

# 基于噪声样本渐近修正的中医舌色分类方法

孙亮亮<sup>1,2</sup>, 李艳萍<sup>1,2\*</sup>, 张 辉<sup>1,2</sup>, 卓 力<sup>1,2</sup>

(1. 北京工业大学信息学部, 北京 100124; 2. 北京工业大学计算智能与智能系统北京重点实验室, 北京 100124)

**摘要:** 基于深度学习的中医舌色分类模型具备良好的性能, 但是依赖大量正确标注的样本. 由于人工标注样本费时费力, 不可避免地存在错误标注, 导致模型在训练过程中对噪声样本过拟合, 使其泛化能力变差. 为此, 本文提出了一种基于噪声样本渐近修正的中医舌色分类方法. 首先, 根据舌色分类的特点, 提出了一种全局-局部特征融合方法, 将其嵌入到 ResNet18 骨干网络中, 构建了舌色分类网络, 并采用集成学习范式, 提高分类模型的可靠性和稳定性; 其次, 针对噪声样本下的舌色分类网络训练问题, 提出了样本注意力机制和噪声样本标签重新标注机制, 在训练过程中对干净样本和噪声样本加以区分, 赋予不同的权重, 并逐步对噪声样本标签进行修正; 最后, 采用 Bootstrapping 损失函数降低模型对噪声样本的关注度, 抑制噪声样本对分类性能的影响. 将提出的方法在两个自建的舌色分类数据集上进行了实验验证, 结果表明, 该方法通过渐进地对噪声标签进行校正, 可以获得比现有的有噪声样本下图像分类方法更高的分类精度, Acc 指标分别达到了 94.6% 和 93.65%.

**关键词:** 中医舌色分类; 噪声样本; 样本注意力机制; 重新标注机制; Bootstrapping 损失

**基金项目:** 国家自然科学基金 (No.61871006); 国家中医药管理局中医药创新团队及人才支持计划项目 (No. ZYYCXTD-C-202210)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2024)05-1450-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220742

## A TCM Tongue Color Classification Method via Progressively Correcting Noisy Samples

SUN Liang-liang<sup>1,2</sup>, LI Yan-ping<sup>1,2\*</sup>, ZHANG Hui<sup>1,2</sup>, ZHUO Li<sup>1,2</sup>

(1. Faculty of Information, Beijing University of Technology, Beijing 100124, China;

2. Beijing Key Laboratory of Computational Intelligence and Intelligent Systems, Beijing University of Technology, Beijing 100124, China)

**Abstract:** Auto tongue color classification is an important research topic in the study of TCM (Traditional Chinese Medicine) objectification. Affected by various factors such as doctor's experience and illumination conditions, there often exist errors in the manually annotated labels, that is, noisy labels. Noisy labels will cause the model not to converge in the training process and the generalization ability will be poor. Therefore, in this paper, a TCM tongue color classification method is proposed by progressively correcting noisy samples. First, according to the characteristics of the tongue color classification, a global-local feature fusion method is proposed, which is embedded in the ResNet18 backbone network, constructing a tongue color classification network. The ensemble learning paradigm is adopted to improve the reliability and stability of the classification model. Next, for the classification network training problem under noisy samples, a sample attention mechanism and a re-labeling mechanism are proposed. During the training process, different weights are assigned to clean samples and noisy samples, and the noisy samples are gradually adjusted. Finally, the network model is optimized and trained with the Bootstrapping loss function to suppress the impact of noisy samples on the classification performance. The experimental results on two tongue color classification datasets SIPL-A and SIPL-B show that, the proposed method can effectively correct noisy labels, thereby, significantly improving the tongue color classification accuracy. Compared with the existing image classification methods under noisy samples, the proposed method can achieve a higher classification accuracy, reaching 94.6% and 93.65%, respectively.

**Key words:** TCM tongue color classification; noisy sample; sample attention mechanism; re-labeling mechanism;

bootstrapping loss

Foundation Item(s): National Natural Science Foundation of China (No.61871006); Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (No.ZYYCXTD-C-202210)

## 1 引言

舌诊是中医辨证施治的核心环节,也是中医区别于其他医学系统的独特之处。医生通过观察舌象的变化,来了解人体的生理功能和病理变化。中医理论将舌象侦察特征归结为 36 种,舌色是其中最为直观且最重要的一种。中医舌色通常分为淡红、红、暗红、紫等 4 种,因此中医舌色分析可以看作是一个分类问题。

近年来,随着深度学习的迅猛发展,学者们深入开展了基于深度学习的中医舌色分类研究工作。文献[1]通过构建多任务联合学习模型,深入挖掘舌色、苔色之间的潜在关联关系,实现了舌色、苔色的协同分类。文献[2]对 CaffeNet 网络进行改进,应用到中医舌色分类上,取得了较高的分类精度。文献[3]设计了轻型卷积神经网络(Convolutional Neural Network, CNN),以较少的参数量,取得了优异的舌色分类性能。文献[4]将整张舌图像划分为舌色、苔色区域,并分别输入到不同的轻型 CNN 网络中,取得了较高的舌色、苔色分类精度。

研究工作表明,利用 CNN 进行中医舌色分类,取得的分类性能远远超过了传统方法,但网络训练时需要大量、高质量的人工标记样本。由于人工标注样本费时费力,不可避免地会出现错误的标签,这些噪声标签使得网络模型在训练过程中对噪声样本过拟合,导致模型的泛化能力变差。因此,如何尽可能降低噪声样本对分类的影响,提升分类的鲁棒性就成为利用深度学习进行中医舌色分类时需要重点解决的问题。

