使得代码更加适用于我们的数据集。最后我们进行消融实验以及结果可视化，结果表明改进后模型的分割结果优于原始模型。本次实验由于考虑欠缺，在程序计算精度时，没有对多种测量指标进行计算，只计算了 Dice 精度，这是本次实验的不足。而且由于本次复现的工作主要体现在创新点部分，所以对实验没有做足够多的消融实验，与其它一些先进的半监督学习方法进行比较，进一步验证实验结果的有效性，这也是本次实验不足的一点。

在实验过程中，发现了数据集少量数据标注错误的问题，该情况会给数据带来噪声，同时也可能会使网络学习错误的数据导致分割结果的不准确。但由于在搜索相关资料之后并未发现有他人的工作对数据集的标签进行修改，并且得知少量的噪声在训练过程中或许不会产生太大的影响，由于时间关系没有对其进行验证，在未来的学习中将会在这个方向上进行进一步研究。此外，在实验过程中，发现模型对于数据集中的恶性图像分割精度比良性图像的分割精度低，通过分析得知是因为恶性图像中病灶区域的边界比较模糊，而良性图像中病灶区域的边界相对清晰，从而导致了分割精度的偏差。通过阅读许多文献得知，可以在模型中添加边缘检测的分支或对图像的边缘添加惩罚项，使得模型更加关注图像的边缘区域，这是我们在未来主要研究的方向。

## 参考文献

- [1] HUIFENG Y, XIAOWEI H, XIAOMENG L. Enhancing Pseudo Label Quality for Semi-supervised Domain-Generalized Medical Image Segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 36: 3. 2022: 3099-3107.
- [2] LIANG-CHIEH C, GEORGE P, IASONAS K, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.
- [3] CHENG B, CHENGLANG Y, JIEXIANG W, et al. Uncertainty-aware domain alignment for anatomical structure segmentation[J]. Medical Image Analysis, 2020, 64: 101732.
- [4] FAN Z, ZHUQING J, CHANGJIAN S, et al. Domain generalization via optimal transport with metric similarity learning[J]. arXiv preprint arXiv:2007.10573, 2020.
- [5] TERESA Y, FATIH K O, AMIR Z. Robustness via cross-domain ensembles[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 12189-12199.
- [6] JOSH T, RACHEL F, ALEX R, et al. Domain randomization for transferring deep neural networks from simulation to the real world[C]//2017 IEEE/RSJ international conference on intelligent robots and systems (IROS). 2017: 23-30.
- [7] QINWEI X, RUIPENG Z, YA Z, et al. A fourier-based framework for domain generalization[C]//

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14383-14392.

- [8] QUANDE L, CHENG C, JING Q, et al. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 1013-1023.
- [9] DONG-HYUN L, et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Workshop on challenges in representation learning, ICML: vol. 3: 2. 2013: 896.
- [10] ANTTI T, HARRI V. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.
- [11] WALID A D, MOHAMMED G, HUSSEIN K, et al. Dataset of breast ultrasound images[J]. Data in brief, 2020, 28: 104863.
- [12] ZHANGHAN K, DI Q, KAICAN L, et al. Guided collaborative training for pixel-wise semi-supervised learning[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. 2020: 429-445.