

# 基于深度学习的舌象自监督聚类

罗思言<sup>1</sup>, 王心舟<sup>2</sup>, 饶向荣<sup>1</sup>

1. 中国中医科学院广安门医院肾病科, 北京 100053; 2. 同济大学电子与信息工程学院, 上海 201804

**【摘要】目的:**解决人工智能舌诊领域数据标注成本较高且带有较强主观性的问题。**方法:**基于深度学习中的对比学习技术,对舌象进行自监督聚类。该方法首先利用卷积神经网络将不同数据增强模式下的舌象映射到潜在空间,并在学习同类实例之间共同特征的同时尽可能区分非同类实例;随后利用高斯混合模型对网络提取的特征向量进行聚类。**结果:**在无需引入先验知识的情况下,利用300张舌象仪采集的无标签图像取得了52.54%的聚类纯度。**结论:**该方法一定程度上将医疗工作者从费事费力的数据标注工作中解放出来。除应用于自动化舌象分类外,该方法还可进一步针对不同病症的特殊舌象症候群进行聚类分析,其提取的舌象特征也可作为舌体分割、舌色分类、苔质分区等下游任务提供预训练的参考。

**【关键词】**自监督学习;中医舌诊;图像聚类;高斯混合模型

**【中图分类号】**R318;TP18

**【文献标志码】**A

**【文章编号】**1005-202X(2023)01-0120-06

## Self-supervised clustering of tongue images based on deep learning

LUO Siyan<sup>1</sup>, WANG Xinzhou<sup>2</sup>, RAO Xiangrong<sup>1</sup>

1. Department of Nephropathy, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China; 2. College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

**Abstract: Objective** To solve the problem of high cost of data annotation and evident subjectivity in the artificial intelligence-based tongue diagnosis. **Methods** The proposed method adopted the contrastive learning of deep learning for realizing the self-supervised clustering of tongue images. The convolutional neural network was used to map tongue images under different data enhancement patterns into the latent space, and to distinguish non-similar instances as much as possible while learning the common features among similar instances. The feature vectors extracted by the network were subsequently clustered using Gaussian mixture model. **Results** A clustering accuracy of 53% was obtained using 300 unlabeled images collected by the instrument for tongue image analysis without introducing prior knowledge. **Conclusion** The method somewhat frees medical practitioners from the laborious and tedious task of data labeling. In addition to its application to automated tongue classification, the method can further be used for the clustering analysis of special tongue syndromes of different diseases; and its extracted tongue features can provide pre-trained references for downstream tasks such as tongue segmentation, tongue color classification, and moss texture partitioning.

**Keywords:** self-supervised learning; tongue diagnosis; image clustering; Gaussian mixture model

## 前言

中医至今已发展了数千年,并取得了辉煌的成果,舌诊是望闻问切中望诊的重要环节<sup>[1]</sup>,其非侵入性和便捷性使得其获得广泛的应用<sup>[2-5]</sup>。随着医疗系统的信息化程度不断提高、医疗数据的不断积累,人

工智能舌诊领域也取得了飞速的发展<sup>[6-8]</sup>。目前人工智能舌诊主要聚焦于舌体分割、舌象分类、苔质分离等下游任务。得益于其对图像输入的稳健性、可移植性等特点,基于卷积神经网络(Convolutional Neural Network, CNN)的深度学习模型在大多数任务上的性能已经超越了传统的图像处理算法。在舌体分割任务上,如Mask R-CNN<sup>[9]</sup>、UNet<sup>[10]</sup>等深度学习算法能够以媲美人类的精度将舌体与外部环境(如牙齿、面部、嘴唇等)区分,并取得91%的平均交并比精度。舌象分类是深度学习应用较为成功的一个方向,Li等<sup>[11]</sup>利用深度学习模型Yolo V3对舌色、苔色、齿痕舌、裂纹舌等舌象属性进行分类,并取得

