

文章编号:1007-2780(2024)08-1014-10

基于双模型互学习的半监督中医舌诊 图像分割方法

李方旭¹, 徐望明¹, 徐 雪², 贾 云^{3*}

(1. 武汉科技大学 信息科学与工程学院, 湖北 武汉 430081;

2. 武汉科技大学 医学院, 湖北 武汉 430065;

3. 中国地质大学(武汉)医院, 湖北 武汉 430074)

摘要:舌体图像的准确分割是中医舌诊客观化分析的重要前提,目前广泛采用的全监督分割方法需要对大量像素级标注的样本进行训练,而基于单一模型的半监督分割方法对学习到的错误知识缺乏自我纠正能力。针对这一问题,本文提出一种新颖的基于双模型互学习的半监督舌体图像分割方法。首先,模型A、B分别在有标签数据集上进行监督训练。之后,模型A、B进入互学习阶段,使用本文设计的互学习损失函数,根据双模型对无标签数据预测的分歧而赋予不同的权重。先由模型A对无标签数据集生成伪标签,模型B在有标签数据集和伪标签数据集上进行微调,然后模型B再对无标签数据集生成伪标签,模型A以同样的方式进行微调。双模型微调结束后,选择性能更佳的一个作为最终的舌体图像分割模型。实验结果表明,本文方法的有标签数据比例为1/100、1/50、1/25、1/8时,mIoU分别达到96.67%、97.92%、98.52%、98.85%,优于对比的其他典型半监督方法。本文方法仅需少数标签数据便可达到较高的舌体图像分割精度,可为后续的中医舌色、舌形等舌象分析应用奠定坚实基础,推动中医诊疗数字化进程。

关键词:半监督;互学习;舌体图像分割;损失函数;中医数字化

中图分类号:TP391 文献标识码:A doi:10.37188/CJLCD.2023-0308

Semi-supervised tongue image segmentation method for traditional chinese medicine based on mutual learning with dual models

LI Fangxu¹, XU Wangming¹, XU Xue², JIA Yun^{3*}

(1. School of Information Science and Engineering, Wuhan University of Science and Technology,
Wuhan 430081, China;

2. School of Medicine, Wuhan University of Science and Technology, Wuhan 430065, China;

3. Affiliated Hospital, China University of Geosciences (Wuhan), Wuhan 430074, China)

Abstract: Accurate tongue image segmentation is a crucial prerequisite for objective analysis in tongue diagnosis in traditional Chinese medicine (TCM). At present, the widely-used full-supervised segmentation methods require a large number of pixel-level annotated samples for training, and the single-model-based

收稿日期:2023-09-22;修订日期:2023-10-16.

基金项目:国家自然科学基金(No.51805386);国家重点研发计划(No.3502300, No.3502302)

Supported by National Natural Science Foundation of China (No.51805386); National Key R&D Program of China (No.3502300, No.3502302)

*通信联系人, E-mail:jiayun@cug.edu.cn

semi-supervised segmentation methods lack the ability to self-correct the learned error knowledge. To address this issue, a novel semi-supervised tongue image segmentation method based on mutual learning with dual models is proposed. Firstly, model A and B undergo supervised training on the labeled datasets. Subsequently, model A and B enter the mutual learning phase, utilizing a designed mutual learning loss function, in which different weights are assigned based on the disagreement between predictions of the two models on the unlabeled data. Model A generates the pseudo-labels for the unlabeled dataset, and model B fine-tunes on both the labeled dataset and the pseudo-labeled dataset. Then, model B generates the pseudo-labels for the unlabeled dataset, and model A fine-tunes in the same manner. After the dual-model fine-tuning process, the model with better performance is selected as the final tongue image segmentation model. Experimental results show that with labeled data proportions of 1/100, 1/50, 1/25, and 1/8, the mean intersection over union (mIoU) achieved by the proposed method is 96.67%, 97.92%, 98.52%, and 98.85%, respectively, outperforming other typical semi-supervised methods compared. The proposed method achieves high precision in tongue image segmentation with only a small number of labeled data, laying a solid foundation for subsequent applications in TCM such as tongue color, tongue shape and other tongue image analysis, which can promote the digitization of TCM diagnosis and treatment.

Key words: semi-supervised; mutual learning; tongue image segmentation; loss function; digitization of TCM

1 引言

舌诊是中医望诊中一种重要的诊断方法,医生通过观察舌体质地、颜色、形状等特征,作为诊断患者健康状况的依据。传统的舌诊依赖于医生用眼睛观察舌头,从而做出诊断。这种诊断方式严重依赖于医生的临床经验和主观视觉感知,因此传统舌诊的诊断结果是主观的。随着人工智能技术的发展,传统舌诊结合人工智能技术实现舌诊数字化、客观化诊断成为主要研究趋势^[1]。然而,通过计算机分析舌体图像面临的首要问题是,在舌体图像采集过程中,不可避免地会拍摄到嘴唇、下巴、脸部等干扰部位,将舌头从这些复杂背景中准确分割出来是后续对舌象进行分析的重要前提和保证。

