

## 有噪声标注情况下的中医舌色分类方法

卓力\* 孙亮亮 张辉 李晓光 张菁

(北京工业大学信息学部 北京 100124)

(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:** 舌色是中医(TCM)望诊最关注的诊察特征之一,自动准确的舌色分类是舌诊客观化研究的重要内容。由于不同类别舌色之间的视觉界限存在模糊性以及医生标注者的主观性等,标注的舌象数据中常含有噪声,影响舌色分类模型的训练。为此,该文提出一种有噪声标注情况下的中医舌色分类方法:首先,提出一种两阶段的数据清洗方法,对含有噪声的标注样本进行识别,并进行清洗;其次,设计一种基于通道注意力机制的轻型卷积神经网络,通过增强特征的表达能力,实现舌色的准确分类;最后,提出一种带有噪声样本过滤机制的知识蒸馏策略,该策略中加入了由教师网络主导的噪声样本过滤机制,进一步剔除噪声样本,同时利用教师网络指导轻型卷积神经网络的训练,提升了分类性能。在自建的中医舌色分类数据集上的实验结果表明,该文提出的舌色分类方法能以较低的计算复杂度,显著提升分类的准确率,达到了93.88%。

**关键词:** 中医舌色分类; 噪声标注样本; 数据清洗; 知识蒸馏; 轻型卷积神经网络

中图分类号: TN911.73; TP391

文献标识码: A

文章编号: 1009-5896(2022)01-0089-10

DOI: 10.11999/JEIT210935

## TCM Tongue Color Classification Method under Noisy Labeling

ZHUO Li SUN Liangliang ZHANG Hui LI Xiaoguang ZHANG Jing

(Signal and Information Processing Laboratory, Beijing University of Technology, Beijing 100124, China)

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

**Abstract:** Tongue color is one of the most concerned diagnostic features of tongue diagnosis in Traditional Chinese Medicine (TCM). Automatic and accurate tongue color classification is an important content of the objectification of tongue diagnosis. Due to the vagueness of the visual boundaries between different types of tongue colors and the subjectivity of the doctors, the annotated tongue image data samples often contain noises, which has a negative effect on the training of the tongue color classification model. Therefore, in this paper, a tongue color classification method in TCM with noisy labels is proposed. Firstly, a two-stage data cleaning method is proposed to identify and clean noisy labeled samples. Secondly, a lightweight Convolutional Neural Network (CNN) based on the channel attention mechanism is designed in this paper to achieve accurate classification of tongue color by enhancing the expressiveness of features. Finally, a knowledge distillation strategy with a noise sample filtering mechanism is proposed. This strategy adds a noise sample filtering mechanism led by the teacher network to eliminate further noise samples. At the same time, the teacher network is used to guide the training of the light convolutional neural network to improve the classification performance. The experimental results on the self-established TCM tongue color classification dataset show that the proposed method in this paper can significantly improve the classification accuracy with lower computational complexity, reaching 93.88%.

**Key words:** Traditional Chinese Medicine (TCM) tongue color classification; Noise labeling samples; Data cleaning; Knowledge distillation; Light Convolutional Neural Network (CNN)

收稿日期: 2021-09-03; 改回日期: 2021-12-22; 网络出版: 2021-12-27

\*通信作者: 卓力 zhuoli@bjut.edu.cn

基金项目: 国家自然科学基金(61871006)

Foundation Item: The National Natural Science Foundation of China (61871006)

## 1 引言

舌诊是中医(Traditional Chinese Medicine, TCM)望诊中的核心内容。舌色作为舌象诊察重要特征之一,在中医上一般分为淡红舌、红舌、暗红舌、紫舌等4类,其颜色的变化反映了人体脏腑气血津液的虚实和邪正盛衰<sup>[1]</sup>。长期以来,舌诊主要是通过医生的目测观察、经验来判断病证。但在临床实践中,舌象的诊察会受到光线、温度等外界客观因素以及医生主观性的影响,进而影响后续的症候诊断及依证施方。利用计算机对中医舌象进行定量化、客观化和标准化的处理与分析,已成为当前舌诊客观化研究中的一个基础性问题。

中医舌诊客观化的研究起步于20世纪80年代初期,其中以中医舌质、苔质颜色自动分析为代表的工作可以分为两个阶段:第1阶段是传统方法,普遍采用人工特征+浅层机器学习的框架,对舌色、苔色等进行分析和处理;第2阶段是基于深度学习的方法,采用端到端的框架,实现中医舌象的智能分析与处理。

在第1阶段的研究工作中,通常提取舌图像的人工特征,如颜色直方图、颜色集、结构元素描述符<sup>[2]</sup>、颜色共生矩阵<sup>[3]</sup>、图像矩<sup>[4,5]</sup>等颜色特征,用于舌色分类。文献<sup>[6]</sup>提出了一种多因素相关的块截断编码用于提取图像的颜色特征,并在图像分类中取得较好的结果。文献<sup>[7]</sup>提出了一种AdaBoost级联分类框架,提高了舌质颜色分类的准确率。文献<sup>[8]</sup>在RGB空间提取了舌图像的颜色特征,采用支持向量机作为分类器,实现了舌色分类。文献<sup>[9]</sup>利用主色调直方图作为颜色特征,在多标记学习算法下对舌质与舌苔进行了分类。文献<sup>[10]</sup>在先验模板和标准的色块作为辅助基础上,利用K近邻(K-Nearest Neighbor, KNN)作为分类器进行分类。王爱民等人<sup>[11]</sup>采用Fisher比率作为色度空间选择的依据,实现了舌色和苔色的自动分类。

这一阶段的研究工作普遍人工提取舌图像的颜色特征进行分类,由于颜色特征对于光照变化特别敏感,因此这些方法的鲁棒性较差,分类准确性难以令人满意。

2012年以来,深度学习在图像分类、自然语言处理等领域取得了巨大突破。在第2阶段的研究工作中,研究者尝试将深度学习应用于中医舌诊研究中,利用深度学习强大的特征学习和语义表达能力,从大量的样本数据中自动学习舌图像的特征,实现舌色的分类。文献<sup>[12]</sup>采用改进后的CaffeNet网络结构对预处理后的舌图像进行分类。文献<sup>[13]</sup>结合图像预处理和深度学习,在舌图像数据集中验

