

DOI:10.3969/j.issn.1005-202X.2024.04.018

医学人工智能

基于度量学习的多分支舌象识别网络

任思羽¹, 吴瑞², 罗庆林³, 肖开慧⁴, 王艺凡², 利节²

1. 成都开放大学教学部, 四川 成都 610000; 2. 重庆科技学院智能技术与工程学院, 重庆 401331; 3. 重庆科瑞制药有限公司, 重庆 400060; 4. 首都医科大学附属北京友谊医院中医科, 北京 100050

【摘要】为提升舌象识别效率与精准度,通过度量学习研究辅助医生识别舌象表征的方法。首先,收集舌诊图像111例,数据按照7:3的比例随机分为训练集和测试集。然后,设计一种基于度量学习的多分支舌象识别网络。深度学习网络被分为两个部分,前半部分为共享权重层,采用基于度量学习的舌象特征编码损失函数,以获得精准的特征;后半部分针对中医舌象的分类分为4个舌象识别辅助分支,降低舌象识别难度,提升准确率。此外,构建多标签残差映射,增加类间距,减小类内距,提升最终识别的准确度。本文方法在舌象数据集的测试集上进行测试时获得84.8%的识别精度,表明多分支网络架构可以很好地降低舌象识别难度,特别是特征类别较多的舌形和苔质。同时,舌象特征编码损失函数可以有效地提取舌象特征;舌象多标签残差映射可以减少各类别之间的干扰,从而提升识别准确度。

【关键词】舌象识别; 多分支网络架构; 特征编码; 损失函数; 多标签残差映射

【中图分类号】R318; R241.25

【文献标志码】A

【文章编号】1005-202X(2024)04-0521-08

Metric learning based multi-branch network for tongue manifestation recognition

REN Siyu¹, WU Rui², LUO Qinglin³, XIAO Kaihui⁴, WANG Yifan², LI Jie²

1. Teaching Department of the Open University of Chengdu, Chengdu 610000, China; 2. School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China; 3. Chongqing Kerui Pharmaceutical Co., Ltd., Chongqing 400060, China; 4. Department of Traditional Chinese Medicine, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China

Abstract: Based on metric learning, a novel method of assisting doctors in identifying tongue manifestation is proposed to improve the efficiency and accuracy of tongue manifestation recognition. A total of 111 tongue images are collected, and the data are randomly divided into training set and test set at a ratio of 7:3. Subsequently, a metric learning based multi-branch tongue manifestation recognition network is designed. The deep learning network is divided into 2 parts. The first part is the shared weight layer which employs metric learning based loss function in tongue manifestation feature coding to obtain accurate features. In order to reduce the difficulty of tongue manifestation recognition and improve the accuracy, the latter part is split into 4 branches for tongue manifestation recognition which correspond to the classification of tongue manifestation in traditional Chinese medicine. Additionally, a multi-label residual mapping is constructed to increase inter-class distance and reduce intra-class distance, so as to enhance the accuracy of final recognition. The proposed method achieves a recognition accuracy of 84.8% on the test set of tongue manifestation dataset, indicating that multi-branch network architecture can lower the difficulties in tongue manifestation recognition, especially for the tongue shape and coating nature with multiple feature categories. The loss function in tongue manifestation feature coding can effectively extract tongue features, while multi-label residual mapping can reduce the interference between different categories, which improves the recognition accuracy.

Keywords: tongue manifestation recognition; multi-branch network architecture; feature coding; loss function; multi-label residual mapping

【收稿日期】2023-12-10

【基金项目】国家科技部“科技助力经济2020”重点专项(SQ2020YFF0405970); 重庆市自然科学基金(cstc2020jcyj-msxmX0683); 重庆市教委科学技术研究项目(KJQN201901507); 重庆科技学院硕士研究生创新计划(ZNYKJCX2022023)

【作者简介】任思羽, 硕士, 副教授, 研究方向: 计算机网络技术、信息安全, E-mail: 528505463@qq.com

【通信作者】利节, 博士, 副教授, 研究方向: 人工智能算法模型、故障诊断、深度学习等, E-mail: jiel@cqust.edu.cn

前言

中医讲究望闻问切四诊合参,其中舌诊在中医诊断上有着举足轻重的地位。五脏六腑都直接或间接地通过经络、经筋与舌象相联系;同时舌还为心之苗、脾之外候;气血津液充养于舌,舌形质及舌色与气血的盈亏和运行有关,舌的润燥与津液的盈亏、输布有关^[1]。舌象主要从以下几个方面进行观察:舌色、舌形、苔色和苔质;然而,由于舌象组成复杂,每个部分又由多种特征组成,且每部分的特征数量不平衡,特征之间的扰动较大,最终使得判别难度进一步提升。同时,舌象图片难以获取,导致数据样本少,训练难度提升,识别准确率不高。现有研究提出多种针对小样本的深度学习方案,如度量学习^[2]、元学习^[3]和迁移学习^[4];同时,多标签分类也有许多相关研究,如二元相关性框架^[5]、标签域增强^[6]和多分支网络^[7]。如何设计舌象识别网络,采用何种小样本检测方案和多标签分类框架,从而达到最好的舌象识别准确率是需要探究的一大问题。