**【收稿日期】**2022-06-22

**【基金项目】**国家自然科学基金(81973683)

**【作者简介】**罗思言,硕士研究生,研究方向:中西医结合, E-mail: luosiyang8888@163.com

**【通信作者】**饶向荣,主任医师,教授,研究方向:中西医结合, E-mail: raoyisheng@163.com

92.63%的准确率;Wang等<sup>[12]</sup>利用ResNet对齿痕舌进行分类,达到91.63%的准确率;芮迎迎等<sup>[13]</sup>将迁移学习与Mask Scoring R-CNN结合起来,识别齿痕舌达到了93%的准确率。相较于前两者,舌苔舌质分离任务是一个更加细粒度的任务,常见的分离方法是基于颜色空间的聚类方法。韩立博等<sup>[14]</sup>将伽马校正与K-means聚类算法相结合,在CIELAB颜色空间进行苔质分离。李兆龙等<sup>[15]</sup>则将颜色空间特征的选取进一步优化,并利用K-means进行苔质分离。

现有研究方法往往依赖于人工标注数据或人工设计特征<sup>[16-17]</sup>,相较于传统的数据标注,中医图像的标注成本更加昂贵,且其一定程度上也受到数据标注者主观因素的影响,这与中医客观化的宗旨相违背。因此,本研究将深度学习领域的图像自监督聚类算法引入中医舌诊领域。自监督算法是无监督算法的一个分支,旨在没有标签的情况下学习数据特征并应用于下游应用。目前中医领域的无监督聚类算法主要基于传统的信号处理或机器学习算法,如冯冰等<sup>[18]</sup>利用小波变换对脉诊信号进行无监督聚类;张世君等<sup>[19]</sup>探讨了主成分分析、聚类分析等方法在中医证候规范化中的运用,而对于图像无监督聚类这一方法在中医中的应用尚处在探索阶段。

本研究提出一个端到端的舌象自监督聚类流程,通过对比学习自监督学习舌象的流形,并利用高斯聚类模型将潜在空间中的舌象特征进行聚类,该方法达到53%的分类纯度。此外该模型也可为其他下游任务,如舌象分类、点刺检测等提供编码器或者预训练权值。

1 实验方法

本研究算法流程图如图1所示,主要分为3个阶段:(1)利用基于深度学习模型Swin-Transformer<sup>[20]</sup>的舌体分割网络将舌体与嘴唇、面部、牙齿等无关因素分离开,得到舌体图像;(2)将舌体图像传入SimCLR<sup>[21]</sup>模型,利用对比学习将无标签的舌体图像映射到潜在空间,得到特征向量;(3)利用高斯混合模型将不同舌象对应的特征向量分为不同的簇,以达到无监督聚类的效果。

1.1 数据采集与标注

本研究的数据采用哈尔滨工业大学提供的公开数据集<sup>[22]</sup>,该数据包含300张768×576像素的RGB舌象图像(<https://github.com/BioHit/TongueImageDataset>)。数据采集设备为带有内部照明设备和相机的半封闭暗箱。照明设备采用CIE(Commission Internationale de l'Eclairage)推荐的D50自然光照明设备,得益于其封闭稳定的照片环境,该设备采集的图像无需进行色彩或白平衡校准。4名志愿者对舌体边缘进行仔细的标注,

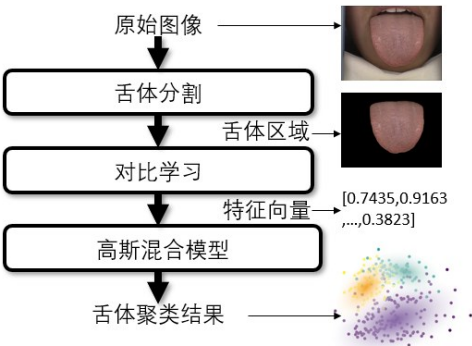


图1 舌象无监督聚类流程图  
Figure 1 Unsupervised clustering of tongue images

并选取其中最准确的一个标注作为数据集的标签。此外,为了验证本文自监督聚类算法对舌象属性分类的准确性,3名具有执业医师证的研究生对每张图片的舌象进行标注,标注的8种属性包括:正常、点刺、裂纹、齿痕、厚腻苔、老舌、嫩舌及淤点。