有噪声样本下的图像分类一直是机器学习领域的研究热点。研究者们从损失函数、网络结构设计、正则化、样本筛选等方面出发,开展了研究工作。下面对这些研究进展进行综述。

损失函数。这类工作通过设计鲁棒的损失函数,尽可能降低噪声样本对训练过程的干扰。文献[5]设计了一种广义交叉熵损失(Generalized Cross Entropy, GCE),该损失函数可以看作是平均绝对误差(Mean Absolute Error, MAE)和类别交叉熵(Categorical Cross Entropy, CCE)损失的一种改进,可以提升有噪声标签情况下的分类性能。文献[6]提出了反向交叉熵(Reverse Cross Entropy, RCE)损失,将预测概率分布与标签概率分布进行互换,在有噪声样本的情况下,可以取得更优的分类性能,将 RCE 与 CE 相结合,形成了对称交叉熵(Symmetric Cross Entropy, SCE)损失,可以有效提升模型对于噪声样本的鲁棒性。文献[7]提出了焦点损失(Focal

Loss, FL),通过给样本赋予不同的权重,可以降低噪声样本对分类性能的干扰。文献[8]提出了课程损失(Curriculum Loss, CL),可以自适应地选择样本用于模型训练,使得噪声样本不参与训练过程。

网络结构设计。这类方法通过设计专门的网络结构来解决有噪样本下的分类问题。文献[9~13]设计了一个噪声适应层,将其嵌入到网络结构中,来应对噪声样本,但存在易陷入局部最优的问题。文献[14,15]提出了一种概率模型,通过为样本分配可信度,来处理标签噪声。

正则化。这类方法有助于解决有噪样本下网络模型泛化能力变差的问题。文献[16]通过裁剪梯度,实现了一种正则化方法,不仅简化了模型,也提升了模型的抗噪能力。文献[17]设计了一种双层正则化方法,解决模型对于噪声样本的过度拟合问题,可以减轻噪声样本的干扰。

样本筛选。这类方法筛选出有噪样本后,仅利用干净的样本训练网络。文献[18,19]设计了一种 Co-teaching 网络训练策略,通过两个网络的协同训练,为彼此的新一轮训练挑选干净样本。噪声样本往往具有较高的损失,文献[20]根据这一特点,对每次迭代的损失值进行排序,去除损失值大的有噪样本。文献[21]将训练过程分为两个阶段:第一个阶段是播种阶段,利用所有样本训练网络,以收集干净样本的种子;第二个阶段是进化阶段,仅使用收集到的干净本来训练网络,缓解了模型对有噪样本的过拟合问题。文献[22]提出了一种噪声样本过滤方法,利用数据的高阶拓扑信息,筛选出干净样本,滤除噪声样本。

元学习。元学习通过设计专门的训练策略,可以确保模型不过度拟合噪声样本。同时可以学习损失调整规则,降低噪声样本的负面影响。文献[23]提出了一种元权重网络,能够在少量噪声元数据的引导下,更好地更新学习过程。文献[17]根据训练样本的梯度方向来学习权重。在训练数据集中包含一个小的、仅包含干净样本的验证数据集,并对样本的后向损失重新进行加权,使得验证数据集上的损失最小。

半监督学习。这类方法将错误标签样本视为无标签样本,然后采用半监督学习的方法进行模型训练。文献[24]将自集成学习引入半监督学习中,以滤除噪声样本。文献[25]通过选择性翻新、利用可以高精度校正的噪声样本等手段,增加训练过程中干净样本的数量,不断提升模型的训练性能。

上述的大量研究结果表明,针对噪声样本设计相应的策略或者机制,可以有效提升分类性能.处理噪声样本时,需要首先筛选出噪声样本,然后采用相应的办法对其进行处理.直接去除噪声样本的方法虽然简单,但同时也会减少样本的数量,导致模型性能的下降.对于舌色分类任务来说,有标注的舌象样本难以获取,样本量少,因此如何充分利用噪声样本,提升分类性能,值得进行深入研究.

本文提出了一种基于噪声样本渐近修正的中医舌色分类方法,用于解决有噪样本下的中医舌色分类问题.该方法在网络训练过程中,不断发现并修正噪声样本的标签,无需去除噪声样本,实现舌色的鲁棒、准确分类.主要的创新点如下.

(1)根据舌色分类任务的特点,提出了一种全局-局部特征融合(Global-Local Feature Fusion, GLFF)方法,将低层和高层特征进行融合,提升特征的表达能力.将GLFF嵌入到ResNet18<sup>[26]</sup>中,构建了GLFF-ResNet18网络结构,用于舌色分类;考虑到有标注的舌象样本数量少,采用了集成学习范式,提高分类模型的可靠性和稳定性.

(2)针对有噪样本下的舌色分类网络训练问题,提出了样本注意力机制和噪声样本标签重新标注机制,在训练过程中对干净样本和噪声样本分别赋予不同的权重,并逐步对噪声样本的标签进行修正;利用Boosting损失函数<sup>[27]</sup>降低模型对噪声样本的关注度,抑制噪声样本对分类性能的影响.

在2个自建的中医舌色分类数据集上的实验结果表明,本文提出的方法能获得比现有的噪声样本下图像分类方法更高的分类准确率.

## 2 噪声样本渐近修正的中医舌色分类方法

本文提出的基于噪声样本渐近修正的中医舌色分类方法整体流程如图1所示.其中, $\{F_i\}_{i=1}^{N_b}$ 表示每个批次输入的样本, $N_b$ 为每个批次的样本数量,本文设置为 $N_b=64$ .

假设人工标注标签用 $y$ 表示,称为硬标签;分类模型得到的类别预测概率用 $P$ 表示,称为软标签;预测概率最大值表示为 $p, p=\max\{P\}$ , $p$ 对应的类别称为伪标签 $\hat{y}$ .