本研究将舌色、舌形、苔色和苔质分开检测,并通过特征图共享与特征图分离方法构建多分支舌象

识别网络,实现舌象中4个部分的单独检测,降低每一部分之间的相互干扰,提升分类准确率;同时设计舌象特征编码损失,生成提取舌象特征的预训练模型,提升各分支的识别准确率;最后设计舌象多标签残差映射空间,提升最终识别准确度。

1 数据集与方法

1.1 构建数据集

在中医舌诊过程中收集舌象图像111例,并对其进行分类。训练与测试数据按7:3的比例划分舌象数据集。数据集的具体分布如表1所示,通过数据分析可以发现:(1)数据集中特征标签分布不均衡,数量多的如苔色类别中的黄苔有58个,数量少的如舌形类别中的老舌只有1个,这为舌象识别带来了挑战;(2)不同类别的舌象的特征个数差异大,其中舌色包含7种特征,舌形包含9种特征,苔质包含12种特征,而苔色只有3种特征,这将会造成特征提取困难;此外,各个舌象的特征区别也比较小,进一步加大识别的难度^[8]。不同类别舌象图像及其意义如图1所示,不同的舌象有不同的临床意义,多个类别携带的特征共同展示人的机理表现。

表1 数据集分布情况
Table 1 Distribution of dataset

类别	特征	特征数量
舌色	淡红舌;淡白舌;红舌;舌尖红;舌两边红赤;绛舌;青紫/暗红舌	28;11;39;2;2;9;21
舌形	老舌;嫩舌;胖大舌;肿胀舌;瘦薄舌;点刺舌;裂纹舌;齿痕舌;瘀点、瘀斑舌	1;10;15;8;8;8;14;18;1
苔质	无苔;少苔;薄苔;厚苔;润苔;滑苔;燥苔;糙苔;腐苔;腻苔;剥苔;偏苔	7;5;24;23;4;7;3;1;10;26;8;3
苔色	白苔;黄苔;灰黑苔	26;58;8

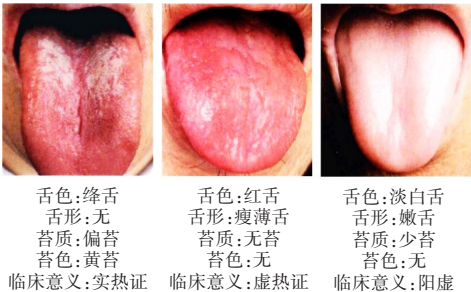


图1 舌象数据集内的部分图像
Figure 1 Some images in the tongue manifestation dataset

1.2 多分支舌象识别基础模型

本研究的多分支舌象识别模型见图2。舌象标签分为舌色、舌形、苔色和苔质,相应地,舌象识别模型分为4个分支。舌象上的舌色和苔色的特征较为

明显,苔质的特征较不清晰,所以使用ResNet18^[9]作为网络的基础架构,残差块可以帮助网络获得浅层信息,防止因网络加深带来的浅层特征丢失。此外,由于舌色和苔色将颜色作为识别根据,识别时不需要细节特征,识别难度较低。为简化模型结构,降低计算资源需求,检测舌色和苔色的分支使用块1和块2生成的特征图,在块2后,网络使用卷积层和最大池化层进行下采样,然后通过全连接层进行识别。舌形和苔质的特征较为复杂,两个分支共有21种标签,识别难度较大。所以,在经过块2后,舌形和苔质两个识别分支继续使用残差块提取特征,以获得更精准的识别结果;之后,使用全连接层进行识别。