证了自搭建的卷积神经网络(Convolutional Neural Network, CNN)的分类性能。文献<sup>[14]</sup>为了减小在舌色、苔色分类时,两者之间的相互干扰,将整张舌图像按照舌色、苔色分布的区域分离开来分别送入轻型CNN网络进行分类。在图像分类任务中,不同特征的融合往往会取得更好的分类结果,文献<sup>[15,16]</sup>提出了不同的视觉与文本特征融合的方法,并在各自的分类任务中取得了更优的性能。

研究表明,基于深度学习的方法可以获得远超过传统方法的性能,舌色分类精度有了显著的提升。但是,目前基于深度学习的中医舌色分类方法还存在以下问题需要解决:

(1)深度神经网络模型结构复杂,计算复杂度和存储需求会大大增加。同时,需要大量的有标注样本对网络进行训练,才能获得理想的性能。而在实际应用中,由于有经验的医生标注成本昂贵,数据采集困难,往往难以获得大规模、有标注的舌象样本。如何利用小规模的数据集,有针对性地设计深度网络模型,提升舌色分类的准确性,还需要进行深入的研究;

(2)受医生的知识水平、诊断经验的限制,也因为光线、环境等客观因素的影响,医生的标注样本中经常会出现错误,标注的样本数据中不可避免地会存在一定的噪声,噪声样本的存在会对分类网络模型的训练造成影响,导致舌色分类的鲁棒性和准确性难以令人满意。针对有噪声标注样本情况下的分类问题,学者开展了深入的研究工作,取得了有意义的研究进展。比如,文献<sup>[17]</sup>提出一种简单而有效的注意力特征混合(Attentive Feature Mixup, AFM)策略,利用注意力机制降低噪声样本的权重,进而抑制噪声样本对分类模型的影响。文献<sup>[18]</sup>提出了置信学习的思想,用于识别和过滤错误样本。在过滤错误样本后,调整样本类别权重,重新训练网络。文献<sup>[19]</sup>提出了一种框架用来构建健壮的损失函数,有效解决了损失函数对噪声样本的欠拟合问题。文献<sup>[20]</sup>提出了一种基于噪声分类器预测的标签校正算法,提升了分类网络模型的性能。文献<sup>[21]</sup>构建了网络参数和标签联合优化的框架,利用网络输出结果更新标签,提升分类性能。文献<sup>[22]</sup>通过构建一种自治愈网络实现对噪声样本的抑制和噪声标注的修正,提高了分类的精度。文献<sup>[23]</sup>提出了一种不一致的潜在真实伪标注框架IPA2LT,该框架在一定程度上解决了人脸表情注释的不一致性带来的问题,提升了表情分类的精度。

上述的研究结果表明,对有噪声的样本进行识

别并加以清除，可以有效提升分类模型的性能。但是目前，针对有噪声样本下的舌色分类的研究工作很少，当前舌色分类的鲁棒性和准确性难以令人满意，严重阻碍了舌诊客观化研究的进展。

为此，本文提出了一种有噪声标注情况下的中医舌色分类方法，目的是实现舌色的鲁棒、准确分类。主要的创新点包括：(1)提出了一种基于2阶段的数据清洗方法，用于对噪声标注样本进行识别并加以清除；(2)针对舌象样本数据规模小的问题，设计了一种带通道注意力机制(Channel Attention, CA)<sup>[24]</sup>的轻型卷积神经网络，以实现舌色的准确分类；(3)提出了一种带有噪声样本过滤机制的知识蒸馏策略，该策略中加入了由教师网络主导的噪声样本过滤机制，进一步剔除噪声样本，同时利用教师网络指导轻型卷积神经网络的训练，提升了分类模型的性能。在自建的中医学舌色分类数据集上的实验结果表明，与现有的分类网络模型相比，本文提出的方法能以较低的计算复杂度，显著提升分类性能，分类准确率达到93.88%。

## 2 有噪声标注情况下的中医舌色分类方法

本文提出的有噪声标注情况下的中医舌色分类方法整体框图如图1所示。下面介绍本文提出的两阶段数据清洗方法、轻型CNN网络结构、带有噪声样本过滤机制的知识蒸馏策略等几个关键部分的具体实现细节。

### 2.1 两阶段数据清洗方法

如上所述，由于舌色类别之间并没有一个严格的界限和医生标注者的主观性等因素的影响，舌象数据集中存在一些带有噪声标注的样本，影响分类模型的训练。为此，本文提出了一种基于2阶段的数据样本清洗方法，其处理流程如表1所示。在第

1阶段，采用ResNet18筛选出数据中的噪声样本。其基本思想是利用ResNet18网络从大部分干净样本中学习到规律，筛选出噪声样本。通过多次随机划分和迭代，实现对数据集的初筛。由于在第1阶段筛选数据的过程中，训练集中有噪声样本的存在，分类模型有可能会误将干净的样本筛选掉。因此，本文又设计了第2阶段，利用经过第1阶段清洗后的数据集微调ResNet18, VGGNet等CNN模型(均在ImageNet上进行了预训练)，对筛选出的噪声样本进行修正，尽可能避免误筛干净的样本。清洗后的样本将用于训练舌色分类网络模型。

### 2.2 轻型CNN网络结构

深度学习依赖大数据作为支撑。然而，由于医生标注数据的成本高，临床数据采集困难，因此难以获得大规模、有标注的中医舌色样本数据。训练样本数据量的不足很容易导致过拟合现象的发生，影响网络模型的训练性能。因此，本文设计了一种轻型CNN网络结构，以防止由于训练样本过少带来的过拟合问题，同时也加入了通道注意力机制，获得更具表达能力的深度特征。

#### 2.2.1 整体网络结构

本文设计的轻型CNN网络结构如图2所示。这是一个端到端的架构，输入为舌图像，输出为预测的舌色类别。网络由9层卷积层和2层全连接层组成。为了减少网络的参数量、避免过拟合现象的发生，除了第1层使用 $7 \times 7$ 的卷积核，其余卷积层均采用了 $3 \times 3$ 的卷积核。此外，每个卷积层之间均使用了批量正则化(Batch Normalization, BN)和ReLU激活操作；其次，为了增强梯度传播和网络的泛化能力，采用了残差网络的连接方式。具体来说，从第2个卷积层开始，每两个卷积层组成一个残差块；最后，在网络中加入了通道注意力机制，