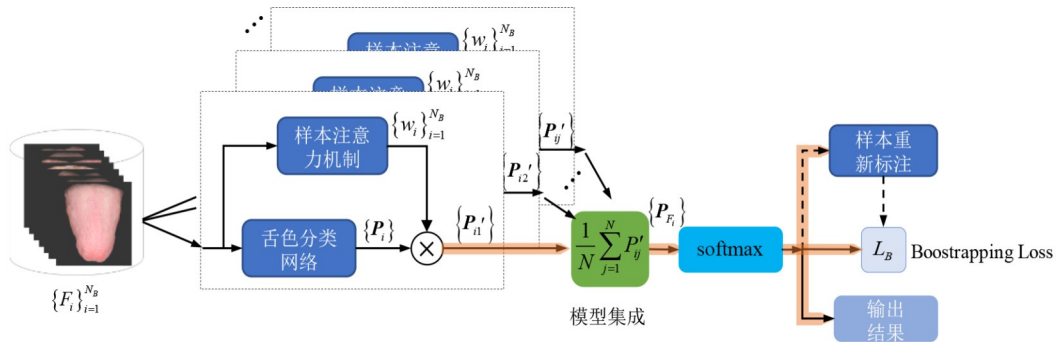


图1 本文方法整体流程

将样本数据按批次输入到分类网络和样本注意力机制模块中,得到每个样本的软标签 $\{P_i\}_{i=1}^{N_b}$ 和样本权重 $\{w_i\}_{i=1}^{N_b}$ .将二者相乘,得到 $\{P'_i\}_{i=1}^{N_b}$ ,可以表示为

$$P'_i = w_i \times P_i \quad (1)$$

由于有标注的样本数量不足,为了提升分类模型的稳定性和可靠性,本文采用集成学习的方式对每个模型的加权软标签求平均,得到 $P_{F_i}$ ,如式(2)所示:

$$P_{F_i} = \frac{1}{N} \sum_{j=1}^N P'_{ij} \quad (2)$$

其中, $N$ 代表模型数量,本文设置 $N=2$ .

将 $P_{F_i}$ 输入到Softmax中,输出为每个类别的预测概率.之后将其送入到噪声样本重新标注模块中,判断该样本是否为噪声样本.如果是,则用该样本的伪标签,替换原有的硬标签,用于下一轮的网络训练.

逐批次进行上述操作,反复迭代,直到训练结束,

得到舌色分类模型.

下面首先介绍基于集成学习的舌色分类网络结构,然后介绍有噪样本下的舌色分类网络训练策略.

### 2.1 基于集成学习的舌色分类网络

图2所示的是本文提出的舌色分类网络结构,命名为“GLFF-ResNet18”.采用ResNet18作为骨干网络,将提出的GLFF模块嵌入其中,并在最后添加了3个全连接层.

#### 2.1.1 GLFF

众所周知,CNN网络包括多个卷积层,其中低层特征中主要包含颜色、纹理、轮廓等信息,而高层特征中则主要包含的是抽象的全局语义信息.而舌色分类时主要利用的是舌体的颜色信息,因此本文提出了GLFF,具体结构如图3所示.GLFF充分发挥不同层特征的作用,使得融合后的特征中既包含低层的视觉信息,也包含高层的语义信息,增强特征的表达能力,从

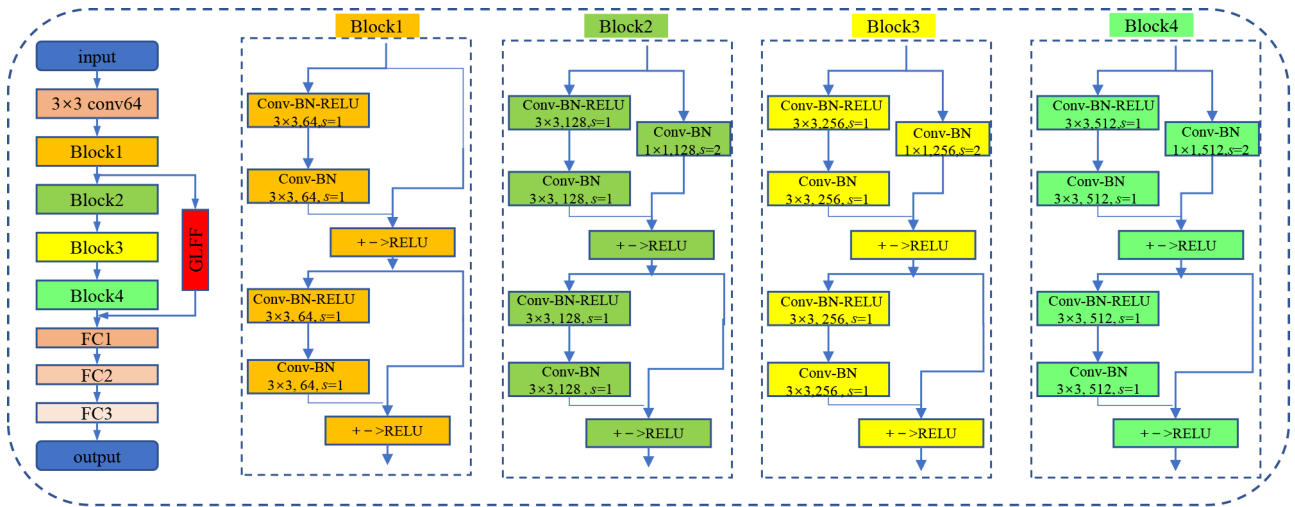


图2 GLFF-ResNet18网络结构

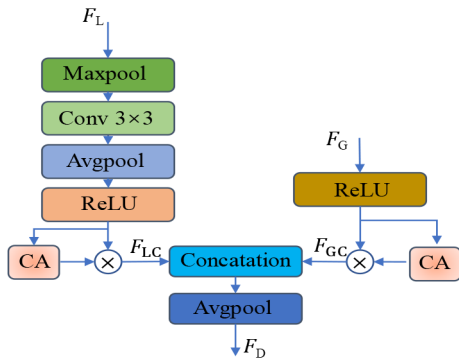


图3 GLFF网络结构

而提升舌色分类性能。

下面分别介绍 GLFF 和模型集成学习的实现细节。

GLFF 的输入分别是 ResNet18 第一个和最后一个卷积层的输出特征,用  $F_L$  和  $F_G$  表示。GLFF 的操作过程如下。

(1)对  $F_L$  进行 Maxpool 操作,目的是从  $F_L$  中筛选出分类区分性更好的特征,同时也降低特征图的尺寸。

(2)进行  $3 \times 3$  卷积、Avgpool 和 ReLU 操作。其中  $3 \times 3$  卷积用于对特征图进行平滑处理,而 Avgpool 用于对特征图进行下采样,使其与  $F_G$  具有相同的尺寸大小,ReLU 函数则用于解决梯度消失问题。将处理后的特征图与通道注意力(Channel Attention, CA)模块输出的权重相乘,得到更新后的特征图,记为  $F_{LC}$ 。其计算过程如式(3)所示:

$$F_{LC} = CA(\text{ReLU}(\text{Avgpool}(\text{Conv}_{3 \times 3}(\text{Maxpool}(F_L)))) \otimes \text{ReLU}(\text{Avgpool}(\text{Conv}_{3 \times 3}(\text{Maxpool}(F_L)))) \quad (3)$$