舌象图片输入后经过一次下采样并输入由残差块组成的块1和块2。接着,块1和块2作为共享权重层连接后续的4个分支网络,舌形和苔质使用2个由

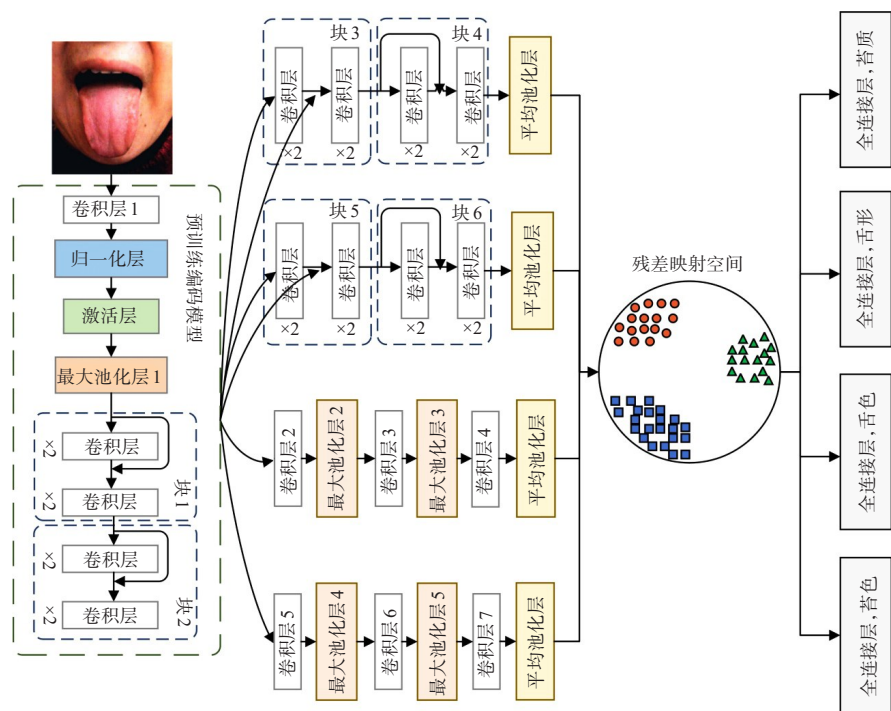


图2 多分支舌象识别模型

Figure 2 Multi-branch tongue manifestation recognition model

残差块组成的分支,舌色和苔色经过由3组卷积层和最大池化层组合的下采样组,最终经由各自的平均池化层和全连接层得到识别结果。具体网络参数见表2。

表2 网络参数表

Table 2 Network parameters

类别	属性	输出
卷积层1	7×7卷积核, 步长2, 填充3	112×112×64
最大池化层1	3×3最大池化	56×56×64
残差块1	{3×3卷积核, 步长1, 填充1}×4	56×56×64
残差块2	3×3卷积核, 步长2, 填充1; {3×3卷积核, 步长1, 填充1}×3	28×28×128
残差块3,5	3×3卷积核, 步长2, 填充1; {3×3卷积核, 步长1, 填充1}×3	14×14×256
残差块4,6	3×3卷积核, 步长2, 填充1; {3×3卷积核, 步长1, 填充1}×3	7×7×512
卷积层2,5	3×3卷积核, 步长2, 填充1	28×28×256
最大池化层2,4	2×2最大池化	14×14×256
卷积层3,6	3×3卷积核, 步长2, 填充1	14×14×512
最大池化层3,5	2×2最大池化	7×7×512
卷积层4,7	3×3卷积核, 步长2, 填充1	7×7×1024

由于舌象识别模型分为4个分支,所以多分支网络的损失函数可以被定义为:

$$L_{\text{Total}} = L_{\text{SS}} + L_{\text{SX}} + L_{\text{TS}} + L_{\text{TZ}} \tag{1}$$

其中, L_{Total} 为整体损失, L_{SS} 为舌色分支损失, L_{SX} 为舌形分支损失, L_{TS} 为苔色分支损失, L_{TZ} 为苔质分支损

失。各个分支的损失函数均为二元交叉熵损失(BCE Loss),定义如下:

$$L_{\text{BCE}} = -\frac{1}{n} \sum_{i=1}^n (y_i \times \ln(\hat{y}_i) + (1 - y_i) \times \ln(1 - \hat{y}_i)) \tag{2}$$

其中, y_i 是真实标签, \hat{y}_i 是预测值。

1.3 舌象特征编码

为了给4个分支提供更便于舌象识别的特征,本研究结合基于度量学习的孪生网络框架,利用样本在不同映射方式下的方差、协方差以及不变性构成舌象特征编码损失 L_{code} 。首先,对于一张给定的舌象图片 I ,通过随机翻转产生新的舌象图片 I' ,将 I 和 I' 分别放入由ResNet-18主干网络组成的编码器 Q 中,生成新的特征图 q 和 q' ,接着将 q 和 q' 分别放入由两层全连接层组成的扩展编码器 eQ 中,生成新的特征图 f 和 f' 。舌象编码损失架构图见图3。

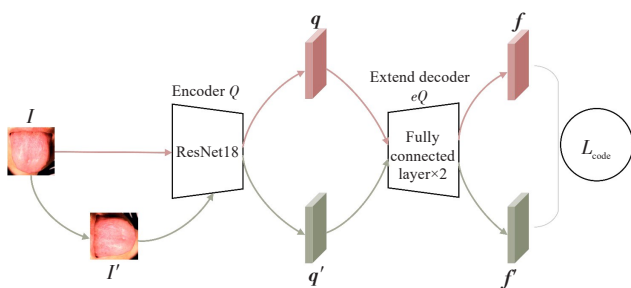


图3 舌象编码损失架构图

Figure 3 Loss architecture of tongue manifestation encoding

损失函数 L_{code} 经由 f 和 f' 计算得到。 L_{code} 可以被定义为:

$$L_{\text{code}} = a[L_v(f) + L_v(f')] + b[L_{\text{cov}}(f) + L_{\text{cov}}(f')] + L_{\text{ed}}(f) \quad (3)$$