近年来,卷积神经网络(Convolutional Neural Network, CNN)凭借其强大的特征学习能力被广泛应用于舌体图像的语义分割。这些方法^[1-3]利用精细标注的数据进行全监督训练并取得了非常高的性能。张新峰等^[4]提出一种基于DCNN和全连接条件随机场的舌体图像分割算法,有效细化了舌体分割边缘。颜建军等^[5]使用Mask R-CNN实现了舌体图像的分割,其分割效果好于传统的阈值分割算法。徐武等^[6]基于DeepLabv2和生成对抗网络构建了舌体分割网络,使用生成

对抗网络对舌体图像进行去噪处理,提升了舌体图像的分割精度。上述方法推动了深度学习技术在舌体图像分割领域的应用,但这些方法均为全监督语义分割,其训练需要大量的像素级标注数据,标注这些像素级标签需要耗费大量的人力物力。为了降低标注成本,研究人员尝试使用较少的标记数据建立高性能的语义分割模型,包括弱监督^[7]、自监督^[8]、半监督^[9]等方法。本文主要研究半监督舌体图像分割方法,即利用少量的标注数据和大量的未标记数据构建高性能的分割模型。Zhou等^[10]结合对抗学习提出了一种舌体图像半监督分割方法,使用少量标注达到了较好的效果。半监督方法又可以分为对抗训练方法、一致性正则化方法和伪标记方法。半监督方法的关键之一是模型如何从无标签数据中学习 to 更多的正确的知识。目前半监督方法广泛使用单一模型,单一的模型一旦学习到错误知识便无法自我纠正,这对于舌体图像分割任务是非常不利的,即模型可能错误分割嘴唇和舌体的边缘。因此,本文基于双模型构建了舌体图像分割模型。双模型采用互学习机制,每个模型为另一个模型生成无标签数据的伪标签,即每个模型都可以从另一个模型获得无标签数据的知识。双模型同时犯一样的错误概率远小于单一模型,避免了单一模型无法自我纠正错误的情况。本文的

主要贡献如下:(1)与之前半监督方法广泛使用的单一模型不同,本文提出一种基于双模型互学习的半监督舌体图像分割方法;(2)为了适应双模型互学习的训练流程,本文提出了模型分歧损失,与交叉熵损失函数共同构成互学习损失函数;(3)本文将所构建的舌体图像数据集划分了1/100、1/50、1/25、1/8四个标签比例并经过大量实验验证,本文方法在所收集的舌体图像数据集上优于目前典型的半监督分割方法结果。

2 相关工作

2.1 图像分割

传统的图像分割方法主要利用图像的像素值特征实现,其可以分为区域分割、边缘检测以及结合特定理论工具的分割方法。由于舌体图像较为复杂,嘴唇、脸部与舌体颜色较为接近,传统方法的分割准确性、鲁棒性较差,难以达到实际应用需求。随着深度学习在语义分割领域的迅速发展,科研工作者提出了一系列基于CNN的语义分割网络。全卷积网络^[11]是语义分割领域的重要里程碑,它将传统的卷积神经网络的全连接层替换为全卷积层,以接受任意大小的输入图像。U-Net^[12]是一种编码器-解码器网络,被广泛应用于医学图像分割领域,其结合了高分辨率的特征图和低分辨率的语义信息,通过跳跃连接将编码器的特征与解码器的特征相结合,以提高分割精度。DeepLab是由Google团队开发的一系列语义分割网络,DeepLabv3+^[13]是目前最新版本,其采用编码器-解码器结构和空洞卷积,利用多尺度信息和ASPP模块,进一步提高了图像语义分割性能。

2.2 半监督图像分割

近年来提出的半监督分割方法可以分为对抗训练方法、一致性正则化方法和伪标记方法^[14]。基于对抗训练的方法^[15]将分割模型作为生成器,并构建判别器用于区分生产的分割掩膜和真实的分割掩膜,生成器和判别器之间的对抗学习使分割结果趋向真实掩膜的形状分布。一致性正则化方法基于平滑度假设和聚类假设,即如果对一个无标签样本施加扰动,网络对它的预测结果不应发生显著变化^[16]。该方法在输入层或者特征层对未标注数据引入扰动,这样对同一个无