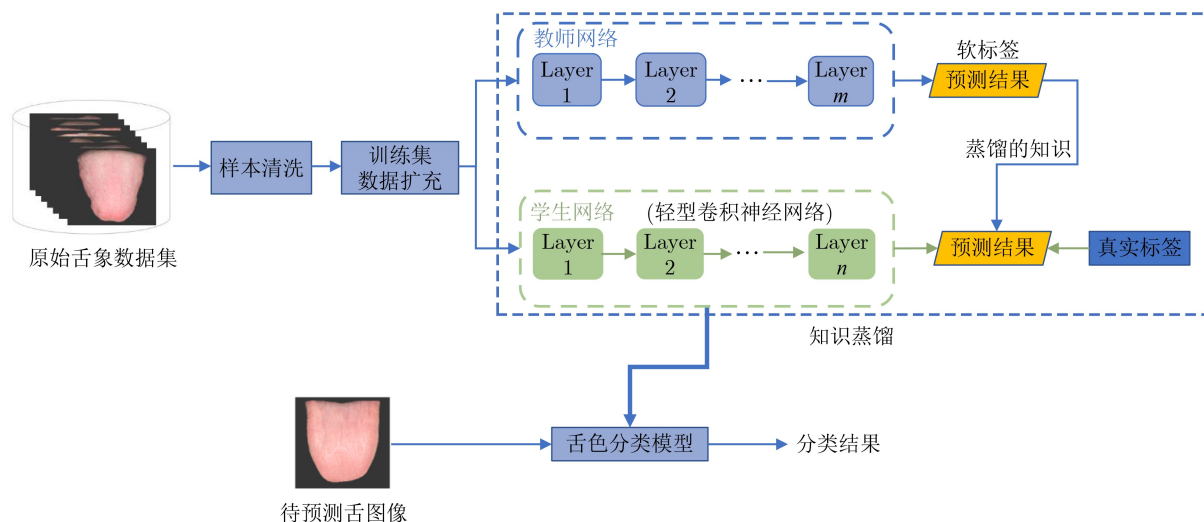


图1 有噪声标注情况下的中医舌色分类方法整体框图

表1 有噪声标准舌图像的数据清洗流程

输入: 原始数据集 $D_o = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$ ; 准确率阈值 $\text{Acc}_{\text{th}} = 0.9$ ; 变量 $a = 0$ ;
过程:
第1阶段:
(1) 将 $D_o$ 按照4:1随机划分为训练集和测试集;
$D_{\text{train}} = \{(x_{r+1}, y_{r+1}), (x_{r+2}, y_{r+2}), \dots, (x_{r+\text{len}(D_{\text{train}})}, y_{r+\text{len}(D_{\text{train}})})\}$
$D_{\text{test}} = \{(x_{e+1}, y_{e+1}), (x_{e+2}, y_{e+2}), \dots, (x_{e+\text{len}(D_{\text{test}})}, y_{e+\text{len}(D_{\text{test}})})\}$
(2) 用 $D_{\text{train}}$ 训练ResNet18网络;
(3) 用训练好的ResNet18模型对 $D_{\text{test}}$ 中样本进行预测得到 $\hat{y} = (\hat{y}_{e+1} \dots \hat{y}_{e+\text{len}(D_{\text{test}})})$ ;
(4) for $i$ in range ( $\text{len}(D_{\text{test}})$ ):
if $\hat{y}_{e+i} \neq y_{e+i}$ :
$D_{\text{test}}.\text{remove}(x_{e+i})$ ;
$D_{\text{delete}}.\text{add}(x_{e+i})$ ; // 将删除的数据放在一起组成
$D_{\text{delete}}$
$\text{Acc} = \text{len}(D_{\text{test}}) / (\text{len}(D_{\text{test}}) + \text{len}(D_{\text{delete}}) - a)$ ;
$a = a + \text{len}(D_{\text{delete}})$ ;
(5) while $\text{Acc} \leq \text{Acc}_{\text{th}}$ do
$D_{\text{onew}} \leftarrow D_{\text{train}} + D_{\text{test}}$ ; // 将去除噪声样本的
$D_{\text{train}}$ 与 $D_{\text{test}}$ 组合形成 $D_{\text{onew}}$
将步骤(1)中的 $D_o$ 换成 $D_{\text{onew}}$ , repeat 步骤(1)—步骤
(4)操作;
end while
第2阶段:
(6) 将最终的 $D_{\text{onew}}$ 训练ResNet18和VGG16;
(7) 利用各个网络训练的最优网络模型对 $D_{\text{delete}}$ 中的样本进行预测得到:
$\hat{y}_r = (\hat{y}_{r+1}, \hat{y}_{r+2}, \dots, \hat{y}_{r+\text{len}(D_{\text{delete}})})$ // ResNet18模型
的预测值
$\hat{y}_v = (\hat{y}_{v+1}, \hat{y}_{v+2}, \dots, \hat{y}_{v+\text{len}(D_{\text{delete}})})$ // VGG16模型
预测值
for $i$ in range ( $\text{len}(D_{\text{delete}})$ ):
if $\hat{y}_{r+i} == \hat{y}_{v+i}$ : // 当三者一致时, 将预测结果作为
样本新的标注
$y_i = \hat{y}_{v+i}$
$D_{\text{onew}}.\text{add}(x_{r+i})$
输出: 经过清洗后的舌象数据 $D_{\text{onew}}$

引导网络更关注对分类任务重要的特征通道, 抑制无关的通道。最后两个全连接层的神经元个数分别设置为128和4。

### 2.2.2 通道注意力机制

为了提升轻型CNN网络的特征提取与表达能力, 本文在网络中加入了通道注意力机制。通道注意力机制认为网络中每层不同通道的特征的重要性不同, 通过向重要的通道赋予更高的权重使得网络更加注重对分类任务有用的信息, 结构如图3所示。