其中, $L_v(\cdot)$ 为方差损失, $L_{\text{cov}}(\cdot)$ 为协方差损失, $L_{\text{ed}}(\cdot)$ 为不变性损失, a 和 b 为超参数。

若最终扩展编码器 eQ 的输出 $f = \{x_1, x_2, \dots, x_i\}, i \in n, f' = \{x'_1, x'_2, \dots, x'_i\}, i' \in m$,其中, n 和 m 均为图像分批处理时一个批次的图像数量,且 x_i 和 x'_i 的维度为 d 。方差损失 $L_v(f)$ 表示如下:

$$L_v(f) = \frac{1}{d} \sum_{j=1}^d \max(0, 1 - S(x^j, \varepsilon)) \quad (4)$$

其中, $S(x^j, \varepsilon) = \sqrt{V(x^j) + \varepsilon}$, x^j 为 f 上第 j 维的向量, ε 是一个极小值, $V(x^j)$ 计算 x^j 的方差。 $L_v(f)$ 可以保证 f 内沿着每个维度的方差为1,防止输入映射至向量中包含0导致崩溃。

此外,协方差损失定义如下:

$$L_{\text{cov}}(f) = \frac{1}{d} \sum_{i \neq j} C_{i,j}^2 \quad (5)$$

其中, C 为协方差矩阵:

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (6)$$

其中, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$,使用协方差矩阵 C 设计损失函数可

以利用非对角系数的平方和,并使用维度系数 d 进行缩放。

协方差损失将 f 的不同维度去相关,防止特征编码的相似。同时,为保证特征的不变性,设计不变性损失为:

$$L_{\text{ed}}(f) = \frac{1}{n} \sum_i \|x_i - x'_i\|_2^2 \quad (7)$$

通过方差损失 $L_v(f)$ 保证单个维度方差为1,通过协方差损失 $L_{\text{cov}}(f)$ 保证每个维度之间都是去相关的,通过不变性损失 $L_{\text{ed}}(f)$ 保证特征的不变性;同时通过 a 和 b 的调整获得更有效的舌象图像特征,以此提升识别准确度。

1.4 舌象多标签残差映射

由于单个舌象图片的特征较多,且特征之间互有联系,这将导致舌象识别精准度的下降。为此,本研究设计舌象多标签残差映射,通过舌象图像数据训练出残差映射图,在辅助分支输出舌象类别后,利用该残差映射图进一步判断,以提升识别准确度。残差映射的结构见图4,残差映射首先使用K近邻算法获取单个舌象样本的扩展数据空间,即把有相似标签的单个舌象样本和之前的样本进行组合,以获得相似样本之间的关系,接下来利用计算残差值获得每个样本组的残差值组。其中,残差值组的计算公式如下:

$$r = y - Px \quad (8)$$

其中, x 表示样本特征组, y 表示样本标签, P 是最小二乘逼近。 P 可以通过如下公式获得:

$$\min_{P \in \mathbb{R}^{p \times q}} \frac{1}{2} \|Y - P^T X\|_F^2 \quad (9)$$

其中, Y 表示标签组, X 表示特征组, p 表示特征的数量, q 表示标签的数量, $\|\cdot\|_F$ 表示弗罗伯尼范数。

在获得原样本及其邻居样本(即有相似标签的舌象样本)的残差值组后,需进一步缩小样本间距离,其可以被视为约束优化问题:

$$\operatorname{argmin}_{V \in \mathbb{R}^{q \times d}} \left\| (Y - P^T X)^T V \right\|_F^2, \text{ s.t. } V^T Y Y^T V = I \quad (10)$$

其中, V 是度量矩阵, $V = (v_1, \dots, v_d)$, $v_i (i = 1, \dots, d)$ 表示映射方向。通过学习低秩度量矩阵 V ,将数据映射到映射空间来实现标签空间的降维,从而更好地区分样本。

由于单个样本存在邻居,所以,邻居样本的优化约束可以定义如下:

$$\operatorname{argmin}_{V \in \mathbb{R}^{q \times d}} \sum_{i=1}^k \left\| (\hat{Y}_i - P_i^T \hat{X}_i)^T V \right\|_F^2, \text{ s.t. } V^T \left(\sum_{i=1}^k \hat{Y}_i \hat{Y}_i^T \right) V = I \quad (11)$$