其中,  $\otimes$  代表逐像素乘法操作。

(3)对  $F_G$  进行 ReLU 和 CA 操作,得到  $F_{GC}$ ,如式(4)所示:

$$F_{GC} = CA(\text{ReLU}(F_G)) \otimes \text{ReLU}(F_G) \quad (4)$$

(4)将  $F_{LC}$  和  $F_{GC}$  拼接后,进行 Avgpool 操作,得到融合后的特征  $F_D$ ,如式(5)所示:

$$F_D = \text{Avgpool}(\text{Concatation}(F_{LC}, F_{GC})) \quad (5)$$

### 2.1.2 舌色分类网络的集成学习

集成学习通过某种策略将多个模型集成起来,通过群体决策来提高决策的准确性。本文通过 2 个 GLFF-ResNet18 网络的集成学习,来提升舌色分类性能。

## 2.2 有噪样本下的舌色分类网络训练策略

大量的研究表明,在图像分类任务中,干净样本能够正向地引导模型训练,而噪声样本会加大模型训练的难度,甚至导致模型训练过程无法收敛,不利于模型性能的提升。因此,在模型训练中,要首先对干净样本和噪声样本进行区分,在此基础上对噪声样本进行校正,抑制其对分类性能的影响。

本文设计了样本注意力机制和噪声样本重标注机制,并通过 Bootstrapping 损失函数来发挥不同样本的作用。下面对这 3 个部分进行详细的介绍。

### 2.2.1 样本注意力机制

本文设计样本注意力机制的目的是,在训练过程中为同一批次内的每个样本分配不同的权重,避免模型对于噪声样本过度拟合。具体来说,在训练过程中给干净样本赋予较高的权重,而对噪声样本则赋予较低的权重,尽可能避免其对网络训练的干扰。

由于舌象样本的规模有限,本文结合两个不同网络的输出结果,来给每个样本分配权重。图 4 所示的是样本注意力机制的网络结构。该网络包含两个分支,其中一个分支是将 ShuffleNetV2<sup>[28]</sup> 与 3 个全连接层相连,每层神经元个数分别为 512、128 和 1;另一个则是将 MobileNetV2<sup>[29]</sup> 与 3 个全连接层相连,每层神经元个数分别是 1 000、128 和 1。

对两个分支输出的最大概率值求平均,作为该批

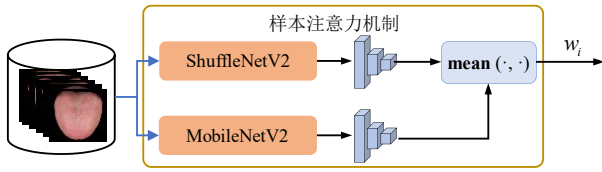


图4 样本注意力机制结构

次每个样本的权重  $\{w_i\}_{i=1}^{N_s}$ 。统计结果表明,对于干净的样本,其输出的最大概率值通常比噪声样本的大。因此,干净样本会具有较高的权重,而噪声样本则具有较低的权重。

为了进一步拉大干净样本和噪声样本权重值之间的差值,本文设计了一个正则化模块。正则化处理通过损失函数  $L_R$  来加以实现。具体来说,对每个批次的样本权重进行降序排列,并按照某一阈值将样本分为高权重组和低权重组。对两组权重分别求平均,得到  $\bar{p}_h$  和  $\bar{p}_l$ ,则  $L_R$  的计算公式如下:

$$L_R = \max \left\{ 0, \varepsilon - \sqrt{\bar{p}_h^2 - \bar{p}_l^2} \right\} \quad (6)$$

其中,  $\varepsilon$  是一个预设值,本文设置为 0.15。可以看出,  $\bar{p}_h \gg \bar{p}_l$  时,  $L_R \approx 0$ ; 而当  $\bar{p}_h \approx \bar{p}_l$  时,  $L_R > 0$ 。在网络训练的过程中,  $L_R$  的值会逐渐减小。因此,通过  $L_R$  就可以逐步拉大  $\bar{p}_h$  和  $\bar{p}_l$  之间的差值。

### 2.2.2 噪声样本标签重新标注机制

噪声样本标签重新标注机制用于纠正错误的标签。为此,需要首先判断该样本是否为噪声样本,然后对其标签重新进行标注,判决条件可以用式(7)表示:

$$\sqrt{p_{\max}^2 - p_{\text{gt}}^2} > \delta \quad (7)$$

其中,  $p_{\max}$  和  $p_{\text{gt}}$  分别表示样本类别预测概率的最大值和样本原始硬标签对应的预测概率值;  $\delta$  是预设的阈值,本文设  $\delta=0.5$ 。当  $p_{\max}$  和  $p_{\text{gt}}$  之间的标准差大于阈值  $\delta$  时,意味着该样本的伪标签与硬标签不一致,将该样本视为噪声样本,并对其标签进行修正。采用伪标签替换硬标签,用于模型的迭代训练。本文采用标准差的目的是拉大  $p_{\max}$  和  $p_{\text{gt}}$  之间的差距,更有利于模型发现噪声样本,并及时对其标签进行修正。

