

基于深度学习的小样本中药材粉末显微图像识别

王一丁¹, 郝晨宇¹, 李耀利², 蔡少青^{2*}, 袁媛³

(1. 北方工业大学信息学院, 北京 100144; 2. 北京大学药学院, 北京 100191; 3. 中国中医科学院中药资源中心, 北京 100700)

(* 通信作者电子邮箱 sqcai@bjmu.edu.cn)

摘要:针对中药材种类繁多、数据量稀少以及难以对其导管进行分类的问题,提出一种基于多通道颜色空间与注意力机制模型的卷积神经网络改进方法。首先,采用多通道颜色空间将RGB颜色空间与其他颜色空间合并为6通道作为网络输入,使网络学习亮度、色调和饱和度等特征信息,弥补数据量的不足;其次,在网络中加入注意力机制模型,其中通道注意力模型将两个池化层紧密连接到一起,空间注意力模型将多尺度空洞卷积结合到一起,使网络将注意力聚焦于小样本中关键的特征信息。实验结果表明,针对34种中药材样本的8774张导管图像,采用多通道颜色空间和注意力机制模型的方法,与原始ResNet网络相比,准确率分别提升了1.8个百分点和3.1个百分点,将二者结合后准确率提升了4.1个百分点,说明所提方法对小样本分类的准确率有着大幅度的提升。

关键词:小样本数据;注意力机制;颜色空间;卷积神经网络;中药材粉末

中图分类号:TP391.41 **文献标志码:**A

Microscopic image identification for small-sample Chinese medicinal materials powder based on deep learning

WANG Yiding¹, HAO Chenyu¹, LI Yaoli², CAI Shaoqing^{2*}, YUAN Yuan³

(1. School of Information, North China University of Technology, Beijing 100144, China;

2. School of Pharmaceutical Sciences, Peking University, Beijing 100191, China;

3. National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China)

Abstract: Aiming at the problems that a wide variety of Chinese medicinal materials have small samples, and it is difficult to classify the vessels of them, an improved convolutional neural network method was proposed based on multi-channel color space and attention mechanism model. Firstly, the multi-channel color space was used to merge the RGB color space with other color spaces into 6 channels as the network input, so that the network was able to learn the characteristic information such as brightness, hue and saturation to make up for the insufficient samples. Secondly, the attention mechanism model was added to the network, in which the two pooling layers were connected tightly by the channel attention model, and the multi-scale cavity convolutions were combined by the spatial attention model, so that the network focused on the key feature information in the small samples. Aiming at 8774 vessel images of 34 samples collected from Chinese medicinal materials, the experimental results show that by using the multi-channel color space and attention mechanism model method, compared with the original ResNet network, the accuracy is increased by 1.8 percentage points and 3.1 percentage points respectively, and the combination of the two methods increases accuracy by 4.1 percentage points. It can be seen that the proposed method greatly improves the accuracy of small-sample classification.

Key words: small-sample data; attention mechanism; color space; Convolutional Neural Network (CNN); Chinese medicinal materials powder

0 引言

中药材粉末显微鉴定是通过在显微镜下观察药材粉末中组织、细胞和内含物等细微特征,对药材进行鉴定的一种方法。自1838年德国学者Schleiden阐明细胞是构成植物体的基本单位后,英国、德国、美国和日本等国家对于常用药材或国家药典及药品处方集中所收录的药材进行了显微鉴定研究。

徐国钧院士作为我国中药材显微鉴定领域的奠基人之一,于1951年,就发表了包含101种中药材的《粉末生药鉴定检索表》;于1986年,出版了研究专著《中药材粉末显微鉴定》^[1]。1977年版的《中华人民共和国药典》^[2]首次收录了中药材显微鉴别项目。1999年,赵中振等^[3]出版了《中华人民共和国药典中药材粉末显微鉴别彩色图集》,并于2005年和2016年,先后出版了《中药显微鉴别图鉴》^[4]和《中药显微鉴定图鉴》^[5]两本