标签样本会得到两个不同的预测结果,最后在损失函数上添加一致性正则化,使两个预测结果尽可能保持一致。基于一致性正则化的方法^[17-18]之间的主要区别在于它们对数据扰动方式的不同,扰动方式可以分为输入扰动、特征扰动、网络扰动、混合扰动等。基于输入扰动的方法使用数据增强技术直接应用于输入图像。基于特征扰动的方法将扰动引入分割网络内部,从而获得变动的特征。基于网络扰动的方法使用不同的网络,网络之间的差异构成所得到的模型预测中的扰动。使用不同的预训练权重初始化网络也是常用的网络扰动方法。基于混合扰动的方法综合使用上面3种扰动类型,例如有研究^[19]提出了将输入、特征和网络扰动相结合的方法。伪标记方法是最早出现的半监督方法之一,这种类型的方法将现有的监督模型扩展到半监督场景,使用其处理未标注的数据。伪标记方法的基本思路是:首先在有标签数据上对模型进行训练,再根据训练得到的模型做出的预测生成未标注数据的伪标签。然后,合并有标签数据和伪标签作为新数据集,并在新数据集上训练模型。典型的伪标签方法便是自训练方法^[20-21],其仅使用一个有监督的模型。在伪标签方法中,生成的伪标签的质量高低决定了模型的性能优劣,质量较低的伪标签可能会使模型学到错误的知识,导致模型性能下降。为了提升伪标签质量,大多方法选择多次迭代,并过滤置信度较低的伪标签。但自训练方法采用的是单一模型,单一模型一旦学习到错误的知识便很难自我纠正,且错误知识会随着迭代次数的增加不断累积,最终可能导致模型性能的下降。在舌体图像分割任务中,舌体边缘与嘴唇相邻且颜色纹理相近,若使用自训练这种缺乏检测其自身错误机制的方法可能造成嘴唇和舌体边缘分割界限不清晰。此外,舌体图像数据的采集较为繁琐且耗时耗力。选择过滤置信度较低的伪标签也对舌体图像无标签数据造成了极大浪费。鉴于自训练方法存在的缺陷,有研究^[22-23]尝试使用多模型互学习。互学习方法在自训练方法基础上进行了扩展,其使用多个模型,每个模型不是从自己的预测中学习,而是使用其他模型生成的伪标签进行训练。这种互学习的机制有效解决了自训练方法陷入错误累积

的问题。Yan 等^[24]针对多模型半监督训练中难以学习到稳定知识的问题,提出一种伪标签增强策略。方超伟等^[25]提出一种用于心脏结构、肿瘤分割等医学图像分割的双模型半监督方法,并在仅采用 30% 的标注比例时,取得了较好的效果。

受到上述研究启发,本文使用双模型互学习的方法构建了舌体图像分割模型,在舌体图像分割任务上取得了较好的效果。

3 本文方法

3.1 双模型结构

本文提出的双模型结构由结构相同、初始化权重不同的两个模型组成。网络结构采用语义分割中常用的 DeepLabv3+ 网络。DeepLabv3+ 是基于编码解码结构和空洞空间卷积池化金字塔(ASPP)的语义分割模型,编码器负责特征提取与下采样,解码器则负责上采样得到输出结果。本文以 ResNet-50 作为编码器主干网络提取特征,然后使用 ASPP 获取不同感受野下的图像特征。解码器将经过 ASPP 模块后得到的特征图进行 4 倍上采样,与骨干网络中提取的相同尺寸特征图进行融合,再通过 3×3 卷积和 4 倍上采样,输出与原图大小相等的预测图像。

本文构建了舌体图像数据集 D , 数据集 D 包括有标签数据集 D^l 和无标签数据集 D^u , 即 $D = \{D^l, D^u\}$ 。整体模型结果如图 1 所示。两个模型分别记为模型 A 和模型 B, 首先对模型 A 和模型 B

使用不同的预训练权重以获得模型初始化分歧。然后在标签数据集 D^l 上使用监督损失函数分别训练 A 和 B, 之后固定 A 并使用 A 在无标签数据集 D^u 上生成伪标签, 再在数据集 D 上使用互学习损失函数训练模型 B。同理, 再固定 B 并使用 B 在无标签数据集 D^u 上生成伪标签, 并在数据集 D 上使用互学习损失函数训练模型 A。模型训练伪代码如算法 1 所示。由于两个模型同时学到同一个错误的可能性很低, 两个不同的模型在分类决策上有分歧, 可能会识别彼此的错误。因此, 模型 A 和模型 B 之间互学习有效避免了使用单一模型在学习到错误知识无法自我纠正的情况。训练完成后, 选取模型 A 和模型 B 中性能最佳的一个作为舌体图像分割模型。

算法 1: 双模型互学习伪代码

输入: 有标签数据集 D^l , 无标签数据集 D^u

输出: 性能最优的模型 M

初始化:

使用 ImageNet 预训练权重初始化模型 A

使用 COCO 预训练权重初始化模型 B

全监督训练:

在 D^l 上训练模型 A

在 D^l 上训练模型 B

互学习:

固定模型 A, 在 D^u 上生成伪标签

在 D^l 和 D^u 上对模型 B 进行微调

固定模型 B, 在 D^u 上生成伪标签

在 D^l 和 D^u 上对模型 A 进行微调

返回: 最优模型 M 为模型 A 和模型 B 中 mIoU 较高者

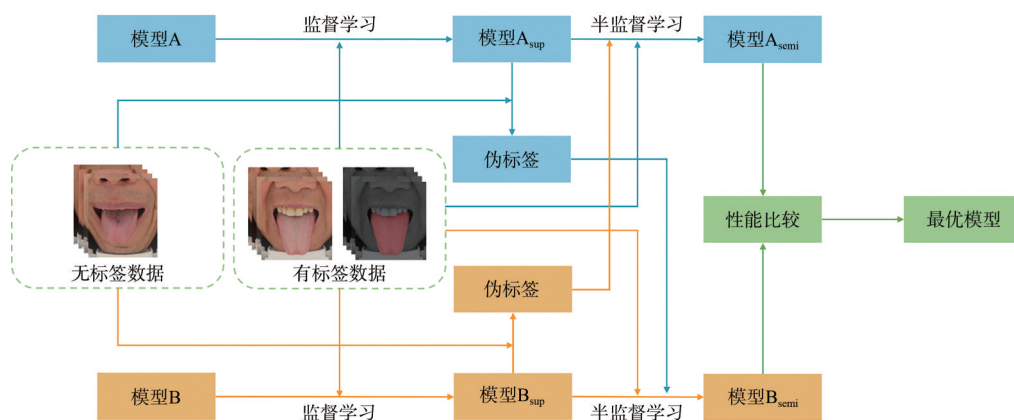


图 1 双模型结构图

Fig. 1 Structure diagram of dual model

3.2 初始化分歧

双模型能够解决单一模型无法自我纠错的一大原因便是双模型预测之间存在分歧,一个模型可以纠正另外一个模型。两个模型犯同一个错误的概率是很小的,不同模型都已高置信度预测错误类的可能性更小。通过这一点,量化不同模型的预测之间的分歧,允许模型根据分歧情况对损失函数进行加权,最大化利用分歧来提升模型的性能和鲁棒性。利用双模型预测间的分歧的关键是如何初始化足够不同的模型。目前,初始化分歧有使用不同的预训练权重、对数据集引入不同的噪声等方式。使用不同的预训练权重是广泛采用的方法,其在模型初始化时引入不同数据集(ImageNet和COCO)下训练的预训练权重。在舌体图像分割任务中,引入噪声虽然可以让模型学习到不同的知识,在预测时存在足够的分歧,但是,引入不必要的噪声极易导致舌体图像分割质量的下降,例如舌与嘴唇边缘分割模糊,面部等其他位置错误分割为舌等。因此,本文使用不同的预训练权重来初始化模型分歧,对于双模型分别引入了在ImageNet和COCO数据集下训练的权重。

3.3 互学习损失函数

在训练进入到双模型互学习阶段时,如果对有标签数据和无标签数据均使用交叉熵损失函数,则双模型对无标签数据的利用性较低,无法有效利用模型A和模型B对无标签数据预测的分歧。为保证双模型中每个模型既可以充分学习有标签图像,也可以通过另一个模型学习到无标注图像的知识,本文将损失函数设计为有监督损失和模型分歧损失两个部分。有监督损失针对数据集中的有标签数据使用,本文采用二元交叉熵损失函数构建有监督损失。

为了适应双模型互学习结构,本文提出了一种模型间分歧损失,用于模型A和模型B互学习无标签数据时使用。假设模型A和模型B已经完成了在有标签数据集 D' 的训练,进入互学习阶段。以模型A正在训练模型B为例,即模型A固定,并生成无标签数据集 D'' 的伪标签供模型B训练。模型B在数据集 D , $D = \{D', D''\}$ 上训练,记 l, u 分别为 D', D'' 中的一个样本; p_l 和 p_u 分别为样本 l, u 中一个像素; y_A 为模型A对 p_u 的预测

结果,即为模型B的伪标签; y_B 为模型B对 p_u 的预测结果。则 y_A 与 y_B 存在两种情况,即 $y_A = y_B$ 和 $y_A \neq y_B$,对这两种情况赋予不同的权重 w_u ,即:

$$w_u = \begin{cases} p_B y_A = y_B \\ p_B^\gamma y_A \neq y_B \end{cases}, \quad (1)$$

其中: p_B 为模型B对无标签数据每个像素点的伪标记类别 y_A 的预测概率,用以量化模型A和模型B预测的一致性,较高的 p_B 对应着模型A和模型B具有较高的预测一致性; γ 为超参数,用于动态调整双模型间分歧的差异。 $y_A = y_B$ 时,说明模型A和模型B对 p_u 的预测一致,以 p_B 作为损失的加权系数; $y_A \neq y_B$ 时,说明模型A和模型B对 p_u 的预测不一致,使用超参数 γ 减小权重的大小,从而强调了模型的互学习。对于尺寸为 $h \times w$ 的输入图像,其每个像素点的权重 w_u 将构成一个 $h \times w$ 的权重矩阵,且对于每张无标签图像,权重矩阵都会因模型对该像素点的预测结果不同而不同。

因此,模型间分歧损失 L_u 定义为:

$$L_u = \frac{1}{N} \sum_{u, y_A \in D''} w_u \text{BCE}(y_A, B(u)), \quad (2)$$

其中,BCE是二元交叉熵损失函数,其定义为:

$$\text{BCE} = - \sum_{c=1}^2 y_c \log(p_c). \quad (3)$$

在互学习中,对于有标签样本,仍使用二元交叉熵损失函数 L_s , L_s 定义为:

$$L_s = \frac{1}{N} \sum_{l, y \in D'} \text{BCE}(y, B(l)). \quad (4)$$

因此,互学习损失函数 L 定义为:

$$L = L_s + L_u. \quad (5)$$

4 实验结果及分析

4.1 实验环境

本文实验平台的处理器为Intel(R) Xeon(R) Gold 6226R CPU@2.90 GHz, RAM容量为256 GB。GPU使用Tesla V100,显存容量为32 GB。Python版本为3.8.13。Pytorch版本为1.13.0。