### 2.2.3 损失函数

本文在网络训练时采用的损失函数如式(8)所示:

$$\text{Loss} = L_B + \sum_{i=1}^m L_R^i \quad (8)$$

其中,  $L_B$  表示 Bostrapping 损失函数;  $\sum_{i=1}^m L_R^i$  是指多个集成模型中的正则化损失之和,  $L_R$  的表达式如式(6)所示,  $m$  表示集成模型的数目。

传统的 CE 损失函数使得模型在训练过程中更关注于困难样本,这使得模型较难拟合出更好的效果。与

CE 不同,Bostrapping 使得模型同时关注硬标签与伪标签,其表达式如式(9)所示:

$$L_B = - \sum_{i=1}^{N_g} \left( (1-w_i)y_i + w_i\hat{y}_i \right)^T \log(h_i) \quad (9)$$

其中,  $y_i$  表示硬标签;  $\hat{y}_i$  表示伪标签;  $h_i$  表示软标签;  $w_i$  表示权重,取值为 0.2。可以看出,Bostrapping 损失函数把模型预测得到的伪标签也同时加以考虑,这样就会直接降低噪声样本的损失,使得模型减少对噪声样本的关注。而对于干净的样本,  $\hat{y}_i$  带来的影响会相对较小,有利于网络的优化训练。

本文将网络训练分为两个阶段:第一阶段,仅对网络进行迭代训练,不加入噪声样本标签重新标注机制;第二阶段,加入噪声样本标签重新标注机制,此时趋于稳定的模型可以对噪声样本逐步进行修正,提升模型的训练性能。

## 3 实验结果与分析

本文在两个自建的舌色分类数据集上进行了实验,以充分验证所提方法的有效性。下面介绍实验结果,并对其进行分析。

### 3.1 数据集和参数设置

数据集。实验中采用了 SIPL-A 和 SIPL-B 两个自建数据集,其中的舌图像利用自行研制的中医舌象分析仪采集得到,每个舌象样本均由具有多年临床经验的中医专家手工进行标注。两个数据集的部分示例图像如图 5 和图 6 所示。



图5 SIPL-A 舌色样本的示例



图6 SIPL-B 舌色样本的示例

在实验之前,首先利用文献[30,31]中的数据清洗方法对两个数据集进行了清洗,得到清洗后的数据集,用 SIPL-A-C 和 SIPL-B-C 表示。表 1 所示的是数据集清洗前后每个类别的数目。

在实验过程中,首先将清洗后的数据集按照 4:1 的比例进行随机划分,其中 20% 用作测试集。训练集由清洗前的数据集去除测试集后组成,目的是保证测试集中的样本均是干净样本,在统计分类准确率时,可以

表1 SIPL-A 与 SIPL-B 清洗前后每个类别的数量

	SIPL-A					SIPL-B				
	淡红舌	红舌	暗红舌	紫舌	总数	淡红舌	红舌	暗红舌	紫舌	总数
清洗前	134	135	116	45	430	250	250	185	36	721
清洗后	126	128	113	42	409	236	233	182	36	687

排除噪声样本的干扰.

分类性能评价指标 本文采用准确率、召回率及 F1-score 三项指标来度量舌色分类性能. 准确率 Acc、召回率 Recall 及 F1 score 的定义如式(10)、式(11)和式(12)所示:

$$\text{Acc} = (\text{TP} + \text{TN})/N \quad (10)$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{TN}) \quad (11)$$

$$\text{F1-score} = 2 \times \text{TP}/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (12)$$

假设 C 表示舌色类别中的一类, TP 表示 C 类中的样本被正确预测为类 C 的数量, TN 表示不属于 C 类的其他类别样本而被预测为其对应类别的数量, N 表示测试集中样本的总数量.

数据扩充. 本文采用随机水平和垂直翻转、在  $-45^\circ$  和  $45^\circ$  之间随机旋转、调整图像大小等多种数据扩充方法来扩大训练样本的规模. 输入图像像素大小均为  $224 \times 224$ .

参数设置. 本文在 Pytorch 平台上进行舌色分类网络的搭建、训练和测试, 硬件配置为 NVIDIA GeForce TX 1080 Ti GPU. 训练中, 采用了 Adam 算法对模型进行优化, batchsize 设为 64、权重衰减设为  $10^{-4}$ . 此外, 本文采用自定义策略在训练过程中调整学习率,  $\text{lr} = \text{lr}_0 \times (0.95)^{\text{epoch}}$ , 其中  $\text{lr}_0$  值为 0.001, epoch 的值为 200. 每 20 轮计算一次均值, 将所得的结果作为测试准确率.

### 3.2 与其他有噪样本下的分类方法对比

为了验证所提出方法的有效性, 在 SIPL-A 和 SIPL-B 两个数据集上, 将其与 6 种代表性的有噪样本下图像分类方法进行了对比, 包括 Label Smoothing<sup>[32]</sup>、GCE (Generalized Cross Entropy)、PENCIL<sup>[33]</sup>、AFM (Attentive Feature MixUp)<sup>[34]</sup>、Co-teaching、Co-teaching+ 等. 每种方法训练和测试所使用的数据以及参数设置均相同, 如 3.1 节所述. 两个数据集上的实验结果如表 2 所示.

从表 2 中可以看出, 相比于现有的方法, 本文方法可以得到最优的分类性能. 在 SIPL-A 数据集上, 分类召回率提升了 2.71% 以上, 达到了 94.26%; 分类 F1-score 提升了 2.66% 以上, 达到了 93.93%; 分类准确度提升了 2.66% 以上, 达到了 94.60%. 在 SIPL-B 数据集上, 分类召回率提升了 1.68% 以上, 达到了 92.94%; 分类 F1-score 提升了 1.34% 以上, 达到了 93.27%; 分类准确度分别提升了 1.24% 以上, 达到了 93.65%.