其中, $\hat{Y}_i (i = 1, \dots, k)$ 表示样本标签向量,相应地, $\hat{X}_i (i = 1, \dots, k)$ 表示样本向量。将式(10)和式(11)进行组合,可以得到:

$$\operatorname{argmin}_{V \in \mathbb{R}^{q \times d}} \left[\left\| (Y - P^T X)^T V \right\|_F^2 + \lambda \sum_{i=1}^k \left\| (\hat{Y}_i - P_i^T \hat{X}_i)^T V \right\|_F^2 \right],$$
$$\text{s.t. } V^T \left(Y Y^T + \lambda \sum_{i=1}^k \hat{Y}_i \hat{Y}_i^T \right) V = I$$

(12)

其中, λ 是一个常数, 可以调整邻居样本的权重。通过上述公式可以最小化映射空间内样本的输入和输出之间的距离, 从而获得更好的识别结果。

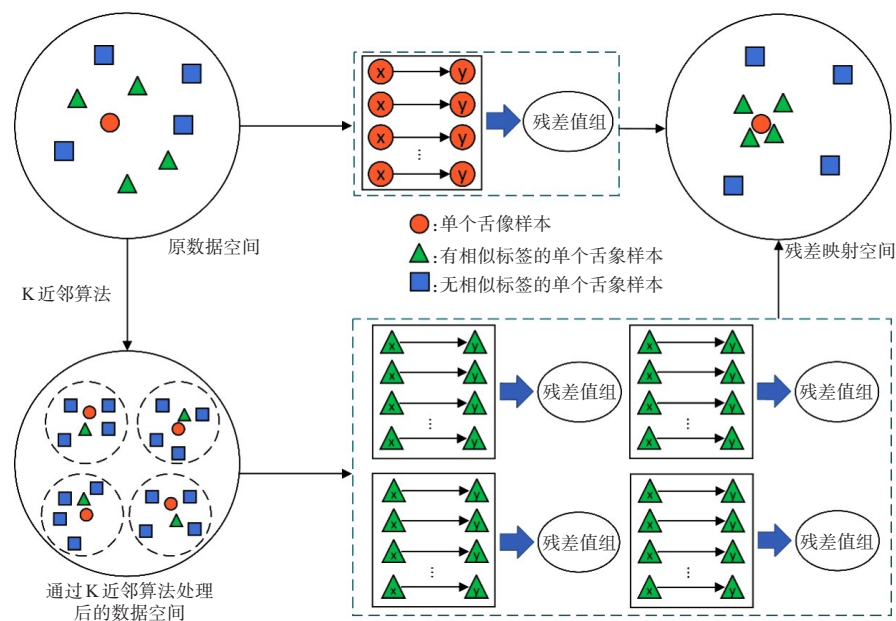


图4 多标签残差映射架构图
Figure 4 Multi-label residual mapping architecture

2 实验

2.1 训练参数

实验均在 Ubuntu 20.04 系统上运行。具体硬件采用型号为 GeForce RTX 3060 12 GB 的 GPU 和型号为 R7 5700X 的 CPU。训练参数见表 3。

表3 训练参数
Table 3 Training parameters

参数	取值
优化器	SGD 优化器
邻居权重 λ	0.1
学习率	0.001
学习率更新机制	Cosine Annealing Warm Restarts
权重衰减	$1e^{-6}$
训练批次	120
批次大小	16

2.2 识别精确率

本研究使用精确率 Precision 作为评价指标:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(13)

其中, TP 表示实际为真且预测为真的个数, FP 表示实际为假但预测为真的个数。

为进行对比, 同时将本研究模型和其他的多标签分类模型进行对比, 具体包括 WILDCAT^[10]、CSRA^[11]、TDRG^[12] 和 ASL^[13]。WILDCAT 采用弱监督的方式定位物体的特征, 使用 MultiLabelMarginLoss 作为损失函数, ResNet101 作为骨干网络。CSRA 通过空间注意力得分为单一类别设计专属特征, 并将其与平均池化层输出特征结合。TDRG 通过基于 Transformer 的二元关系学习框架构建结构关系图和语义关系图之间的关联。ASL 针对阳性样本较少的问题, 通过平衡阳性样本和阴性样本之间的概率提升准确度。

在实验中, WILDCAT、TDRG 和 ASL 获得较低的精确率, 分别只有 9.1%、25.0% 和 27.5%。CSRA 获得 60.0% 的精确率, 相较于前三者有较大的提升。CSRA 使用 Vision Transformer^[14] 构建每个类别的专属特征, 结合平均池化层输出的特征, 以此实现比 ILDCAT、TDRG 和 ASL 高的精确率。WILDCAT、TDRG 和 ASL 均采用 ResNet 作为主干网络, 并且都在 ImageNet1000^[15] 上进行预训练。预训练后, WILDCAT、TDRG 和 ASL 均在舌象数据集上进行下