在图3中, 输入的特征图 ( $H \times W \times C$ ) 分别经过最大池化和平均池化, 得到两个  $1 \times 1 \times C$  的特征图。接着, 将其分别送入一个多层感知机 (Multi-Layer Perception, MLP) 中。MLP主要用于压缩输入特征图的空间维数和生成对应的权重, 共有两层: 第1层神经元个数为  $C/r$  ( $r$ 为减少率, 本文中  $r$ 设为16), 激活函数为ReLU; 第2层神经元个数为  $C$ , 这两层的神经网络权值是共享的。之后, 将MLP输出的特征分别进行加法和Sigmoid激活操作, 生成最终的通道注意力特征, 即  $M_c(F)$ 。

整个通道注意力模块的计算过程可以表示为

$$F_{\text{mc}} = \text{Conv}_{1 \times 1}^C(\text{ReLU}(\text{Conv}_{1 \times 1}^{C/16}(\text{Max\_pooling}(F)))) \quad (1)$$

$$F_{\text{ac}} = \text{Conv}_{1 \times 1}^C(\text{ReLU}(\text{Conv}_{1 \times 1}^{C/16}(\text{Avg\_pooling}(F)))) \quad (2)$$

$$M_c(F) = \text{Sigmoid}(F_{\text{mc}} + F_{\text{ac}}) \quad (3)$$

### 2.3 带有噪声样本过滤机制的知识蒸馏策略

为了进一步筛选数据集噪声样本, 同时提高轻型CNN网络的分类性能, 本文提出了一种带有噪声样本过滤机制的知识蒸馏策略, 通过一个复杂的教师网络去指导轻型学生网络的学习。教师网络中加入噪声样本的过滤机制, 进一步滤除噪声样本。该过滤机制是基于噪声样本相对于干净样本的分类损失值会更大的思路, 利用教师网络计算每个批次中样本的分类损失值, 然后按照一定的比例过滤掉损失值较大的样本, 只把损失值小的样本用于训练学生网络。该策略可以在教师网络将知识传递

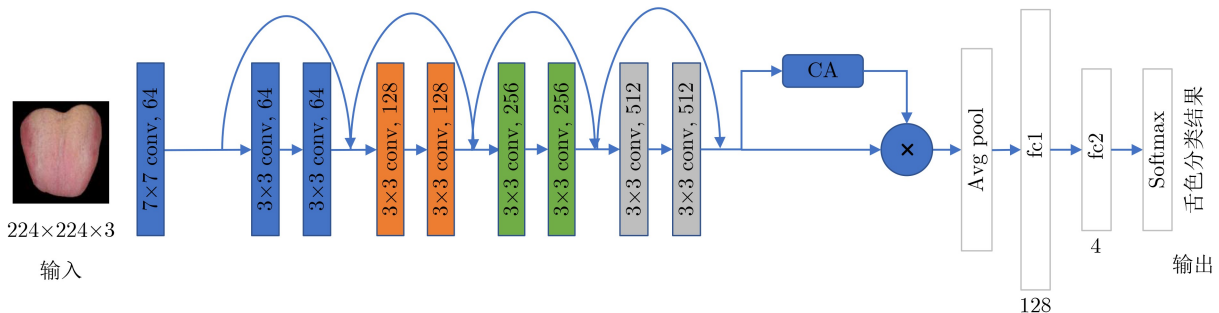


图2 轻型CNN网络结构图

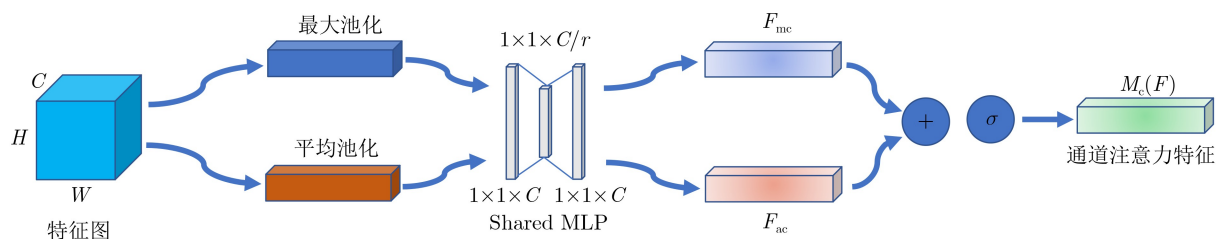


图3 通道注意力网络结构

到学生网络的同时，降低噪声样本对学生模型训练的干扰，使学生网络尽可能逼近亦或是超过教师网络的性能，从而以较低的网络复杂度，获得较高的分类性能。噪声样本的过滤流程如表2所示。

教师网络采用ResNet50<sup>[25]</sup>+SeNet<sup>[26]</sup>的网络结构，而学生网络则是本文设计的轻型CNN网络。知识蒸馏策略的整体结构如图4所示。采用在ImageNet上训练好的ResNet50+SeNet对输入样本进行分类，输出结果作为软标签，去指导轻型CNN网络的训练。相比于硬标签来说，软标签携带了更多有用的信息，可以有效提升学生网络的训练性能。

在知识蒸馏中，本文采用的目标函数为交叉熵损失函数与K-L散度损失的加权和，具体为

$$\text{Loss} = \alpha \text{Loss}_1 + (1-\alpha) \text{Loss}_2 \quad (4)$$

其中，Loss1表示交叉熵损失函数，Loss2表示K-L散度损失， $\alpha$ 表示加权值。Loss1和Loss2分别定义为

$$\text{Loss}_1 = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n p(x_{ij}) \log_2(q(x_{ij})) \quad (5)$$

$$\text{Loss}_2 = D_{\text{KL}}(\hat{q}||q) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n \hat{q}(x_{ij}) \log_2 \frac{\hat{q}(x_{ij})}{q(x_{ij})} \quad (6)$$

其中， $m$ 是样本总数， $n$ 是类别数， $p(x_{ij})$ 是第 $i$ 个样本第 $j$ 个类别的真实标签； $q(x_{ij})$ 是学生网络经过Softmax计算出的第 $i$ 个样本第 $j$ 个类别的概率值， $\hat{q}(x_{ij})$ 是教师网络经过Softmax计算出的第 $i$ 个样本第 $j$ 个类别的概率值。