本文构建的舌体图像数据集共3246张,其中2596张作为训练集,650张作为测试集。实验时将训练集分别按照1/100(25张),1/50(51张),1/25(103张),1/8(324张)的标签比例划分为有标签数据集 D' 和无标签数据集 D'' ,模型的性能

指标统一在测试集上进行评估。

4.2 评价指标

本文采用的评价指标为平均交并比(mean Intersection over Union, mIoU)。IoU是真实值与预测值的交集与并集之比,mIoU是交集与并集之比的平均值,其结果可以直接反映舌体分割模型的性能。mIoU计算公式如式(6)所示:

$$\text{mIoU} = \frac{1}{m+1} \sum_{i=0}^m \frac{p_{ii}}{\sum_{j=0}^m p_{ij} + \sum_{j=0}^m p_{ji} - p_{ii}}, \quad (6)$$

其中: m 为类别个数, p_{ii} 为将第 i 类正确地分割为第 i 类的像素点数量, p_{ij} 为将第 i 类错误地分割为第 j 类的像素点数量, p_{ji} 为将第 j 类错误地分割为第 i 类的像素点数量。

4.3 消融实验及中间结果对比

为了验证本文方法所使用的双模型和互学习损失函数两个策略的有效性,在舌体图像分割数据集上进行消融实验。

本文共设计4组消融实验。第一组不使用双模型和互学习损失函数,仅使用单一模型和二元

交叉熵损失函数。单一模型在有标签数据上训练完成后,以对无标签数据生成伪标签并进行微调的方式迭代训练两次。第二组使用双模型构建舌体图像分割网络,但损失函数不使用互学习损失函数,仅使用二元交叉熵损失函数,即 $L = L_s$ 。第三组损失函数增加模型分歧损失,即 $L = L_s + L_u$,但是不采用双模型互学习的方式,仅使用单一模型。单一模型在有标签数据集训练完成后,以对无标签数据生成伪标签并进行微调的方式迭代训练两次。第四组即为本文方法,双模型迭代训练和分歧损失均用于构建模型。

实验结果如表1所示。消融实验结果表明,使用双模型和互学习损失函数都对模型性能提升有积极作用。与不采用任何一种策略相比,使用双模型在1/100、1/50、1/25、1/8标签数据比例下mIoU分别提升了2.92%、2.77%、2.86%和2.16%;使用互学习损失函数的mIoU分别提升了3.12%、3.26%、2.7%和1.98%;同时使用双模型和互学习损失函数时,mIoU分别提升了3.46%、3.4%、3.39%和2.66%。

表1 不同策略下的模型性能(mIoU)对比

Tab. 1 Comparison of model performance (mIoU) under different strategies

是否使用双模型	是否使用互学习损失函数	1/100	1/50	1/25	1/8
×	×	93.21	94.52	95.13	96.19
√	×	96.13	97.29	97.99	98.35
×	√	96.33	97.78	97.83	98.17
√	√	96.67	97.92	98.52	98.85

进一步地,给出本文方法在模型训练及微调过程中的中间实验结果以说明其有效性。以标签数据比例为1/50为例,通过实验记录发现:在有标签数据集上监督训练后,模型A的mIoU为86.88%,模型B的mIoU为87.02%;接着,模型A对无标签数据生成伪标签,模型B在有标签和伪标签数据上再次微调训练后mIoU为97.92%;随后,由模型B再对无标签数据生成伪标签,模型A在有标签和伪标签数据上再次微调训练后mIoU为97.53%,低于模型B的97.92%。因此,最后取两者中性能更佳的模型B作为最终的舌体图像分割模型。图2为数据比例为1/50时伪标签和标签的可视化对比图。可见,由于缺乏训练数据,生成的伪标签在准确性上低于真实标

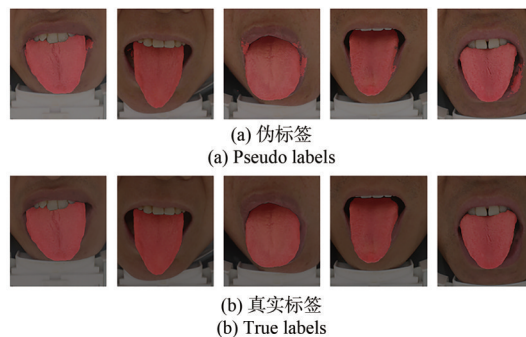


图2 伪标签和真实标签的对比

Fig. 2 Comparison of pseudo labels and true labels

签,甚至存在一些错误,但经过后续的双模型互学习,这些问题在双模型互学习中被消除或缓解,最终能够达到较好的分割效果。

4.4 对比实验

为了验证本文提出的双模型互学习半监督舌体图像分割方法的有效性,在本文构建的舌体图像数据集上进行对比实验。

半监督分割方法可以分为对抗训练方法、一致性正则化方法和伪标记方法。分别选取近年来每类中的典型方法进行对比实验,包括 ClassMix^[17]、Mean Teacher^[26]、CCT^[18]、S4GAN^[15]、ST++^[20]。同时,为了展示出半监督分割方法的优越性,选取全监督语义分割网络 DeepLabv3+ 作为性能比较的基准,训练 DeepLabv3+ 时仅使用有标签数据集。通过实验结果可以看出不同的半监督分割方法在无标签数据集上的有效性。在对比实验中,所有方法均采用相同的训练策略,随机种子设置为 0, epoch 设为 50, batch size 设为 16。本文方法中超参数 γ 设置为 3。