1.2 舌体分割

舌体分割是舌诊客观化方法的第一步。舌体分割可以将舌体区域从包含嘴唇、牙齿、面部等无关信息的图像中分离出来,以便于后续算法对图像进行进一步的处理。舌体分割模型的训练方式基于迁移学习,采用预训练-微调的方式进行。在微软提供的预训练模型基础上,以舌象数据集中的300张舌象图像作为输入,其对应的舌体区域作为标签,对模型进行进一步微调,网络结构如图2所示。其关键部分为基于注意力机制的Swin-Transformer编码器及UperNet解码器。相较于传统的卷积神经网络,Transformer通过自注意力机制可以更好地对图像有一个全局性的认识,从而以更高的精度分割舌体与其他无关区域。

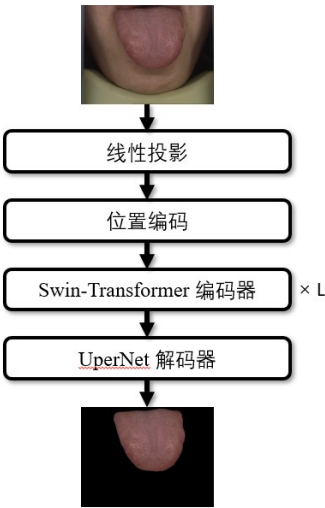


图2 舌体分割网络结构  
Figure 2 Tongue segmentation network structure

### 1.3 自监督对比学习

近年来计算机视觉快速发展一个非常关键的因素便是海量带标注数据集的发布。大量高质量的标注将直接为计算机视觉模型带来性能提升。然而图像的标注往往需要大量的时间和精力,作为专业性更强的医学图像,其标注成本则更加昂贵。在医学领域,每时每刻都在产生大量以文字、数据、图像、音频等形式存储的医疗数据,而其标注的困难性直接制约了绝大多数医学数据在人工智能领域的应用。

对比学习是一种表征学习的方式,其目的是在无标注的情况下依据图像在某些数据增强模式下的不变性,为对应的图像元组打上正或负的伪标签,并依此进行自监督学习。本文采用Google SimCLR模型作为对比学习框架,学习舌象的视觉表征。SimCLR的网络结构如图3所示。首先,将分割后的舌象图像  $x$  输入两个不同的图像增强算子 (Augmentation Operator) 当中,两个图像增强算子  $t$  以及  $t'$  均由图像增强算子族  $T$  中采样得到。 $T$  由随机裁剪、翻转、缩放、色彩失真以及高斯模糊组成,其不同增强方式的内部参数均限定在一定范围内。随后将同一舌象在经过不同增强变换后得到的图像  $\tilde{x}$  以及  $\tilde{x}'$  输入同一个图像编码网络 (Encoder Network), 得到二者在潜在空间上的投影向量  $h$  及  $h'$ 。本文采用预训练残差连接神经网络 ResNet<sup>[23]</sup> 作为图像编码网络。 $h$  及  $h'$  即为对比学习想要获得的图像特征向量。但需要注意的是实验表明如果直接将  $h$  及  $h'$  代入损失函数进行一致性的判断,学到的特征质量非常有限<sup>[21]</sup>, 因此在该步骤之后还加入了一个可学习的非线性变化层对特征向量进行进一步映射。本文选用多层感知机 (Multilayer Perceptron, MLP) 作为非线性映射模型,并使用线性整流函数 (Rectified Linear Unit, ReLU) 作为激活函数以引入非线性。在每一轮训练时,从数据集中抽样出  $N$  张图像,在经过图像增强变换后得到  $2N$  个样本。对每一个样本来说,与其来源于同一张原始图像的样本作为正样本,其余  $2(N-1)$  个样本作为负样本,则对比损失函数<sup>[24]</sup>为:

$$l_{ij} = -\log \frac{\exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right)}{\sum_{k=1}^{2N} 1_{[k \neq i]} \exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right)} \quad (1)$$