利用这种带有噪声样本过滤机制的知识蒸馏策略，可以进一步去除有噪样本，同时提升轻型网络的分类性能。

## 2.4 舌色分类

对待预测舌图像进行预测时，首先利用上述训练好的网络模型提取到舌色特征，接着将特征送入分类器中进行分类。本文采用了Softmax分类器，通过计算当前样本属于每一类的概率，选择具有最高概率的类作为当前样本的类别。本文将Softmax分类器与轻型网络结构相结合，通过一个端到端的网络结构，实现舌色的分类。

表2 噪声样本过滤流程

输入：训练集 $D_t = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ； 教师网络 $\text{Model}_t$ 和已经训练好的模型参数 $\theta$ 以及损失函数 Loss； remember_rate=0.95； 单次送入网络的样本数量为batch_size；
--

过程：

- (1) 将训练集中的样本按照batch\_size大小送入教师网络中；
- (2) 利用教师网络 $\text{Model}_t$ 和训练好的模型参数 $\theta$ 得到batch\_size个样本的预测结果；
- (3) 利用损失函数Loss计算样本的损失值，将batch\_size个样本的损失值按照升序方式排序；
- (4) 设置remember\_rate将样本按照排序的结果保留前batch\_size×remember\_rate个样本，清除掉剩余的样本；
- (5) 将步骤(4)保留的样本更新送入学生网络样本。

输出：干净的数据样本，并将数据送入学生网络

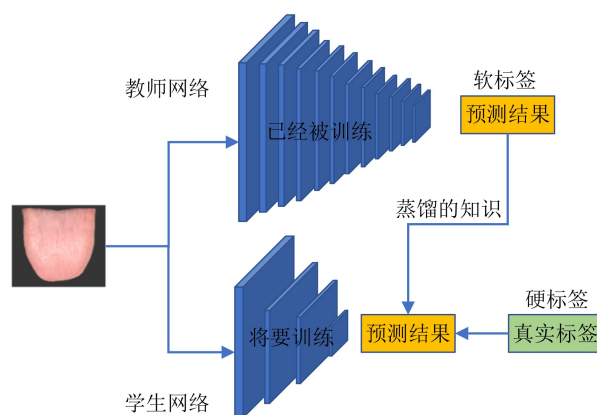


图4 知识蒸馏的结构图

## 3 实验结果与分析

为了验证所提方法的有效性，本文在自建的数据集上进行了实验。在实验中，首先对比了数据清洗前后的分类性能；其次，在清洗后的数据集上，对比了本文设计的轻型CNN网络与其他CNN网络的分类准确率；然后，又比较了采用知识蒸馏策略之后，轻型CNN网络的分类性能；最后，针对两阶段的数据清洗方法和轻型CNN网络分别进行了相关的消融实验，验证各个部分对分类性能的影响。

### 3.1 数据集和参数设置

数据集。本文自建了舌色分类数据集，所有数据样本均采集自首都医科大学附属宣武医院，采集设备为本课题组研制的中医舌象仪，每个样本由中医专家人工进行标注。数据集共包含舌图像样本430例，舌色共分为4类，其中淡红舌148例、红舌133例、暗红舌106例和紫舌43例。部分示例图像如图5所示。

分类性能评价指标。本文采用准确率来度量舌色分类性能，定义为

$$\text{Acc} = (\text{TP} + \text{TN}) / N \quad (7)$$

假设 $C$ 表示舌色类别中的一类，式(7)TP表示类 $C$ 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正确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。在训练轻型CNN网络时，为了提高模型的训练性能，避免过拟合，数据扩充是必不可少的。本文采用了多种数据扩充方法，包括随机水平和垂直翻转、在 $-45^\circ$ 和 $45^\circ$ 之间随机旋转，调整图像的大小等。输入网络的图像被统一调整为 $224 \text{像素} \times 224 \text{像素}$ 。

参数的设置。本文设计的轻型CNN网络是在Pytorch平台上搭建的，实验中配置了NVIDIA GeForce TX 1080 Ti GPU。采用Adam算法对模型进行优化，其中Batch Size设为64、权重衰减为 $10^{-4}$ 。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，损失函数使用的交叉熵损失，epoch的值为200。实验中的测试准确率均为每20轮取1次均值所得。

### 3.2 数据清洗对分类结果的影响

为了验证提出的两阶段数据清洗方法的有效性，本文进行了数据清洗前后的分类性能对比实验。首先，利用表1算法对原始数据集进行了清洗，共清洗出21例噪声样本。其中第1阶段筛选出的样本数量为74，第2阶段校正了53例样本。在清洗后的数据集中，舌图像数量为409例，其中淡红舌126例，红舌128例，暗红舌113例和紫舌42例。图6所示的是清洗出来的部分噪声样本示例。可以看出，这些舌图像在视觉上很难区分其归属舌色类别中哪一类，导致医生标注时很有可能给出不一致的结果。

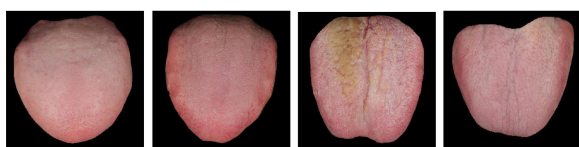


图5 舌色样本的示例

接下来，将清洗后的数据集按照4:1比例随机划分为训练集与测试集。在验证数据清洗对于分类性能的影响时，为了保证测试集的一致性，本文选取清洗后的测试集做测试，而清洗前训练集的划分则是由清洗前的数据集去除测试集之后得到的，之后采用清洗前后的训练集对在ImageNet上预训练后的ResNet18网络进行微调，数据扩充与参数设置均采用3.1节所述方式，得到优化后的分类网络模型。数据清洗前后的分类对比结果如表3所示。

从表3可以看出，数据清洗之后，无论是对于单个类别还是整体的分类性能都有了显著的提升。具体来说，淡红舌与红舌的分类准确率由75%，76.92%提升到了88%，88.46%，分别提升了13%和11.54%。这是因为通过清洗，清除了淡红舌与红舌中带有噪声标注的样本，提升了分类模型的性能。整体准确率在样本清洗前仅为82.34%，而清洗后达到了91.81%，提升了9.47%。上述的实验结果表明，本文提出的数据清洗方法可以有效筛选出带有噪声的标注样本，显著提升了分类准确率。