收稿日期:2019-09-26;修回日期:2019-11-01;录用日期:2019-11-01。

基金项目:中医药行业科研专项(201407003);中央本级重大增减支项目(2060302)。

作者简介:王一丁(1967—),男,北京人,教授,博士,CCF会员,主要研究方向:生物特征识别、机器视觉;郝晨宇(1996—),男,北京人,硕士研究生,主要研究方向:计算机视觉、深度学习;李耀利(1977—),女,陕西宝鸡人,讲师,博士,主要研究方向:中药鉴定和质量评价;蔡少青(1960—),男,宁夏银川人,教授,博士,主要研究方向:生药品种鉴定和品质评价;袁媛(1978—),女,河北保定人,研究员,博士,主要研究方向:中药鉴定、分子药理学。

专著。这些前辈的研究和专著的出版,为中药材粉末显微鉴定提供了大量的图像和文字描述。但是,目前中药材粉末显微鉴定采用人工对比的方式,即将显微镜下所见图像与各类专著中图像和文字进行对比鉴别。因此,此种方法需要专业人士的知识和经验,对于非专业人士来说比较困难,导致中药材粉末显微鉴别方法受到了极大的限制。

迄今记载中药材品种最多的《中华本草》记载了我国药材品种 8 980 个,其中常用的约 500 种。依据天然属性和药用部位,中药材分为植物药、动物药和矿物药。植物药再依不同的药用部位分为根和根茎类、茎木类、皮类、叶类、花类、果实和种子类及全草类等。植物药材的粉末在显微镜下可见多种组织、细胞和内含物,包括导管、纤维、石细胞、淀粉粒、结晶、毛茸、表皮、块状物等,其中,导管依据细胞壁的加厚纹理可分为环纹导管、螺纹导管、梯纹导管、网纹导管以及具缘纹孔导管等,五类导管示意图见图 1。一般中药材粉末中可见 1~3 种导管类型,不同中药材导管的直径以及类型也有所不同。因为导管丰富的纹理细节,所以选其作为本文的研究对象。

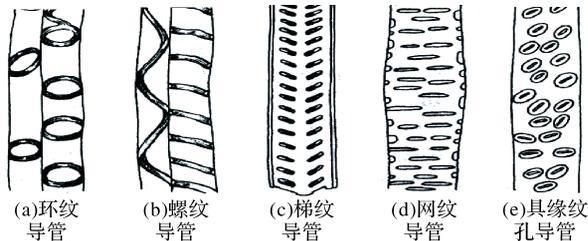


图 1 中药材导管的五种类型

Fig. 1 Five types of Chinese medicinal material vessels

针对导管图像,本文选用基于深度学习的中药材粉末分类识别方法^[6],可以由计算机部分替代专业人士,通过卷积神经网络(Convolutional Neural Network, CNN)的方法对显微特征图像进行自动分类与识别。

深度学习这一概念是由 Hinton 等^[7]于 2006 年提出,它具有的多隐层神经网络具有优异的特征学习能力,通过构建多层网络,对目标采取多层表达的方式,通过高层次特征来表达样本所具有的抽象语义信息。随着深度学习这一方法的盛行,经过大量学者的研究,设计出了多种结构的 CNN,如 LeNet^[8]、VGG^[9]、Inception^[10]、ResNet^[11]等。CNN 作为一种多层的监督学习神经网络,通过大规模样本的自我学习和重复迭代训练以及局部感受野对图像信息的提取,实现了分类器对图像特征的分类。本文选用的网络为 ResNet-v2-101^[12],它具有一定的网络的深度,可以保障充分学习中药材粉末显微特征的高层次信息;并且具有残差结构,可以在一定程度上缓和在网络训练过程中的过拟合现象。

本文将在王一丁等^[6]研究的基础上将中药材种类由 15 种扩增至 34 种,通过 ResNet-v2-101 网络对每类中药材粉末的导管进行分类识别。随着中药材种类的增多,每类中药材样本数量稀少的问题尤为凸显,这也导致了网络学习的特征信息不足,存在难以对导管精确分类的问题,并且不同的显微镜成像系统采集中药材粉末显微特征时,由于光源、滤光片、白平衡和操作人员等不同对图像的亮度等信息造成一定的差异,同样会对导管分类产生一定的误差。