式中,  $1_{[k \neq i]} \in \{0, 1\}$  代表值域是 0 或 1 的指示函数, 当且仅当  $k \neq i$  时其值为 1;  $\tau$  为温度参数;  $\text{sim}(z_i, z_j)$  为两特征向量相似度, 本文采用余弦相似度, 其定义为:

$$\text{sim}(z_i, z_j) = \frac{z_i^T z_j}{\|z_i\| \|z_j\|} \quad (2)$$

SimCLR 可以在没有图像标签的情况下隐式地学习舌象的内在特征, 如颜色、形状、淤点、裂纹等, 并将其学习结果映射为特征向量。该特征向量是舌象的高级特征在数学上的具体化, 对特征向量进行聚类等价于根据舌象的高级特征进行分类。训练时利用非线性映射函数对特征  $h$  进行变换, 辅助编码网络学习。测试时则将该部分丢弃, 只利用编码网络进行推理。

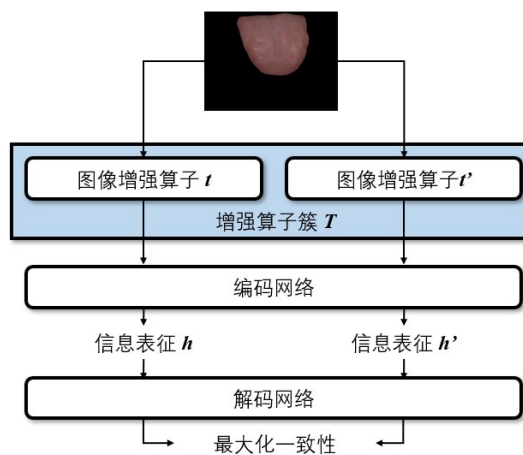


图3 对比学习网络结构

Figure 3 Contrastive learning network structure

### 1.4 舌象特征聚类

高斯混合模型是一种统计学习模型, 模型假设数据服从如下的概率分布:

$$P(x|\theta) = \sum_{k=1}^K \alpha_k \phi(x|\theta_k) \quad (3)$$

其中,  $K$  为模型中子高斯模型的数目,  $\alpha_k$  是观测数据属于第  $k$  个子模型的概率,  $\phi(x|\theta_k)$  为高斯函数,  $\theta_k$  为高斯函数的参数, 即均值  $\mu$  和协方差  $\Sigma$ 。多维输入的高斯函数形式如下所示:

$$P(x|\theta) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2}\right) \quad (4)$$

本文中  $x$  为利用自监督对比学习提取的舌象高级特征, 即一个 128 维的特征向量。舌象特征作为一种非人为属性, 假设其服从高斯分布是合理的。该模型中  $\alpha_k$  及  $\theta_k$  为需要学习的参数, 采用 EM (Expectation-Maximum) 算法可根据数据集迭代学习模型参数, 其步骤如下:

(1) 初始化参数  $\alpha_k$  及  $\theta_k$ 。

Expectation step. 根据数据集计算给定参数的似然性:



$$\gamma_{jk} = \frac{a_k \phi(\mathbf{x}_j | \theta_k)}{\sum_{k=1}^K a_k \phi(\mathbf{x}_j | \theta_k)}, j = 1, 2, \dots, N; k = 1, 2, \dots, K \quad (5)$$

Maximization step。找到使得似然性最大的参数：

$$\mu_k = \frac{\sum_j^N (\gamma_{jk} \mathbf{x}_j)}{\sum_j^N \gamma_{jk}}, k = 1, 2, \dots, K \quad (6)$$

$$\Sigma_k = \frac{\sum_j^N \gamma_{jk} (\mathbf{x}_j - \mu_k)(\mathbf{x}_j - \mu_k)^T}{\sum_j^N \gamma_{jk}}, k = 1, 2, \dots, K \quad (7)$$

$$\alpha_k = \frac{\sum_j^N \gamma_{jk}}{N}, k = 1, 2, \dots, K \quad (8)$$

(2)重复直至  $\alpha_k$  及  $\theta_k$  收敛。

在舌象特征向量数据集上利用EM算法学习GMM模型的各项参数,学习完成后舌象特征向量被分为  $K$  个簇,每个簇内的舌象在高级特征上具备相似性。

## 2 结果

### 2.1 实验平台

神经网络模型的训练和测试都在基于Python的深度学习框架PyTorch(<https://pytorch.org>)上进行,高斯混合模型则基于机器学习工具库Scikit-learn(<http://scikit-learn.github.io/stable>)进行开发。计算平台为基于Ubuntu 16.04 LTS操作系统的服务器。该服务器备有 Intel Xeon(R) E5-2620 CPU, 4 张 NVIDIA RTX2080Ti GPU,以及 128 GB的DDR5内存。