Label Smoothing 方法以 soft-one-hot 代替真实的

表2 与其他方法在 SIPL-A 和 SIPL-B 数据集上的分类性能对比

单位: %

方法	SIPL-A			SIPL-B		
	Recall	F1-score	Acc	Recall	F1-score	Acc
Label Smoothing	88.92	89.23	89.46	88.79	89.11	89.36
GCE	89.58	90.94	91.51	89.47	89.88	90.41
PENCIL	89.04	89.87	90.16	86.82	87.02	87.32
AFM	90.66	91.29	91.65	91.10	91.39	91.84
Co-teaching	89.89	90.18	90.93	87.03	87.45	87.75
Co-teaching+	91.27	91.55	91.96	91.26	91.93	92.41
本文方法	<b>93.93</b>	<b>94.26</b>	<b>94.60</b>	<b>92.94</b>	<b>93.27</b>	<b>93.65</b>

one-hot, 使得标签具有概率值的软标签, 在真实标签上的概率值最大, 其他的概率值很小, 相当于是正则化处理, 因此在相对较小的模型上, 精度提升不是很明显, 结果比较接近.

在 SIPL-A 数据集上, GCE 的准确率、召回率及 F1-score 分别达到了 89.58%、90.94%、91.51%. 在 SIPL-B 数据集上, AFM 的准确率、召回率及 F1-score 分别达到了 91.10%、91.39%、91.84%. Co-teaching 方法通过双网络互学习来抑制噪声样本的影响, 但到了一定的训练轮数后, 性能不再提升. 而 Co-teaching+ 将“分歧更新”策略与原始的 Co-teaching 结合起来, 实验结果表明, 与 Co-teaching 相比, 在 SIPL-A 和 SIPL-B 数据集上, Co-teaching+ 均获得了更优的分类性能.

相比之下, 本文方法通过样本注意力机制、噪声样本标签重新标注机制和 Bootstrapping 损失函数, 可以有效抑制噪声样本对分类性能的影响, 获得了最优的分类性能. 由于 2 个数据集在样本数量、类别数量、干净样本和噪声样本的比例等方面有所不同, 因此分类精度上也存在差异.

### 3.3 不同 CNN 网络的分类结果

为了验证 GLFF-ResNet18 的分类性能, 在清洗后的 SIPL-B-C 数据集上, 将其与 AlexNet<sup>[35]</sup>、Vgg16<sup>[36]</sup>、ShuffleNetV2 和 MobileNetV2 等几种主流的轻型 CNN 网络进行了对比, 网络训练时均采用了 CE 损失函数, GLFF-ResNet18 未采用模型集成方式. 将 SIPL-B-C 数据集按照 4:1 比例随机划分为训练集和测试集, 每个网络均采用 ImageNet 上预训练+微调的策略进行训练.

表 3 所示的是 SIPL-B-C 数据集上分别采用不同的网络得到的分类性能对比结果. 可以看出, 本文设计的

GLFF-ResNet18 网络可以取得最高的分类性能, 准确率、召回率和 F1-score 等三个指标分别达到了 92.06%、91.55% 和 91.39%。其他轻型 CNN 网络中分类性能最高的是 MobileNetV2, 与其相比, GLFF-ResNet18 的 3 个指标分别提高了 0.85%、1.11% 和 1.26%。这是因为 GLFF 通过融合网络不同层的特征, 提升了特征表达能力, 从而提高了分类准确率。

表 3 不同网络在 SIPL-B-C 数据集上的分类性能对比结果

单位: %

网络模型	Recall	F1-score	Acc
AlexNet	88.47	85.11	89.78
Vgg16	88.74	86.63	90.76
ResNet18	90.17	89.93	91.15
MobileNetV2	90.44	90.13	91.21
ShuffleNetV2	88.46	86.55	90.84
本文提出的分类网络	<b>91.55</b>	<b>91.39</b>	<b>92.06</b>

对比表 2 和表 3 的实验结果, 还可以看出, 相比于直接清洗噪声样本, 本文提出的噪声样本抑制策略可以获得更优的分类性能。这是因为清洗样本会减少训练样本数据量, 降低模型的训练性能。

### 3.4 样本注意力机制中不同组合网络的分类性能

本文将各种轻型网络结构进行组合, 以验证样本注意力机制中不同网络结构组合对分类性能的影响。网络结构主要包括 DenseNet121<sup>[37]</sup>、ResNet18、ResNext50<sup>[38]</sup>和 EfficientNet-b4<sup>[39]</sup>等, 实验中使用的数据以及参数设置均如 3.1 节所述。在 SIPL-A 和 SIPL-B 两个数据集上的实验结果如表 4 所示。

从表 4 中可以看出, 与单个网络结构相比, 在样本注意力机制中采用两个网络模型可以获得更优的分类性能。主要原因是采用两个网络可以有效避免预测结果的随机性。而在两个不同网络的组合中, Mobile-

表 4 不同网络组合对舌色分类性能的影响

模型	SIPL-A	SIPL-B
MobileNetV2	90.16	89.81
ShuffleNetV2	90.24	90.35
DenseNet121+ ResNet18	89.96	89.67
ResNext50+ ShuffleNetV2	90.89	91.02
MobileNetV2+ ResNet18	90.66	89.46
EfficientNet-b4+ ResNet18	91.06	90.55
EfficientNet-b4+ MobileNetV2	90.44	89.92
<b>MobileNetV2+ ShuffleNetV2</b>	<b>91.75</b>	<b>91.34</b>

NetV2 与 ShuffleNetV2 组合可以取得最优的分类性能。在 SIPL-A 数据集上, 舌色分类准确率达到 91.75%, 比其他网络组合提升了 0.69% 以上。

### 3.5 消融实验

本节在 SIPL-A 数据集上, 以 ResNet18 为基线网络, 进行了消融实验。基线网络使用 CE 损失进行网络训练。表 5 所示的是分别加入 GLFF、样本注意力机制、分类模型集成学习、噪声样本标签重新标注和 Bootstrapping 损失函数等前后的分类性能对比结果。从表 5 可以看出, 采用 GLFF、召回率、F1-score 及准确率等 3 个指标相对于基线模型分别提升了 1.40%、2.54%、1.06%。加入模型集成学习之后, 3 个指标分别提升了 1.43%、2.08%、1.54%。在加入样本注意力机制后, 3 个指标分别提升了 1.23%、1.49%、0.55%, 达到了 92.44%、91.34%、92.30%。这说明样本注意力机制可以使得模型在训练的过程中更少地关注噪声样本, 降低噪声样本的干扰。