针对中药材粉末显微特征样本所存在样本量稀少以及跨设备采集所造成图像差异的问题,本文提出了一种 ResNet-

v2-101 网络的改进方法。通过在网络输入端对样本的信息扩充,在残差模块(Residual)后加入通道注意力模块(Channel Attention)和空间注意力模块(Spatial Attention),得到权重的分配,从而实现小样本数据分类效果的提升。网络结构的改进如图 2 所示。

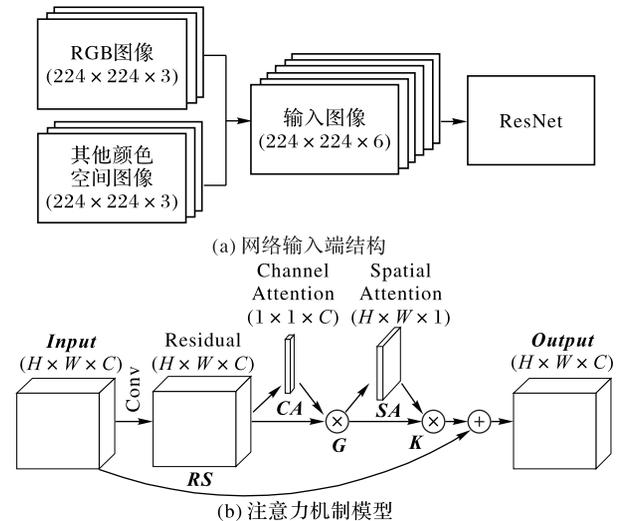


图 2 网络改进结构

Fig. 2 Network improvement structure

本文主要工作分为以下两部分:

1) 提出多通道颜色空间。通过多通道与颜色空间相结合作为网络输入端,实现对小样本数据量不足的补充,如图 2(a) 所示。网络通过 HSV (Hue, Saturation, Value) 颜色空间作为输入信息的补充,其可以充分学习到亮度、色调、饱和度等特征信息,消除不同显微镜成像系统采集所带来的影响。实验证明色彩特征信息对中药材显微图像更为关键。

2) 提出改进的注意力机制模型。通道注意力机制模型和空间注意力机制模型,如图 2(b) 所示。CNN 训练小样本数据的轮次(Epoch)不能太多或太少,否则都会导致识别效果达不到预期,并且不是所有特征信息都需要被网络所关注,因此本文通过此两个模型实现网络对通道和空间上权重的分配,使得网络将注意力集中到重要的通道和空间上,从而学习关键的特征信息。

1 多通道颜色空间

1.1 中药材粉末显微特征图像采集与预处理

按照以前报道的方法,本文收集了 34 种中药材中 8 774 张原始的导管图像,其中每张原始图像的分辨率达到 5 440×3 648。然而大背景下的中药材显微图像由于其显微特征在图像中所占比例过小,且一张图像中不能保证只有一种显微特征,因此并不能直接用于分类网络的训练学习,所以先经过筛选和剪切,得到了导管的最大外接图像,部分中药材导管如图 3 所示。然而即使采用最大外接图像,中药材导管间的差异并不明显,因此需要通过网络学习最大外接图像导管之间的细微差异。

经筛选和裁剪,样本一共有 8 774 张图像,设置原始训练集 7 986 张(91%),测试集 788 张(9%),其中由于中药材类别众多,且存在样本不平衡问题,每类中药材的导管样本数量由几十到几百不等,本文为保障 CNN 在学习过程中不会由数据

极度不平衡问题而产生过拟合问题,测试集中每类中药材图像数量占原始数据集中该类中药材数量的9%。

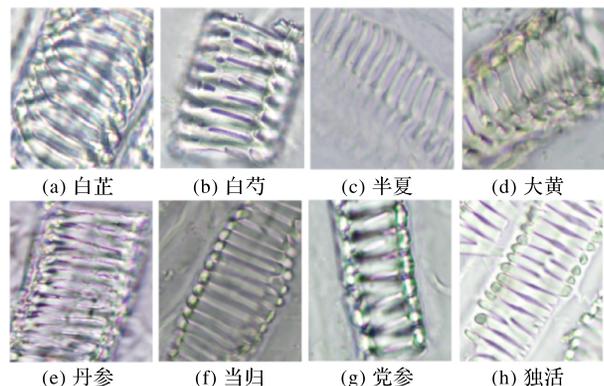


图3 中药材导管的显微特征

Fig. 3 Microscopic features of Chinese medicinal material vessels

在划分训练集和验证集后,需对图像进行预处理操作,其中包含随机裁剪、多角度旋转、去均值和尺度归一化。图像预处理的目的一方面是为了更方便让CNN学习图像间细微区别而进行相应的处理,另一方面则是扩充数据集。本文原始训练集的10%(798张)作为网络训练的验证集,剩余原始训练集在经图像预处理(随机裁剪和多角度旋转)扩充后,训练集数量达到21564张,是原始训练集的2.7倍。