### 3.3 不同CNN网络结构的分类结果

为了验证本文设计的轻型CNN网络的性能，本文将该网络(未采用知识蒸馏策略)与AlexNet<sup>[27]</sup>，VGG16<sup>[28]</sup>，ResNet18和MobileNetV2<sup>[29]</sup>等几种主流的轻型CNN网络结构进行了对比。对比时，将清洗后的数据按照4:1比例随机划分为训练集和测试集。除了本文提出的轻型CNN网络，其他网络均在ImageNet上进行了预训练，采用训练集对网络模型进行微调。训练时采用了相同的数据扩充方法。

对比结果如图7所示。可以看出，上述几种主流的轻型CNN网络分别取得了89.53%，90.69%，90.86%，91.81%和91.86%的分类准确率。而本文设计的轻型CNN网络在没有预训练的情况下，仍然取得了92.87%的准确率，分类准确率提高了1.01%以上。这是因为本文设计的轻型CNN网络中加入了通道注意力机制，使得网络更加关注对分类有用的信息，从而提高了分类的准确率。

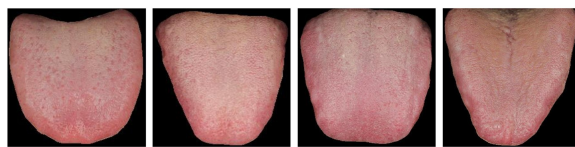


图6 清洗出的部分噪声样本示例

表3 数据清洗前后分类准确率(%)对比结果

	淡红舌	红舌	暗红舌	紫舌	整体精度
清洗前	75	76.92	91.30	88.89	82.34
清洗后	88	88.46	95.65	100	91.81

表4所示的是几种CNN网络结构的参数量。可以看出，相对于AlexNet, VGG16和ResNet18等网络结构，本文提出的轻型CNN网络的参数量小了很多。虽然比MobileNetV2结构的参数量大了1 MB，但是网络的分类精度却提高了1.01%以上。综上，本文设计的轻型CNN网络不仅网络结构简单，同时也能取得更优的分类性能。

### 3.4 知识蒸馏策略对分类结果的影响

为了验证所提带有噪声过滤机制的知识蒸馏策略对于提升轻型CNN网络性能的有效性，本文将采用知识蒸馏策略前后的轻型CNN网络分类性能做了对比。教师网络分别采用了ResNet18+CBAM, ResNet50, ResNext50<sup>[30]</sup>, ResNet101和ResNet50+SeNet等5种复杂的CNN网络结构，网络训练方式如2.3节所述。采用知识蒸馏策略前后得到的实验结果如表5所示。

从表5结果可以看出，在没有使用知识蒸馏的情况下，轻型CNN网络的分类准确率为92.87%。而使用了不同的CNN网络模型作为教师网络时，轻型CNN网络的分类准确率均有所提升，尤其是采用ResNet50+SeNet作为教师网络时，轻型

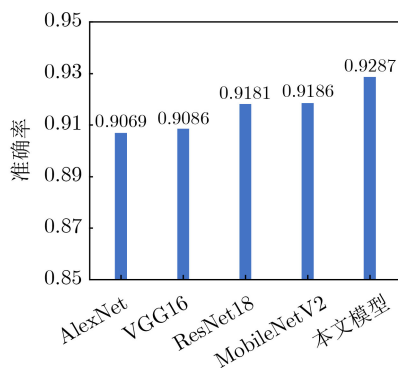


图7 不同分类模型准确率对比

表4 各种CNN网络结构的参数量(MB)

	AlexNet	VGG16	ResNet18	MobileNet V2	轻型CNN网络
参数量	61.1	138.4	11.2	4.2	5.2

CNN网络的分类性能达到了最优，为93.88%，提升了约1%。尽管轻型CNN网络在分类的性能上可能略低于教师网络(低0.59%)，但是在网络参数量方面要远小于教师网络。例如ResNet50+SeNet网络的参数量为28.1 MB，而本文设计的轻型CNN网络只有5.2 MB，仅为18.5%。由以上实验结果可以看出，采用本文提出的知识蒸馏策略，在网络参数量不变的情况下，可以进一步提升轻型网络模型的性能。

### 3.5 不同分类网络的结果

为了验证所提出方法的有效性，本文将其与目前主流的几种分类网络进行了比较，主要包含DenseNet121<sup>[31]</sup>, ResNet18, ResNext50, ShuffleNetV2<sup>[32]</sup>, MobileNetV2, EfficientNet-b4<sup>[33]</sup>等。以上网络均使用3.2节的测试集做测试，训练集由清洗前的数据集除去测试集组成，参数设置均采用3.1节所述方式。具体实验结果如表6所示。

从表6可以看出，本文所提方法取得了最优的分类性能，相对于其他分类网络，提高了10.96%以上。主要原因在于本文所提方法对有噪声标注的样本进行了清洗，避免了这类样本对网络模型产生不利的影响。实验结果再次证明，噪声样本会对分类精度产生非常严重的影响，如果不进行数据清洗，即便采用先进的网络结构，也无法获得令人满意的分类精度。

### 3.6 消融实验

为了验证本文方法的不同部分对分类性能的作用，本文进行了消融实验。下面介绍消融实验结果。

(1)两阶段数据清洗。在实验中，分别利用原始舌象数据集与经过第1阶段和第2阶段清洗后的数据集训练ResNet18网络，然后对比训练后的模型在相同测试集上的分类准确率，结果如表7所示。数据集均按照训练集与测试集4:1比例随机划分，实验参数的设置如步骤3.1所述。从表7可以看出，经过第1阶段后得到的舌象数据集相对于原始数据集在准确率方面提升了7.68%，主要原因是第1阶段

表5 采用知识蒸馏前后的对比实验结果

教师网络			学生网络		
名称	精度	参数量(MB)	名称	精度	参数量(MB)
×	×	×		0.9287	
ResNet18+CBAM	0.9418	11.8		0.9345	
ResNet50	0.9415	25.6	轻型卷积神经网络	0.9356	5.2
ResNext50	0.9423	25.1		0.9345	
ResNet101	0.9435	44.6		0.9352	
ResNet50+SeNet	0.9447	28.1		0.9388	

清洗掉了有噪样本,提升了标签的准确性。而第2阶段在第1阶段的基础上,准确率又提升了1.79%,其主要原因是通过对噪声样本的修正,增加了训练样本的数量,从而获得了比第1阶段更高的分类准确率。