加入噪声样本标签重新标注机制后, 三个指标分别提高了 1.22%、1.08%、1.16%。这是因为噪声样本标签重新标注机制可以在训练的过程中对错误标签进行修正, 因此模型可以在不筛除噪声样本的情况下, 降低噪声样本对分类性能的影响。

表 5 消融实验结果

ResNet18	GLFF	分类模型集成学习	样本注意力机制	噪声样本标签重新标注机制	Boostrapping	Recall/%	F1-score/%	Acc/%
√						88.38	84.03	89.15
√	√					89.78	86.87	90.21
√	√	√				91.21	89.85	91.75
√	√	√	√			92.44	91.34	92.30
√	√	√	√	√		93.66	92.12	93.46
√	√	√	√	√	√	<b>94.18</b>	<b>92.59</b>	<b>94.60</b>

利用 Bootstrapping 替换 CE 后, 3 个指标进一步提升了 0.52%、0.87%、1.14%。这说明该损失函数可以有效

提升分类模型的抗噪性能。可视化混淆矩阵 (Confusion Matrix, CM) 结果如图 7 所示。

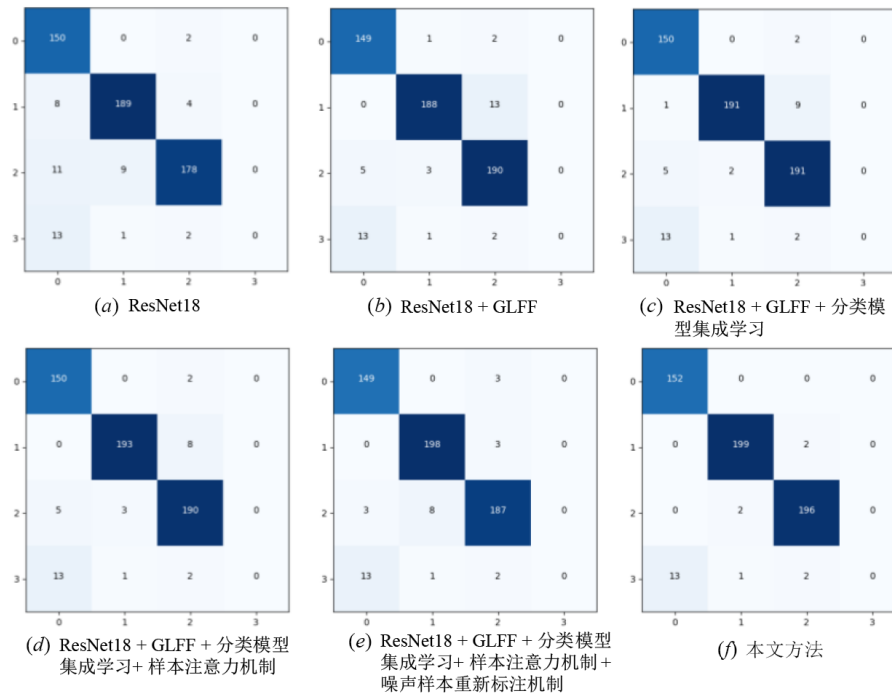


图7 可视化混淆矩阵

#### 4 结束语

为了抑制噪声样本对中医舌色分类性能的影响,本文提出了一种噪声样本标签的渐近修正方法.该方法在不筛除噪声样本的前提下,在网络训练过程中逐步识别噪声样本,对其噪声标签进行修正,从而提升了舌色分类性能.

本文的研究结果表明,针对舌色分类任务的具体特点,设计GLFF,将低层特征与高层特征有效融合,有助于提升分类性能.此外,与直接清洗噪声样本相比,在训练过程中对噪声样本的标签逐步进行修正,可以获得更高的分类准确率.

在未来的工作中,将进一步增大数据集的规模,提升分类准确率和模型的泛化能力,为实际临床应用打下良好基础.

#### 参考文献

- [1] 汤一平, 王丽冉, 何霞, 等. 基于多任务卷积神经网络的舌象分类研究[J]. 计算机科学, 2018, 45(12): 255-261.  
TANG Y P, WANG L R, HE X, et al. Classification of tongue image based on multi-task deep convolutional neural network[J]. Computer Science, 2018, 45(12): 255-261. (in Chinese)
- [2] HOU J, SU H Y, YAN B, et al. Classification of tongue color based on CNN[C]//2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). Piscataway: IEEE, 2017: 725-729.
- [3] FU S Y, ZHENG H, YANG Z J, et al. Computerized tongue coating nature diagnosis using convolutional neural network[C]//2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). Piscataway: IEEE, 2017: 730-734.
- [4] 肖庆新. 基于深度学习的中医舌象特征分类技术研究[D]. 北京: 北京工业大学, 2019.  
XIAO Q X. Research on Classification Technology of Tongue Picture Features in Traditional Chinese Medicine Based on Deep Learning[D]. Beijing: Beijing University of Technology, 2019. (in Chinese)
- [5] ZHANG Z L, SABUNCU M R. Generalized cross entropy loss for training deep neural networks with noisy labels [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 8792-8802.
- [6] WANG Y S, MA X J, CHEN Z Y, et al. Symmetric cross entropy for robust learning with noisy labels[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 322-330.
- [7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2999-3007.
- [8] LYU Y M, TSANG I W. Curriculum loss: Robust learning and generalization against label corruption[EB/OL]. (2019)