一步的训练。由于 ImageNet1000 不同类别之间的差异较大,而舌象数据集各个类别之间的差距较小,样本数量也比较少,最终造成这些方法无法适用在舌象数据集。本研究根据舌象数据集样本少以及类别之间差异度小的特点,提出基于方差、不变性和协方差的舌象特征编码损失,通过自监督学习,获取舌象预处理模型,为舌象识别提供良好的编码特征,提高识别精确度;同时,通过多分支结构学习不同大类之间的语义信息,进一步提高模型的精确度,最终取得 84.8% 的精确率。

2.3 消融实验

本研究针对舌象数据集的小样本和多标签问

题,设计舌象编码损失,以获得更精准的预训练特征;同时引入多分支网络模型,不同分支学习不同舌象类别的语义信息,避免各个舌象大类之间的扰动,提高模型的整体精确率。为验证各个模块的有效性,本研究设计了如下的消融实验。

2.3.1 舌象特征提取 通过对编码器f和扩展编码器e进行训练,舌象特征编码损失从小样本的舌象图片中获取准确的舌象特征。在该实验中,舌象特征编码损失的超参数a、b均设置为1、5、10、15、20、25。使用 Grad-CAM^[16]对网络进行特征可视化,图5为网络最后一次卷积的特征图可视化结果。

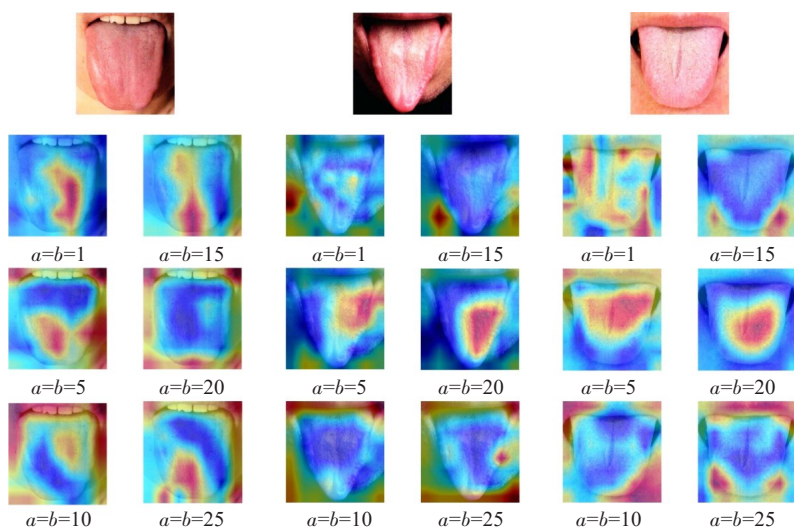


图5 特征可视化
Figure 5 Feature visualization

不同的超参数会造成网络的关注区域不同。当 $a=b=1$ 时,网络关注的区域较为零散,没有关注所谓的核心区域,即舌头的边缘或中心。随着 a 和 b 的增大,网络的关注区域逐渐集中。通过图6可以看出,当 $a=b=25$ 时,准确率达到 84.8%, $a=b=5$ 时准确率最低,为 54.00%。实验证明了舌象编码损失的有效性。

2.3.2 多分支架构 为验证多分支模型对舌象识别任务的准确率的影响,将多分支模型和单分支模型进行对比。在 ImageNet1000 上,使用 ResNet-18 作为骨干网络,分别采用单分支网络架构和多分支网络架构进行舌象识别,分别得到 46.30% 和 50.00% 的准确率。其中,单分支网络指的是提取特征的骨干网络使用 ResNet-18,最后分类部分使用多层感知机对所有的类别进行分类;多分支网络则是依旧使用 ResNet-18 作为骨干网络提取特征,在分类部分使用 4 个单独设计的多层感知机分别对舌色、舌形、苔质

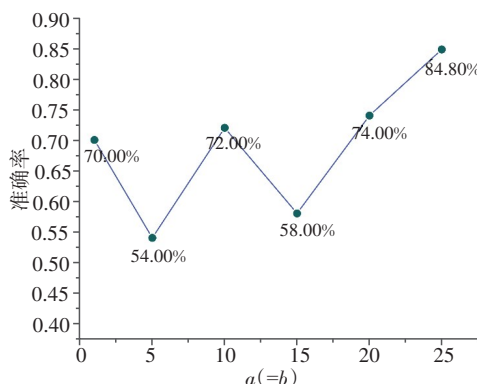


图6 $a(=b)$ 对准确率的影响
Figure 6 Effect of $a(=b)$ on accuracy

和苔色内的子类进行分类。

类似地,将本文方法分为单分支与多分支网络,分别达到 72.6% 和 84.8% 的准确率,进一步证明了针

对舌象数据集的多分支网络的有效性。此处的多分支网络架构指的是在预训练编码模型后,根据舌形、苔质、舌色和苔色的分类难度分别使用2个残差连接和2个普通卷积神经网络来提取特征,使用4个全连接层用作最后的舌象分类;单分支架构则是在预训练编码模型后连接单个残差网络模型,最后的分类阶段使用单个全连接层对所有舌象进行分类。