1.2 多通道颜色空间

现阶段大多数分类网络创新在于网络构架的优化,如多尺度融合、批归一化等,其本质是为了更加有效地学习图像间的特征信息。但是现在分类网络的预处理方式大多只在网络输入端进行简单的平移、翻转和随机裁剪等扩充,对网络准确率提升的帮助有限。本文的中药材粉末显微特征图像存在着样本数量稀少问题,并且不同显微镜成像系统所采集的图像也存在着一定的亮度、色调和饱和度等差异。为了能更多地扩充特征信息,减少不同显微镜成像系统采集的差异,本文将在原有图像预处理的基础上探索多通道颜色空间方式能否提升小样本数据分类网络的识别效果。

颜色空间是对色彩的组织形式,借物理设备的采集,得以通过颜色空间得到色彩的固定模拟和数字表示,其中每一个颜色定义的名字和代码都基于严谨的数学定义。在深度学习中被普遍运用的颜色空间为RGB(Red, Green, Blue)颜色空间,其中红色、绿色和蓝色三种原色作为基础,通过对此三种颜色不同比例配比组成繁多的色彩,从而形成了RGB颜色空间,其中每一种颜色由24b代表,每个通道R、G、B各出8b,因此导致每个通道依据灰阶划分出了256个亮度等级。

近些年不同颜色空间这一概念一直被运用到肤色检测^[13]和手势识别^[14]中。在2018年陈昌红等^[15]提出了基于多通道和卷积神经网络的极光分类,首次将多通道融合运用到分类网络中,但是由于此篇文章所研究的是极光分类问题属于特殊性质的图像分类,多通道包含极光的局部二值模式(Local Binary Pattern, LBP)特征图和变化幅度等信息,并不能推广到普通图像的分类问题上;且对于普通图像的分类问题,三原色调配颜色的方式不能很好体现中药材粉末显微特征图像颜色、形状等特性;因此本文将多通道与不同颜色空间相结合,将其运用到分类网络中,通过实验论证颜色空间的多通道融合可以在有限的样本下提供更多的特征信息,以此来实现提

升小样本数据分类效果的目的。根据2011年江凤兵^[16]提出的关于人体肤色检测论文对于颜色空间的改进,将HSV空间和YCrCb空间相结合形成新的HCrCb空间,与原始图像相结合形成六通道的RGB-H-CrCb作为网络输入进行尝试,希望找出多通道颜色空间的组合方式。

本文首先通过实验比较不同颜色空间对小样本分类效果的影响,选用的颜色空间包括HSV、Lab、YCrCb、YUV、XYZ、HLS(Hue, Luminance, Saturation)、H-CrCb。通过对原始RGB图像进行公式转换得到不同的颜色空间。以HSV颜色空间为例,HSV空间的三种通道的划分十分直观,其源自颜色的特性,因此其也被称呼为六角锥体模型。HSV颜色空间的三个通道分别代表色调(H,范围为 $0^{\circ}\sim 360^{\circ}$)、饱和度(S,范围为 $0\sim 255$)和亮度(V,范围为 $0\sim 255$)。HSV图像通过将R、G、B采用不同的组合方式转换而成,转换公式如式(1)~(6)所示:

$$\begin{cases} R' = R/255 \\ G' = G/255 \\ B' = B/255 \end{cases} \quad (1)$$

$$\begin{cases} Max = \max(R', G', B') \\ Min = \min(R', G', B') \end{cases} \quad (2)$$