### 2.2 实施细节

首先利用微软在包含超过 20 000 张常见物体图片的ADE20K<sup>[25]</sup>数据集上训练的Swin-Transformer作为预训练模型,在舌象数据集上对其语义分割能力进行进一步微调,其中 240 张图片作为训练集,60 张图片作为测试集。将微调后模型分割得到的结果作为新的输入,对SimCLR中的ResNet解码器及MLP映射模型的权值进行训练直至模型收敛。两个模型的参数设置如表1所示。

随后利用GMM对SimCLR抽取分割后舌象得到的128维特征向量集合进行无监督聚类。其超参数采用网格搜索(Grid Search)进行优化,得到的最优参数如表2所示。

### 2.3 结果分析

利用Swin-Transformer对舌体进行分割的结果如图4所示。Swin-Transformer对图片全局语义信息的理解使得其非常适合于进行舌体分割,其交并比

表1 神经网络参数设置

Table 1 Neural network parameter setting

超参数	Swin-Transformer	SimCLR
循环数	20	100
Batch 大小	4	32
优化器	AdamW	LARS
学习率	6e-5	0.375
学习率策略	多项式	多项式
权值惩罚	1e-2	1e-6
损失函数	交叉熵	NT-Xent <sup>[24]</sup>

表2 高斯混合模型参数设置

Table 2 Parameter setting of Gaussian mixture model

超参数	GMM
混合模型数目	9
协方差种类	full
收敛阈值	1e-3
非负正则化	1e-6
EM迭代次数	3 000
参数初始化	K-means
热启动	false

达到了99.47%,可以看到舌体边缘非常清晰,且没有外部离群区域的出现。

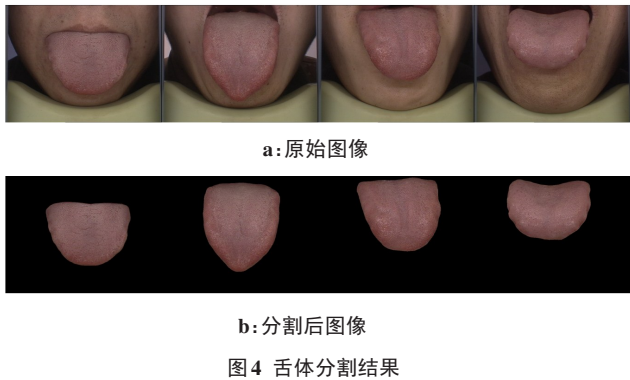


Figure 4 Tongue segmentation results

将舌体分割的结果放入SimCLR中进行自监督对比学习,并利用GMM对特征向量集合进行聚类,根据设置的子高斯模型部件数量,可将300张舌象图片分为  $K=8$  类。纯度(Purity)是无监督聚类学习常用的评价指标,其定义如下：