2.3.3 舌象多标签残差映射 对舌象多标签残差映射进行消融研究。结果显示在没有加入舌象多标签残差映射时,可获得70.0%的识别准确率;通过添加残差映射,成功地将识别准确率提升至84.8%。此外,加入多标签残差映射的提升(即14.8%)大于未加入残差映射时对于CSRA的提升(即10.0%)。实验表明舌象多标签残差映射可以增大舌象样本的类间距,减少舌象样本的类内距,降低分类器分类难度,提升舌象分类精度。

3 讨论

随着计算机视觉及深度学习技术的发展,计算机辅助医生进行诊断,加快诊断效率的技术受到了广泛关注与快速发展。翟鹏博等^[17]通过加入融合注意力机制抑制噪声信息实现舌象颜色的分类。舌形上,Li等^[18]使用R-CNN进行舌上齿痕的识别。芮迎迎等^[19]设计Mask Scoring RCNN,通过残差网络提取舌象特征,构建特征金字塔来生成感兴趣区域,从而得到齿痕预选框,实现齿痕的辨别与定位。罗思言等^[20]利用高斯混合模型对深度学习神经网络提取的特征进行聚类以提高舌象识别的准确度。在针对某一类的舌象的识别上,刘佳丽等^[21]针对裂纹舌,使用YOLO-v3作为骨干网络,同时结合细粒度学习来实现检测。温坤哲等^[22]针对胖瘦舌设计一种K Block卷积块,并应用于K-CNN网络。针对老嫩舌,曹美玲等^[23]结合了K近邻分类器与AdaBoost分类器。

传统中医舌诊依靠中医师的肉眼观察及个体判断,其结果受到该医师中医理论水平、临床经验等影响,具有很大的主观差异性,缺乏客观、定量、定性的评判标准。比如在色彩感受上,红舌、淡红舌、暗红舌之间如何进行色差的界定,临床上很少借助比色卡,这就可能导致医生对患者体质的寒热虚实属性评判上出现误差;再比如润苔与滑苔,两者区别主要在舌苔的湿润度上,反映人体津液的充盛程度不同,帮助医生判断患者体内是否存在湿邪,人为主观判断的偏差将使诊疗原则、处方用药走上不同的道路^[24-26]。本研究通过构建基于度量学习的多分支舌象识别架构,提高诊断准确率,为舌象识别提供一种新的解决方案。针对各舌象间特征差异混杂且特征

类别较多的问题,通过权重共享,在网络前半部分使用相同参数,获取浅层信息;涉及到深层细节信息时,针对4个大类别分别设计4个辅助识别分支,分别为舌色、舌形、苔质和苔色识别分支。同时,由于苔色和舌色的特征类别较少,识别任务较为简单,故使用简单的特征下采样模块;而舌形和苔质较为复杂,故使用残差块作为特征下采样模块,以提升检测精确度。此外,还设计了具有方差、不变性和协方差特性的舌象特征编码损失,利用自监督学习机制,在无标签的情况下训练编码器,获得更精准的特征图,提升精度。最后设计了舌象多标签残差映射空间,通过将样本域进行优化,提升各样本间的差异度,提升识别准确度。

4 结论

本研究提出一种基于度量学习的多分支舌象识别架构,设计舌象特征编码损失,提升特征精度;构造多分支舌象识别模型,学习不同语义的特征,降低检测难度;通过构建舌象多标签残差映射空间增加类间距,减少类内距,对最终识别结果进行进一步优化。实验表明本文方法解决了舌象识别过程中的数据样本少、类别较多,且各类别之间相互干扰的问题,获得较高的舌象识别准确率。