(2)轻型CNN网络。为了验证轻型CNN网络中通道注意力机制模块的有效性,本文在知识蒸馏的基础上,进行了相关的消融研究。结果如表8所示。其中基线模型包含9层卷积层和2层全连接层,然后将通道注意力机制模块嵌入到基线模型中,验证其对于性能提升的作用。从表8可以得到,采用通道注意力机制,分类的准确率提升了1.1%,这说明通道注意力机制可以提升特征的表达能力,从而提升分类性能。

#### 4 结束语

本文针对中医舌色分类中存在的噪声样本问题,提出了一种有噪声标注情况下的中医舌色分类方法。首先,提出一种两阶段的数据清洗方法,可以对噪声样本进行筛选,并加以清除;然后,设计了一种带有通道注意力机制的轻型CNN网络结构,用于舌色分类;此外,为了进一步提升网络的性能,采用了带有噪声样本筛选机制的知识蒸馏策略,由教师网络主导对噪声样本进行过滤,利用较少的网络参数量就可以获得较高的分类性能。实验结果表明,采用本文提出的方法,可以显著提升带

有噪声标注的舌色分类准确性。鉴于实际应用中,有标注的中医舌象样本稀缺,且代价昂贵,未来的工作将考虑采用半监督学习的方式,利用大量的无监督样本,进一步提升舌色分类的性能。

#### 参考文献

- [1] 李乃民. 中国舌诊大全[M]. 北京: 学苑出版社, 1994: 30-170.  
LI Naimin. Complete Collection of Tongue Diagnosis of Traditional Chinese Medicine[M]. Beijing: Academy Press, 1994: 30-170.
- [2] WANG Xingyuan and WANG Zongyu. A novel method for image retrieval based on structure elements' descriptor[J]. *Journal of Visual Communication and Image Representation*, 2013, 24(1): 63-74. doi: 10.1016/j.jvcir.2012.10.003.
- [3] WANG Xingyuan, CHEN Zhifeng, and YUN Jiaojiao. An effective method for color image retrieval based on texture[J]. *Computer Standards & Interfaces*, 2012, 34(1): 31-35. doi: 10.1016/j.csi.2011.05.001.
- [4] WANG Chunpeng, WANG Xingyuan, XIA Zhiqiu, et al. Image description with polar harmonic Fourier moments[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(12): 4440-4452. doi: 10.1109/TCSVT.2019.2960507.
- [5] XIA Zhiqiu, WANG Xingyuan, ZHOU Wenjie, et al. Color medical image lossless watermarking using chaotic system and accurate quaternion polar harmonic transforms[J]. *Signal Processing*, 2019, 157: 108-118. doi: 10.1016/j.sigpro.2018.11.011.
- [6] WANG Xingyuan and WANG Zongyu. The method for image retrieval based on multi-factors correlation utilizing block truncation coding[J]. *Pattern Recognition*, 2014, 47(10): 3293-3303. doi: 10.1016/j.patcog.2014.04.020.
- [7] 王奕然, 张新峰. 基于AdaBoost级联框架的舌色分类[J]. 北京生物医学工程, 2020, 39(1): 8-14. doi: 10.3969/j.issn.1002-3208.2020.01.002.  
WANG Yiran and ZHANG Xinfeng. Tongue color classification based on AdaBoost cascade framework[J]. *Beijing Biomedical Engineering*, 2020, 39(1): 8-14. doi: 10.3969/j.issn.1002-3208.2020.01.002.
- [8] 张艺凡, 胡广芹, 张新峰. 基于支持向量机的痤疮患者舌色苔色识别算法研究[J]. 北京生物医学工程, 2016, 35(1): 7-11. doi: 10.3969/j.issn.1002-3208.2016.01.02.  
ZHANG Yifan, HU Guangqin, and ZHANG Xinfeng. An algorithm study on tongue color recognition of patients with acne based on support vector machine[J]. *Beijing Biomedical Engineering*, 2016, 35(1): 7-11. doi: 10.3969/j.

表 6 不同分类网络的比较结果(%)

	淡红舌	红舌	暗红舌	紫舌	整体精度
DenseNet121	82.60	68.00	84.61	88.89	82.92
ResNet18	75.00	76.92	91.30	88.89	82.34
ResNext50	78.26	84.00	80.76	77.78	80.72
ShuffleNetV2	86.95	76.00	76.92	88.89	81.14
MobileNetV2	86.95	80.00	76.92	77.78	81.32
EfficientNet-b4	86.95	64.00	84.61	77.78	81.45
本文方法	91.30	92.00	96.15	100.00	93.88