$$\text{purity} = \sum_{i=1}^K \frac{m_i}{m} \max(p_{ij}) \quad (9)$$

其中,  $m$  为样本数目,  $m_i$  为第  $i$  类中的样本数目,  $p_{ij}$  为

第*i*类中成员具有属性*j*的概率。本文的高斯混合模型聚类方法可以达到 52.54% 的纯度,而对于每个子模型,其纯度  $\max(p_{ij})$  如表 3 所示。作为对比,采用 K-means 对特征向量进行聚类,则只可达到 30.60% 的纯度,每个簇的纯度如表 4 所示。本文将舌象划分为 8 类,若进行随机标注,则纯度期望值为 12.50%,远低于利用本文提出的算法。每个子模型在没有监督的情况下自发地将舌体依据不同特征聚类起来,如图 5 所示。可以看到尽管部分舌象同时具备多种属性,无监督模型会抓住其中最有价值的特征。从特征空间的角度上来看,模型会将特征向量中与数据集整体特征分布相似性最低的维度作为分类依据,对其对应舌象进行聚类。该方法擅长识别舌体的形态(如齿痕舌)以及舌体上的异常区域(如点刺、裂纹、厚腻舌),而对语义信息层级更高的特征(如老舌、嫩舌)或易混淆特征(如淤点)效果欠佳。经分析其原因有 3 点。首先,SimCLR 采用的图像增强方式为随机裁剪、翻转、缩放、色彩失真以及高斯模糊,这使得其编码更加注重局部信息在变换下的不变性及对噪声干扰的鲁棒性。此外本文的数据规模较为有限,且类别标签不平衡,这导致不同舌象的先验概率不均衡,此时模型参数具有一定的倾向性,基于后验估计的无监督聚类结果也因此受到影响。最后,淤点和点刺在图像特征上较为相似,而数据集中点刺舌的数目远大于淤点舌的数目,因此模型容易将点刺与淤点混淆。对聚类结果中重复出现的类别为点刺及裂纹的子模型进行分析,发现其中的图像除开其最主要的特征之外,次要特征也存在一定的聚类性。子模型进一步的分类为:齿痕、点刺、裂纹、点刺+厚腻苔、厚腻苔、裂纹+点刺、点刺+裂纹+厚腻苔、点刺+厚腻苔。

表 3 高斯混合模型无监督聚类纯度(%)

Table 3 Gaussian mixture model unsupervised clustering purity (%)

子模型								
	1	2	3	4	5	6	7	8
类别	齿痕	点刺	裂纹	点刺	厚腻苔	裂纹	点刺	点刺
纯度	49.88	30.88	39.77	65.87	18.92	23.88	28.90	43.15

3 总结与展望

本文首先利用语义分割网络 Swin-Transform 进行舌体分割,并将结果输入无监督对比学习框架 SimCLR 进行无监督表征学习。对比学习在高维特

表 4 K-means 无监督聚类纯度(%)

Table 4 K-means unsupervised clustering purity (%)

簇								
	1	2	3	4	5	6	7	8
类别	点刺	点刺	裂纹	厚腻苔	点刺	点刺	点刺	点刺
纯度	40.53	41.22	18.98	19.54	23.52	30.48	30.70	38.23

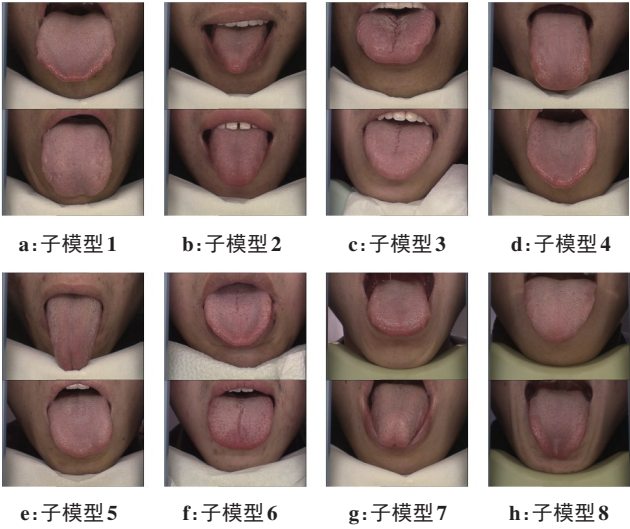


图 5 无监督聚类结果

Figure 5 Unsupervised clustering results

征空间对同一图片不同数据增强模式下的特征向量进行度量学习,保证其在特征空间的欧氏距离较小的同时,尽可能将不相关实例投影到距离较远处。最后利用高斯混合模型进一步对学习到的舌象高维特征进行聚类分析。