实验结果如表 2 所示,可见半监督方法与监督方法之间的性能差异。各种半监督方法相比于 DeepLabv3+ 模型,均增加了不同的对未标注数据的利用策略,因此,各种方法理应在 DeepLabv3+ 性能基础上有性能提升,以此证明其策略的有效性。然而训练中包含的未标注数据并不总是可以提升性能,相反,在标签数据比例非常小时,未标注数据甚至会损害训练过程,造成模型性能下降。在本文选取的对比方法 CCT 中,当标签数据比例为 1/100 时,其模型性能相比于 DeepLabv3+ 发生了下降。其他半监督方法在所有标签数据比例下均实现了不同程度的性能提升。

表 2 本文方法与其他方法在不同标签比例下的 mIoU 对比
Tab. 2 Comparison of mIoUs between the proposed method and other methods under different proportion of annotated data

方法	1/100	1/50	1/25	1/8
DeepLabv3+ ^[13]	85.53	87.02	91.06	94.88
ClassMix ^[17]	94.74	96.14	97.35	98.09
Mean Teacher ^[26]	87.12	89.52	92.28	95.61
CCT ^[18]	83.71	87.98	92.02	95.37
S4GAN ^[15]	90.31	94.74	96.69	97.61
ST++ ^[20]	94.62	95.09	96.08	97.31
本文方法	96.67	97.92	98.52	98.85

本文方法在 1/100、1/50、1/25、1/8 四个标签比例下均取得了最高的 mIoU,相比于基准(DeepLabv3+)性能实现了最大程度的提升。尤其是当标签数据比例减小时,本文方法相比于其他半监督方法对性能的提升更加明显。可以看出,在标签数据比例极小时(1/100、1/50),本文方法的性能已经可以比肩其他半监督分割方法在标签数据比例为 1/25 和 1/8 的性能,从侧面验证了本文方法对舌体图像无标签数据利用得更充分有效。

4.5 舌体图像分割效果对比

本文方法对舌体图像的分割效果如图 3 所示,其中图 3(a)为原始舌体图像,图 3(b)为舌体图像分割图,为了更清楚地展示舌体图像分割的效果,本文将原始舌体图像图片与模型输出的 mask 融合在原始图片上,如图 3(c)所示。后续均使用分割效果图来对比各方法分割效果。

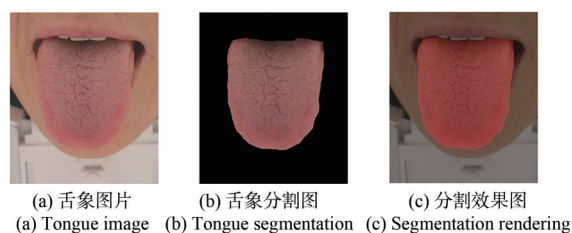


图 3 舌体图像分割示例图

Fig. 3 Example images of tongue segmentation

同时,本文以标签数据为 1/50 为例,选取了 3 张嘴唇与舌体边界不明显的图像,对本文方法与其他半监督方法的舌体图像分割效果进行展示。图 4 展示了对比效果,并分别挑选了一个有代表性的局部区域进行放大,以便更清楚地可视化不同方法的效果差异。

由图 4 可见,本文方法对舌体边缘轮廓分割得更清晰光滑,嘴唇与舌体边界分割更为准确,完整地保留了舌体的边缘轮廓信息,没有出现将嘴唇部位的像素错误分割为舌体像素的情况。本文对比的其他半监督方法所得舌体边缘轮廓处存在明显的毛刺或瑕疵,在舌体边缘轮廓分割的准确性低于本文方法,部分方法还出现了将嘴唇、牙齿或口腔等处像素误分割为舌体像素的情况。这可能与仅使用单一模型进行半监督学习有关,单一模型在学习到错误知识后,在后续的训练过程中无法纠正,导致分割错误的累积。