表 7 基于两阶段数据清洗方法的消融实验结果

	原始数据集	第1阶段	第2阶段	准确率(%)
方式1	√	×	×	82.34
方式2	√	√	×	90.02
方式3	√	√	√	91.81

表 8 轻型CNN网络的消融实验结果

知识蒸馏	基线模型	通道注意力	准确率(%)
√	√	×	92.78
√	√	√	93.88



- issn.1002-3208.2016.01.02.
- [9] 张静, 张新峰, 王亚真, 等. 多标记学习在中医舌象分类中的研究[J]. 北京生物医学工程, 2016, 35(2): 111–116. doi: [10.3969/j.issn.1002-3208.2016.02.01](https://doi.org/10.3969/j.issn.1002-3208.2016.02.01).  
ZHANG Jing, ZHANG Xinfeng, WANG Yazhen, *et al.* Research on multi-label learning in the classification of tongue images in TCM[J]. *Beijing Biomedical Engineering*, 2016, 35(2): 111–116. doi: [10.3969/j.issn.1002-3208.2016.02.01](https://doi.org/10.3969/j.issn.1002-3208.2016.02.01).
- [10] 王永刚, 杨杰, 周越. 中医舌象颜色识别的研究[J]. 生物医学工程学杂志, 2005, 22(6): 1116–1120. doi: [10.3321/j.issn:1001-5515.2005.06.008](https://doi.org/10.3321/j.issn:1001-5515.2005.06.008).  
WANG Yonggang, YANG Jie, and ZHOU Yue. Tongue image color recognition in traditional Chinese medicine[J]. *Journal of Biomedical Engineering*, 2005, 22(6): 1116–1120. doi: [10.3321/j.issn:1001-5515.2005.06.008](https://doi.org/10.3321/j.issn:1001-5515.2005.06.008).
- [11] 王爱民, 赵忠旭, 沈兰荪. 中医舌象自动分析中舌色、苔色分类方法的研究[J]. 北京生物医学工程, 2000, 19(3): 136–142. doi: [10.3969/j.issn.1002-3208.2000.03.002](https://doi.org/10.3969/j.issn.1002-3208.2000.03.002).  
WANG Aimin, ZHAO Zhongxu, and SHEN Lansun. Research on the tongue color classification in automatic tongue analysis of traditional Chinese medicine[J]. *Beijing Biomedical Engineering*, 2000, 19(3): 136–142. doi: [10.3969/j.issn.1002-3208.2000.03.002](https://doi.org/10.3969/j.issn.1002-3208.2000.03.002).
- [12] Hou Jun, Su Hongyi, Yan Bo, *et al.* Classification of tongue color based on CNN[C]. Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis, Beijing, China, 2017: 725–729. doi: [10.1109/ICBDA.2017.8078731](https://doi.org/10.1109/ICBDA.2017.8078731).
- [13] FU Shengyu, ZHENG Hong, YANG Zijiang, *et al.* Computerized tongue coating nature diagnosis using convolutional neural network[C]. The 2017 IEEE 2nd International Conference on Big Data Analysis, Beijing, China, 2017: 730–734. doi: [10.1109/ICBDA.2017.8078732](https://doi.org/10.1109/ICBDA.2017.8078732).
- [14] 肖庆新. 基于深度学习的中医舌象特征分类技术研究[D]. [硕士学位论文], 北京工业大学, 2019.  
XIAO Qingxin. Research on TCM tongue feature classification technology based on deep learning[D]. [Master dissertation], Beijing University of Technology, 2019.
- [15] UNAR S, WANG Xingyuan, and ZHANG Chuan. Visual and textual information fusion using Kernel method for content based image retrieval[J]. *Information Fusion*, 2018, 44: 176–187. doi: [10.1016/j.inffus.2018.03.006](https://doi.org/10.1016/j.inffus.2018.03.006).
- [16] UNAR S, WANG Xingyuan, WANG Chunpeng, *et al.* A decisive content based image retrieval approach for feature fusion in visual and textual images[J]. *Knowledge-Based Systems*, 2019, 179: 8–20. doi: [10.1016/j.knosys.2019.05.001](https://doi.org/10.1016/j.knosys.2019.05.001).
- [17] PENG Xiaojiang, WANG Kai, ZENG Zhaoyang, *et al.* Suppressing mislabeled data via grouping and self-attention[C]. The 16th European Conference on Computer Vision, Glasgow, UK, 2020: 786–802. doi: [10.1007/978-3-030-58517-4\\_46](https://doi.org/10.1007/978-3-030-58517-4_46).
- [18] NORTH CUTT C G, JIANG Lu, and CHUANG I L. Confident learning: Estimating uncertainty in dataset labels[J]. *Journal of Artificial Intelligence Research*, 2021, 70: 1373–1411. doi: [10.1613/jair.1.12125](https://doi.org/10.1613/jair.1.12125).
- [19] MA Xingjun, HUANG Hanxun, WANG Yisen, *et al.* Normalized loss functions for deep learning with noisy labels[C]. The 37th International Conference on Machine Learning, Vienna, Austria, 2020: 6543–6553.
- [20] ZHENG Songzhu, WU Pengxiang, GOSWAMI A, *et al.* Error-bounded correction of noisy labels[C]. The 37th International Conference on Machine Learning, Vienna, Austria, 2020: 11447–11457.
- [21] YI Kun and WU Jianxin. Probabilistic end-to-end noise correction for learning with noisy labels[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 7010–7018. doi: [10.1109/CVPR.2019.00718](https://doi.org/10.1109/CVPR.2019.00718).
- [22] WANG Kai, PENG Xiaojiang, YANG Jianfei, *et al.* Suppressing uncertainties for large-scale facial expression recognition[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 6896–6905. doi: [10.1109/CVPR42600.2020.00693](https://doi.org/10.1109/CVPR42600.2020.00693).
- [23] ZENG Jiabei, SHAN Shiguang, and CHEN Xilin. Facial expression recognition with inconsistently annotated datasets[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 227–243. doi: [10.1007/978-3-030-01261-8\\_14](https://doi.org/10.1007/978-3-030-01261-8_14).
- [24] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional block attention module[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: [10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [25] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [26] HU Jie, SHEN Li, and SUN Gang. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141. doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).
- [27] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E.

- ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012: 1097–1105. doi: [10.5555/2999134.2999257](https://doi.org/10.5555/2999134.2999257).
- [28] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. The 3rd International Conference on Learning Representations, San Diego, USA, 2015: 1409–1556.
- [29] SANDLER M, HOWARD A, ZHU Menglong, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4510–4520. doi: [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474).
- [30] XIE Saining, GIRSHICK R, DOLLÁR P, *et al.* Aggregated residual transformations for deep neural networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5987–5995. doi: [10.1109/CVPR.2017.634](https://doi.org/10.1109/CVPR.2017.634).
- [31] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, *et al.* Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2261–2269. doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [32] MA Ningning, ZHANG Xiangyu, ZHENG Haitao, *et al.* ShuffleNet V2: Practical guidelines for efficient CNN architecture design[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 122–138.
- [33] TAN Mingxing and LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]. The 36th International Conference on Machine Learning, Long Beach, USA, 2019: 6105–6114.
- 卓 力: 女, 1971年生, 教授, 博士生导师, 研究方向为智能化信息处理、图像/视频编码与传输.
- 孙亮亮: 男, 1996年生, 硕士生, 研究方向为人工智能系统设计与集成.
- 张 辉: 男, 1982年生, 副教授, 硕士生导师, 研究方向为视觉目标检测与跟踪.
- 李晓光: 男, 1980年生, 副教授, 硕士生导师, 研究方向为图像视频信号处理、超分辨率图像复原.
- 张 菁: 女, 1975年生, 教授, 博士生导师, 研究方向为多媒体信息检索.

责任编辑: 马秀强