本文聚焦于医学图像标注这一制约人工智能算法在医学领域应用的瓶颈,探索图像无监督聚类在中医舌诊当中的应用,其结果证明无监督学习在医学领域具有广阔应用前景,也在一定程度上将医疗工作者从费事费力的数据标注工作中解放出来。目前各大医院均具备较为成熟的医疗数据库,而本文的方法以一种几乎零成本的方式,在无需数据标注的情况下,便可对舌象数据库中的不同舌象进行聚类,并达到一定精度。除应用与自动化舌象分类外,该方法还可对针对不同病症的特殊舌象症候群进行聚类分析。SimCLR 学习得到的模型权值也可迁移到如舌体分割、舌色分类、苔质分区等下游任务之上。本文证明无监督聚类在舌象数据上的可行性,未来研究可在如下几个方向进行:首先是如何进一步配置图像增强方法,以及如何改善数据集不平衡问题,以提高模型在不同舌象特征下的精度,使其具备真正投入临床使用的可能。其次是模型鲁棒性问题,如何保证模型在舌象采集设备发生变化时保持其性能。最后是迁移性问题,以怎么样的形式组织

与利用源域与目标域数据,将本文学习到的预训练模型迁移到其他任务上。

## 【参考文献】

- [1] 王忆勤. 中医诊断学[M]. 北京: 高等教育出版社, 2012.  
Wang YQ. Diagnostics of Chinese medicine[M]. Beijing: Higher Education Press, 2012.
- [2] 周胜, 曹海燕, 吴传良, 等. 胃溃疡患者客观化舌象与胃内炎症反应的相关性研究[J]. 广西中医药, 2016, 39(5): 15-16.  
Zhou S, Cao HY, Wu CL, et al. Study on the correlation between objective tongue image and gastric inflammation in patients with gastric ulcer[J]. Guangxi Traditional Chinese Medicine, 2016, 39(5): 15-16.
- [3] 许岚, 陈曙, 张云静, 等. 从中医舌诊客观化分析慢性乙型肝炎常见证型[J]. 辽宁中医杂志, 2014, 41(9): 1817-1819.  
Xu L, Chen W, Zhang YJ, et al. Analysis on common syndrome of chronic hepatitis B patients by objectifying tongue diagnosis of traditional Chinese medicine[J]. Liaoning Journal of Traditional Chinese Medicine, 2014, 41(9): 1817-1819.
- [4] 阙翼. 冠心病稳定期患者舌象特征及其客观测量比较研究[D]. 北京: 中国中医科学院, 2018.  
Que Y. Comparative study of tongue features and objective measurements in patients with stable coronary heart disease[D]. Beijing: China Academy of Chinese Medical Science, 2018.
- [5] 王辰光. 糖尿病肾病患者舌象特征及其相关性研究[D]. 北京: 北京中医药大学, 2020.  
Wang CG. Study on the characteristics and correlation of tongue image in diabetic nephropathy patients[D]. Beijing: Beijing University of Chinese Medicine, 2020.
- [6] 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 YX, Li XG, 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.
- [7] 丁成华, 杜建强, 朱明峰, 等. 基于图像分析技术探索中医舌诊研究[J]. 江西中医药, 2013, 44(11): 7-9.  
Ding CH, Du JQ, Zhu MF, et al. Exploring tongue diagnosis in chinese medicine based on image analysis technology[J]. Jiangxi Journal of Traditional Chinese Medicine, 2013, 44(11): 7-9.
- [8] 朱培超, 潘锡明, 阮亚君, 等. 探讨以人工智能诊断输出为目的的中医舌诊与病性证素关系模型构建[J]. 环球中医药, 2021, 14(6): 1033-1038.  
Zhu PC, Pan CM, Ruan YJ, et al. Explore the construction of a model of the relationship between TCM tongue diagnosis and disease syndrome elements for the purpose of artificial intelligence diagnosis output[J]. Global Traditional Chinese Medicine, 2021, 14(6): 1033-1038.
- [9] 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 JJ, Xu Z, Guo R, et al. Research on tongue images segmentation based on Mask R-CNN[J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2020, 22(5): 1532-1538.
- [10] 李宗润, 李炜弘, 井洪. 基于多种机器视觉技术的舌体分割模型研究[J]. 时珍国医国药, 2021, 32(7): 1768-1771.  
Li ZR, Li WH, Jing Q. Research on Tongue Images Segmentation based on multiple machine vision technologies[J]. Lishizhen Medicine and Materia Medica Research, 2021, 32(7): 1768-1771.