【参考文献】

- [1] 许家伦. 中医舌诊彩色图谱[M]. 上海: 上海中医药大学出版社, 2009. Xu JT. Color atlas of Chinese medical tongue diagnosis[M]. Shanghai: Publishing House of Shanghai University of TCM, 2009.
- [2] Heidari M, Fouladi-Ghaleh K. Using Siamese networks with transfer learning for face recognition on small-samples datasets[C]//2020 International Conference on Machine Vision and Image Processing (MVIP). Piscataway, NJ, USA: IEEE, 2020: 1-4.
- [3] Su H, Xiang L, Hu AJ, et al. A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions[J]. Mech Syst Signal Process, 2022, 169: 108765.
- [4] Yan Y, Sun J, Yu JP, et al. Small sample radar target recognition based on metric learning[C]//2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Piscataway, NJ, USA: IEEE, 2020: 441-445.
- [5] Zhang ML, Li YK, Liu XY, et al. Binary relevance for multi-label learning: an overview[J]. Front Comput Sci, 2018, 12(2): 191-202.
- [6] Liu SY, Song XH, Ma ZC, et al. MoRE: multi-output residual embedding for multi-label classification[J]. Pattern Recognit, 2022, 126: 108584.
- [7] Qu XW, Che H, Huang J, et al. Multi-layered semantic representation network for multi-label image classification[J]. Int J Mach Learn Cybern, 2023, 14(10): 3427-3435.
- [8] 裴一萍. 中医舌诊彩色图谱[M]. 北京: 中国中医药出版社, 2010. Gong YP. Chromatic illustrated atlas of tongue diagnosis in Chinese medicine[M]. Beijing: China Press of Traditional Chinese Medicine, 2010.
- [9] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [10] Durand T, Mordan T, Thome N, et al. WILDCAT: weakly supervised learning of deep ConvNets for image classification, pointwise localization and segmentation[C]//2017 IEEE Conference on

- Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 5957-5966.
- [11] Zhu K, Wu JX. Residual attention: a simple but effective method for multi-label recognition[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 184-193.
- [12] Zhao JW, Yan K, Zhao YF, et al. Transformer-based dual relation graph for multi-label image recognition[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 163-172.
- [13] Ridnik T, Ben-Baruch E, Zamir N, et al. Asymmetric loss for multi-label classification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 82-91.
- [14] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03)[2022-12-13]. <https://arxiv.org/abs/2010.11929>.
- [15] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2009: 248-255.
- [16] Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 618-626.
- [17] 翟鹏博, 杨浩, 宋婷婷, 等. 融合注意力机制的多阶段舌象分类算法[J]. 计算机工程与设计, 2021, 42(6): 1606-1613.
- Zhai PB, Yang H, Song TT, et al. Multi-stage tongue image classification algorithm incorporating attention mechanism [J]. Computer Engineering and Design, 2021, 42(6): 1606-1613.
- [18] Li XQ, Zhang Y, Cui Q, et al. Tooth-marked tongue recognition using multiple instance learning and CNN features[J]. IEEE Trans Cybern, 2019, 49(2): 380-387.
- [19] 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.
- [20] 罗思言, 王心舟, 饶向荣. 基于深度学习的舌象自监督聚类[J]. 中国医学物理学杂志, 2023, 40(1): 120-125.
- Luo SY, Wang XZ, Rao XR. Self-supervised clustering of tongue images based on deep learning [J]. Chinese Journal of Medical Physics, 2023, 40(1): 120-125.
- [21] 刘佳丽, 孙自强. 基于Double-D算法的舌像检测[J]. 计算机工程与设计, 2020, 41(7): 2025-2030.
- Liu JL, Sun ZQ. Tongue image detection based on Double-D algorithm [J]. Computer Engineering and Design, 2020, 41(7): 2025-2030.
- [22] 温坤哲, 韦玉科. 基于深度卷积神经网络的舌体胖瘦精细分类[J]. 现代计算机, 2020(34): 87-90.
- Wen KZ, Wei YK. Fine classification of fat and thin tongue based on deep convolution neural network [J]. Modern Computer, 2020(34): 87-90.
- [23] 曹美玲, 张新峰, 沈兰荪. 分类器融合技术在中医舌象老嫩识别中的应用研究[J]. 北京生物医学工程, 2006, 25(6): 644-648.
- Cao ML, Zhang XF, Shen LS. Application survey of information combination in the toughness and tenderness of tongue manifestation recognition [J]. Beijing Biomedical Engineering, 2006, 25(6): 644-648.
- [24] 罗思言, 王心舟, 饶向荣. 人工智能在中医诊断中的应用进展[J]. 中国医学物理学杂志, 2022, 39(5): 647-654.
- Luo SY, Wang XZ, Rao XR. Advances in the application of artificial intelligence in traditional Chinese medicine diagnosis [J]. Chinese Journal of Medical Physics, 2022, 39(5): 647-654.
- [25] 张林子, 周武, 张洪来. 自动舌诊技术的研究进展[J]. 中国中医基础医学杂志, 2023, 29(5): 871-876.
- Zhang LZ, Zhou W, Zhang HL. Research progress of automatic tongue diagnosis technology [J]. Journal of Basic Chinese Medicine, 2023, 29(5): 871-876.
- [26] 吴欣, 徐红, 林卓胜, 等. 深度学习在舌象分类中的研究综述[J]. 计算机科学与探索, 2023, 17(2): 303-323.
- Wu X, Xu H, Lin ZS, et al. Review of deep learning in classification of tongue image [J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(2): 303-323.

(编辑:谭斯允)