- [2022]. <http://arxiv.org/abs/1905.10045.pdf>.
- [9] MENON A, RAWAT A, REDDI S J, et al. Can gradient clipping mitigate label noise?[C/OL]//International Conference on Learning Representations. (2020) [2022]. <https://openreview.net/pdf?id=rkIB76EKPr>.
- [10] GOLDBERGER J, BEN-REUVEN E. Training deep neural-networks using a noise adaptation layer[C/OL]//International Conference on Learning Representations. (2016) [2022]. <https://openreview.net/pdf?id=H12GRgxcg>.
- [11] CHENG L L, ZHOU X Z, ZHAO L M, et al. Weakly supervised learning with side information for noisy labeled images[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 306-321.
- [12] XIA X B, LIU T L, HAN B, et al. Extended T: Learning with mixed closed-set and open-set noisy labels[EB/OL]. (2020)[2022]. <http://arxiv.org/abs/2012.00932.pdf>.
- [13] BEKKER A J, GOLDBERGER J. Training deep neural-networks based on unreliable labels[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2016: 2682-2686.
- [14] JINDAL I, NOKLEBY M, CHEN X W. Learning deep networks from noisy labels with dropout regularization [C]//2016 IEEE 16th International Conference on Data Mining (ICDM). Piscataway: IEEE, 2016: 967-972.
- [15] HAN B, YAO J C, NIU G, et al. Masking: A new perspective of noisy supervision[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 5841-5851.
- [16] YAO J C, WANG J J, TSANG I W, et al. Deep learning from noisy image labels with quality embedding[J]. IEEE Transactions on Image Processing, 2018, 28(4): 1909-1922.
- [17] XIA X B, LIU T L, HAN B, et al. Robust early-learning: Hindering the memorization of noisy labels[C/OL]//International Conference on Learning Representations. (2020) [2022]. [https://openreview.net/pdf?id=Eq15b1\\_hTE4](https://openreview.net/pdf?id=Eq15b1_hTE4).
- [18] NORTH CUTT C, JIANG L, CHUANG I. Confident learning: Estimating uncertainty in dataset labels[J]. Journal of Artificial Intelligence Research, 2021, 70: 1373-1411.
- [19] HAN B, YAO Q M, YU X R, et al. Co-teaching: Robust training of deep neural networks with extremely noisy labels[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 8536-8546.
- [20] YU X R, HAN B, YAO J C, et al. How does disagreement help generalization against label corruption? [EB/OL]. (2019)[2022]. <http://arxiv.org/abs/1901.04215.pdf>.
- [21] HUANG J C, QU L, JIA R F, et al. O2U-net: A simple noisy label detection approach for deep neural networks [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 3325-3333.
- [22] SONG H, KIM M, PARK D, et al. Robust learning by self-transition for handling noisy labels[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM, 2021: 1490-1500.
- [23] WU P X, ZHENG S Z, GOSWAMI M, et al. A topological filter for learning with label noise[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 21382-21393.
- [24] REN M Y, ZENG W Y, YANG B, et al. Learning to re-weight examples for robust deep learning[EB/OL]. (2018) [2022]. <http://arxiv.org/abs/1803.09050.pdf>.
- [25] NGUYEN D T, MUMMADI C K, NGO T P N, et al. SELF: Learning to filter noisy labels with self-ensembling [EB/OL]. (2019)[2022]. <http://arxiv.org/abs/1910.01842.pdf>.
- [26] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [27] REED S, LEE H, ANGUELOV D, et al. Training deep neural networks on noisy labels with bootstrapping[J]. Computer Science, 2014, 14(20): 687-698.
- [28] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[C]//Proceedings of Computer Vision - ECCV 2018: 15th European Conference, Part XIV. New York: ACM, 2018: 122-138.
- [29] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.
- [30] 卓力, 孙亮亮, 张辉, 等. 有噪声标注情况下的中医舌色分类方法[J]. 电子与信息学报, 2022, 44(1): 89-98.
- ZHUO L, SUN L L, ZHANG H, et al. TCM tongue color classification method under noisy labeling[J]. Journal of Electronics & Information Technology, 2022, 44(1): 89-98. (in Chinese)

- [31] 孙亮亮. 有噪声标注下的中医舌色分类研究[D]. 北京: 北京工业大学, 2022.  
SUN L L. Study on Classification of Tongue Color in Traditional Chinese Medicine with Noisy Labeling[D]. Beijing: Beijing University of Technology, 2022. (in Chinese)
- [32] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 2818-2826.
- [33] YI K, WU J X. Probabilistic end-to-end noise correction for learning with noisy labels[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 7010-7018.
- [34] PENG X J, WANG K, ZENG Z Y, et al. Suppressing mislabeled data via grouping and self-attention[C]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 786-802.
- [35] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2012, 60(6): 84-90.
- [36] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, 14(5): 1409-1556.
- [37] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 2261-2269.
- [38] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 5987-5995.
- [39] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 10778-10787.

#### 作者简介



孙亮亮 男, 1996年11月出生于河南省周口市. 现为北京工业大学信息学部2022届工程硕士研究生. 主要研究方向为机器学习和中医信息化处理.  
E-mail: 18845634374@163.com



李艳萍 女, 1997年4月出生于黑龙江省五大连池市. 现为北京工业大学信息学部2023届工学硕士研究生. 硕士期间获优秀毕业生、科技创新奖一等奖等奖项6项. 主要研究方向为机器学习和中医信息化处理.  
E-mail: yplingee@163.com



张辉 男. 北京工业大学信息学部电子科学与技术学院(微电子学院)副教授、硕士生导师. 毕业于北京理工大学信号与信息处理专业, 获工学博士学位. 主要从事视觉信息处理、人工智能系统设计与集成研究.  
E-mail: huizhang@bjut.edu.cn



卓力 女, 1971年10月出生于江苏省徐州市. 现为信息学部电子科学与技术学院人工智能与视觉大数据研究所所长、教授、博士生导师. 获得教育部高等学校科学研究优秀成果奖(科学技术)技术发明奖二等奖和北京市科学技术奖三等奖各1项(均排名第一)、IEEE最佳论文奖等. 在国内外发表学术论文200余篇.  
E-mail: zhuoli@bjut.edu.cn