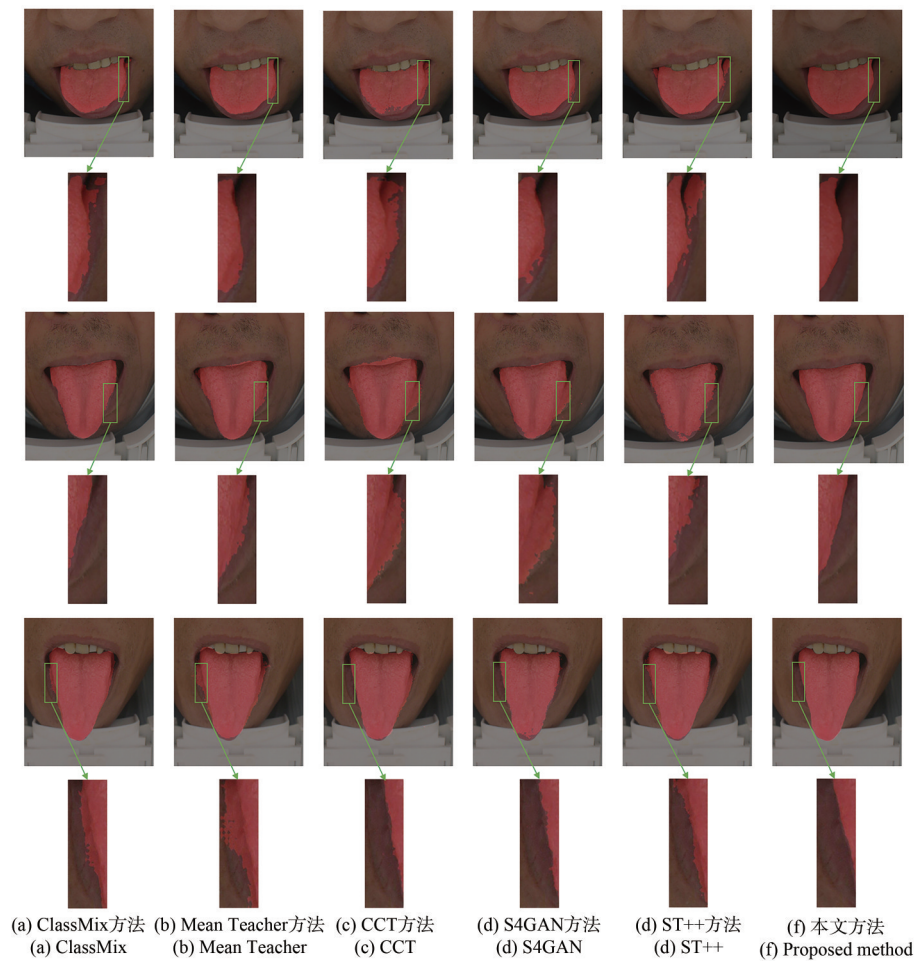


图4 舌体图像分割效果对比

Fig. 4 Comparison of tongue image segmentation effects

5 结 论

目前关于舌体图像分割方法的研究主要使用全监督训练,其需要大量的像素级标注数据,标注这些数据耗时耗力,而现有基于单一模型的半监督分割方法对学习到的错误知识缺乏自我纠正能力。针对这一情况,本文提出一种基于双模型互学习的半监督舌体图像分割方法,仅需极少数标签数据便可实现较好的分割效果。本文方法使用双模型互学习策略,改善了单模型在迭代训练时学习到错误知识的情况。同时,为了有效利用和实现双模型互学习策略,本文重新设计了损失函数,针对双模型对无标签数据预测的分

歧对权重给予不同的加权系数,消融实验证明双模型的使用和互学习损失函数的改进均有效提升了模型的分割性能。实验结果表明,本文方法在1/100、1/50、1/25、1/8四个标签数据比例下的mIoU分别达到96.67%、97.92%、98.52%、98.85%,优于其他典型的半监督方法,尤其在1/100标签数据比例下,即使使用这种极少数的标签数据,本文方法依然取得了较高的性能。本文提出的半监督舌体图像分割方法有着优异的性能,解决了舌体图像分割需要标注大量像素级标签的问题。同时,准确的舌体图像分割为之后的中医舌色、舌形等舌象分析奠定了坚实基础,有望推进中医舌诊客观化研究的进程。

参 考 文 献:

- [1] 卢运西,李晓光,张辉,等. 中医舌象分割技术研究进展:方法、性能与展望[J]. 自动化学报,2021,47(5):1005-1016.