- [11] Li C, Zhang D, Chen S. Research about tongue image of traditional chinese medicine based on artificial intelligence technology[C]//2020 IEEE 5th ITOEC. 2020: 633-636.
- [12] Wang X, Liu J, Wu C, et al. Artificial intelligence in tongue diagnosis: using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark[J]. Comput Struct Biotech, 2020, 18: 973-980.
- [13] 芮迎迎, 孔祥勇, 刘亚楠, 等. 基于Mask Scoring R-CNN的齿痕舌象识别[J]. 中国医学物理学杂志, 2021, 38(4): 523-528.  
Rui YY, Kong XY, Liu YN, et al. Tooth-marked tongue recognition using Mask Scoring R-CNN[J]. Chinese Journal of Medical Physics, 2021, 38(4): 523-528.
- [14] 韩立博, 胡广芹, 张新峰, 等. 基于直方图均衡化的伽马校正和K-means聚类的舌象苔质分离方法[J]. 北京生物医学工程, 2019, 38(1): 1-6.  
Han LB, Hu GQ, Zhang XF, et al. Separation method of tongue coating and body of tongue image based on histogram equalization and gamma correction and K-means clustering[J]. Beijing Biomedical Engineering, 2019, 38(1): 1-6.
- [15] 李兆龙, 苏育挺. 一种基于聚类的舌苔苔质分离方法[J]. 南开大学学报(自然科学版), 2017, 50(4): 52-56.  
Li ZL, Su YT. A method for separating tongue coating and tongue quality based on clustering[J]. Acta Scientiarum Naturalium Universitatis Nankaiensis, 2017, 50(4): 52-56.
- [16] 赵露, 杨燕. 161例婴儿肝炎综合征症状与舌象指纹聚类分析[J]. 辽宁中医药大学学报, 2013, 15(6): 154-156.  
Zhao Q, Yang Y. Cluster analysis on 161 cases with infant hepatitis syndrome and tongue fingerprint[J]. Journal of Liaoning University of Traditional Chinese Medicine, 2013, 15(6): 154-156.
- [17] 薛志鹏, 李记天, 高欢欢, 等. 股骨头坏死舌象特点的聚类分析[J]. 中医正骨, 2020, 32(1): 32-35.  
Xue ZP, Li JT, Gao HH, et al. Tongue manifestation characteristics of osteonecrosis of the femoral head database cluster analysis[J]. The Journal of Traditional Chinese Orthopedics and Traumatology, 2020, 32(1): 32-35.
- [18] 冯冰, 李绍滋. 中医脉诊信号的无监督聚类分析研究[J]. 智能系统学报, 2018, 13(4): 564-570.  
Feng B, Li SZ. Unsupervised clustering analysis of human-pulse signal in traditional Chinese medicine[J]. CAAI Transactions on Intelligent Systems, 2018, 13(4): 564-570.
- [19] 张世君, 齐冬梅, 李运伦, 等. 中医证候规范化研究中无监督方法探讨[J]. 长春中医药大学学报, 2016, 32(1): 1-5.  
Zhang SJ, Qi DM, Li YL, et al. Unsupervised methods in TCM syndrome standardization study[J]. Journal of Changchun University of Chinese Medicine, 2016, 32(1): 1-5.
- [20] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012-10022.
- [21] Chen T, Komblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.
- [22] 崔振超. 计算机舌诊中舌体分割与纹理分类研究[D]. 哈尔滨: 哈尔滨工业大学, 2015.  
Cui ZC. Tongue body segmentation and texture classification in computerized tongue diagnosis[D]. Harbin: Harbin Institute of Technology, 2015.
- [23] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE CVPR. 2016: 770-778.
- [24] Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv e-prints, 2018: arXiv: 1807.
- [25] Zhou B, Zhao H, Puig X, et al. Semantic understanding of scenes through the ADE20K dataset[J]. Int J Comput Vision, 2016, 127(3): 302-321.

(编辑: 陈丽霞)