- LU Y X, LI X G, ZHANG H, *et al.* Review on tongue image segmentation technologies for traditional Chinese medicine: methodologies, performances and prospects [J]. *Acta Automatica Sinica*, 2021, 47(5): 1005-1016. (in Chinese)
- [2] 李家炜. 基于深度学习的中医舌象特征分类方法研究[D]. 太原: 中北大学, 2022.
- LI J W. Research on the classification of tongue features of traditional Chinese medicine based on deep learning [D]. Taiyuan: North University of China, 2022. (in Chinese)
- [3] 王翊越. 基于卷积神经网络的舌象诊断研究[D]. 西安: 西安电子科技大学, 2022.
- WANG Y Y. Diagnosis of tongue images based on convolutional neural networks [D]. Xi'an: Xidian University, 2022. (in Chinese)
- [4] 张新峰, 郭宇桐, 蔡轶珩, 等. 基于 DCNN 和全连接 CRF 的舌图像分割算法[J]. 北京航空航天大学学报, 2019, 45(12): 2364-2374.
- ZHANG X F, GUO Y T, CAI Y H, *et al.* Tongue image segmentation algorithm based on deep convolutional neural network and fully conditional random fields [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2019, 45(12): 2364-2374. (in Chinese)
- [5] 颜建军, 徐姿, 郭睿, 等. 基于 Mask R-CNN 的舌图像分割研究[J]. 世界科学技术——中医药现代化, 2020, 22(5): 1532-1538.
- YAN J J, XU Z, GUO R, *et al.* Research on tongue image segmentation based on mask R-CNN [J]. *Modernization of Traditional Chinese Medicine and Materia Medica—World Science and Technology*, 2020, 22(5): 1532-1538. (in Chinese)
- [6] 徐武, 陈盈君, 汤弘毅, 等. 基于 DeepLab 改进网络下的舌图像分割算法[J]. 计算机仿真, 2022, 39(6): 210-213, 489.
- XU W, CHEN Y J, TANG H Y, *et al.* Improved tongue image segmentation algorithm based on DeepLabv2 network and OSIM network [J]. *Computer Simulation*, 2022, 39(6): 210-213, 489. (in Chinese)
- [7] LEE J, KIM E, LEE S, *et al.* FickleNet: weakly and semi-supervised semantic image segmentation using stochastic inference [C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA: IEEE, 2019: 5262-5271.
- [8] ZHUANG X R, LI Y X, HU Y F, *et al.* Self-supervised feature learning for 3D medical images by playing a Rubik's cube [C]//*Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*. Shenzhen, China: Springer, 2019: 420-428.
- [9] MIYATO T, MAEDA S I, KOYAMA M, *et al.* Virtual adversarial training: a regularization method for supervised and semi-supervised learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(8): 1979-1993.
- [10] ZHOU C G, FAN H Y, ZHAO W, *et al.* Reconstruction enhanced probabilistic model for semisupervised tongue image segmentation [J]. *Concurrency and Computation: Practice and Experience*, 2020, 32(22): e5844.
- [11] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015: 3431-3440.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Munich, Germany: Springer, 2015: 234-241.
- [13] CHEN L C, ZHU Y K, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer, 2018: 833-851.
- [14] PELÁEZ-VEGAS A, MESEJO P, LUENGO J. A survey on semi-supervised semantic segmentation [J/OL]. *arXiv*, 2023: 2302.09899.
- [15] MITTAL S, TATARCHENKO M, BROX T. Semi-supervised semantic segmentation with high- and low-level consistency [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(4): 1369-1379.
- [16] CHAPELLE O, SCHOLKOPF B, ZIEN A. Semi-supervised learning [J]. *IEEE Transactions on Neural Networks*, 2009, 20(3): 542.

- [17] OLSSON V, TRANHEDEN W, PINTO J, *et al.* ClassMix: segmentation-based data augmentation for semi-supervised learning [C]//*Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa, HI, USA: IEEE, 2021: 1368-1377.
- [18] OUALI Y, HUDELOT C, TAMI M. Semi-supervised semantic segmentation with cross-consistency training [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 12671-12681.
- [19] LIU Y Y, TIAN Y, CHEN Y H, *et al.* Perturbed and strict mean teachers for semi-supervised semantic segmentation [C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA, USA: IEEE, 2022: 4248-4257.
- [20] YANG L H, ZHUO W, QI L, *et al.* ST++: make self-training work better for semi-supervised semantic segmentation [C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA, USA: IEEE, 2022: 4258-4267.
- [21] LI H L, ZHENG H C. A residual correction approach for semi-supervised semantic segmentation [C]//*Proceedings of the 4th Chinese Conference on Pattern Recognition and Computer Vision*. Beijing, China: Springer, 2021: 90-102.
- [22] KE Z H, WANG D Y, YAN Q, *et al.* Dual student: breaking the limits of the teacher in semi-supervised learning [C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE, 2019: 6727-6735.
- [23] ZHANG Y, XIANG T, HOSPEDALES T M, *et al.* Deep mutual learning [C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 4320-4328.
- [24] 方超伟,李雪,李钟毓,等. 基于双模型交互学习的半监督医学图像分割[J]. 自动化学报,2023,49(4):805-819.
FANG C W, LI X, LI Z Y, *et al.* Interactive dual-model learning for semi-supervised medical image segmentation [J]. *Acta Automatica Sinica*, 2023, 49(4): 805-819. (in Chinese)
- [25] ZHOU Y, JIAO R Y, WANG D L, *et al.* Catastrophic forgetting problem in semi-supervised semantic segmentation [J]. *IEEE Access*, 2022, 10: 48855-48864.
- [26] TARVAINEN A, VALPOLA H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates Inc., 2017: 1195-1204.

作者简介:



李方旭,男,硕士研究生,2021年于河南工业大学获得学士学位,主要从事图像处理、图像语义分割、深度学习等方面的研究。E-mail:li_fangxu@163.com



贾云,女,硕士,主治医师,2009年于华中科技大学获得硕士学位,主要从事中西医结合临床方面的工作。E-mail:jiayun@cug.edu.cn



徐望明,男,博士,正高级实验师,2013年于武汉科技大学获得博士学位,主要从事图像处理与模式识别等方面的研究。E-mail:xuwangming@wust.edu.cn