$$\Delta = Max - Min \quad (3)$$

$$H = \begin{cases} 0^{\circ}, & \Delta = 0 \\ 60^{\circ} \times [(G' - B')/\Delta] + 0^{\circ}, & Max = R \text{ 且 } G \geq B \\ 60^{\circ} \times [(G' - B')/\Delta] + 360^{\circ}, & Max = R \text{ 且 } G < B \\ 60^{\circ} \times [(B' - R')/\Delta] + 120^{\circ}, & Max = G \\ 60^{\circ} \times [(R' - G')/\Delta] + 240^{\circ}, & Max = B \end{cases} \quad (4)$$

$$S = \begin{cases} 0, & Max = 0 \\ \Delta/Max = 1 - Min/Max, & Max \neq 0 \end{cases} \quad (5)$$

$$V = Max \quad (6)$$

经过公式转换,不同颜色空间的通道信息都可由RGB颜色空间得到。通过公式转换得到的通道信息之间有着一定的关联,但也都有着其独有的特征信息,所以本文通过结合不同颜色空间来尝试对中药材粉末显微特征图像分类问题的适用性,为网络提供最合适用于小样本中药材数据分类识别的多通道颜色空间。多通道颜色空间可以是两种颜色空间的组合,甚至是多种颜色空间的组合,但是其他颜色空间信息的生成需要经过RGB颜色空间的转化,需要大量时间进行处理,并且极大占用计算机内存。为节省网络初始化时图像预处理时间以及减轻电脑配置需要,本文将采用6通道的颜色空间作为多通道颜色空间的输入,在保证扩充小样本数据通道信息的同时,降低图像预处理的时间,最终实现通过六通道颜色空间的方式为中药材图像提供更多重要的特征信息,提升网络对小样本数据的准确率。

2 改进的注意力机制模型

CNN基于其丰富的表现能力,极大地促进了图像识别领域在深度学习方向的发展。近些年来针对CNN性能的提升,大多侧重于深度、宽度和基数此三个影响因素。从ResNet18到ResNet152网络证明通过跳跃连接和堆叠排列的残差卷积块可以做到分类精度的提升;GoogLeNet的提出证明了提升模型性能也可以通过增加网络宽度的方式;在之后Xception^[17]和ResNeXt^[18]的提出则表明增加基数也可以达到提升网络学习性能的效果,并且性能的提高还要优于深度和宽度二者;自

从2017年ImageNet竞赛冠军Hu等^[19]提出了基于通道间权重的分配来实现通道间重要性选择的算法SEnet(Squeeze and Excitation Networks),从而引起了对于CNN在通道注意力上的关注;在2018年,Woo等^[20]就以此为理论基础,提出了在通道和空间的注意力机制模型CBAM(Convolutional Block Attention Module),此研究表明不仅是在通道间,在空间中也有着相应的内在联系,因此其设计的CBAM结构在兼顾通道和空间权重选择后,网络的分类效果又得到进一步的提升。本文存在着样本量稀少的问题,这导致网络学习的Epoch不能过多,否则会出现过拟合的情况,因此如何在有限的Epoch中使得网络能够更快降低损失函数(Loss Function, loss)是所需要研究的问题。本文通过注意力机制模型与ResNet-v2-101网络结合,对不同信息权重进行分配,使得网络能够将更多的注意力放到关键的通道和空间中,实现小样本数据分类效果的提升。

本文发现在通道和空间上的注意力模型并未进一步考虑其在深度、宽度、基数或内在联系这四方面的展开。经过进一步的研究表明可以分别在通道和空间注意力机制上结合内在联系和宽度这两个重要因素,以达到让网络着重学习关键信息的目的。

本文所提出的注意力机制模型位于ResNet-v2-101网络中每个残差模块之后,通过依次对残差模块结果采取通道和空间上的权重分配,从而实现对小样本关键特征的着重学习,如图2(b)所示。 $Input \in \mathbb{R}^{H \times W \times C}$ 为输入层,其在经过残差模块后得到结果 $RS \in \mathbb{R}^{H \times W \times C}$ 。通过通道注意力模块可以得到所有通道权重 $CA \in \mathbb{R}^{1 \times 1 \times C}$,残差模块结果 RS 与通道权重相乘得到结果 $G \in \mathbb{R}^{H \times W \times C}$ 。 G 通过空间注意力模块得到空间权重 $SA \in \mathbb{R}^{1 \times 1 \times C}$,再与 SA 相乘得到新的残差模块 $K \in \mathbb{R}^{H \times W \times C}$ 。将 K 与 $Input$ 相加得到最后的输出结果 $Output \in \mathbb{R}^{H \times W \times C}$ 。注意力模型的计算过程如式(7)所示:

$$Output = (CA \cdot RS) \cdot SA + Input \quad (7)$$

2.1 通道注意力模型

SEnet结构分别采取全局平均池化和全局最大池化操作用于注意力机制模型,相较于原始训练结果都有一定的提升,但是CBAM所提出的通道注意力模型的训练结果并未优于其两者。本文结果表明,CBAM所采取将全局平均池化和全局最大池化简单相加的方式并未充分考虑其二者的内在联系,使得其并未优于最好结果。

因此为了更好地将全局平均池化和全局最大池化内在关系联系,本文设计了一种新的通道注意力模型——串联注意力模型(Tandem Attention Model, TAM),通过全连接层将二者更加紧密地联系在一起,以更好提升关键特征的权重分配。TAM结构如图4所示。

TAM分别通过对 $H \times W$ 尺寸图像采取全局平均池化和全局最大池化操作,实现搜集 $Input$ 的通道信息 $\mathbb{R}^{1 \times 1 \times C}$,其中 $AvgPool$ 和 $MaxPool$ 分别代表全局平均池化和全局最大池化的操作符。两类通道信息都分别经过全连接层得到输出 $\mathbb{R}^{1 \times 1 \times C/d}$,其中 Mc 为第一层全连接层操作符。 d 则代表压缩比例,此操作原因在于批尺寸(Batch Size, BS)设置过大,如果不经过压缩会导致全连接层因参数量过大而产生显存溢出问题,因此通过设置压缩比减少网络参数量,同时也可以提升网络的运算效率。最后经过一层全连接层,其目的在于将二者

的信息充分融合,并且将输出恢复为原始尺寸 $\mathbb{R}^{1 \times 1 \times C}$,从而得到通道权重,其中 Mc' 代表第二层全连接层操作符。本文所提出的TAM通过将全局平均池化和最大池化串联,使得全连接层学习其两者间内在联系,实现通过卷积神经网络去深层次调整通道上权重配比的目的,从而达到更好的识别效果。TAM计算过程如式(8)所示:

$$CA = \sigma(Mc'(w(Mc(AvgPool(Input))) \oplus w(Mc(MaxPool(Input)))))) \quad (8)$$

式(2)中: CA 表示通道注意力权重, σ 代表Sigmoid激活函数, w 代表ReLU激活函数, \oplus 代表一维向量的串联。

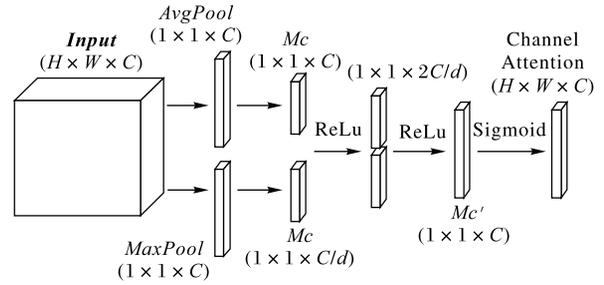


图4 串联注意力模型

Fig. 4 Tandem attention model

2.2 空间注意力模型

CBAM网络对于空间注意力模型构思的提出是对通道注意力机制的补充,将其思路从通道上权重的分配延续到空间上,分别将输入层从纵向对特征图进行全局平均池化和全局最大池化操作,得到输出 $\mathbb{R}^{H \times W \times 1}$ 。将两者并联得到 $\mathbb{R}^{H \times W \times 2}$,最后通过 7×7 卷积得到空间注意力权重 $\mathbb{R}^{H \times W \times 1}$ 。但是经过导管图像样本进行实验论证,其效果远不及原始训练效果。本文认为可能与平均池化和最大池化并联有关,因此对CBAM空间注意力模型分别验证两者的可行性。

经实验发现,单独的全局平均和最大池化相较于CBAM的空间注意力模型都有一定提升,认为其可能并非适用于二者的联立,使得二者互相干扰导致影响实验效果,因此本文选择效果最好的全局平均池化作为接下来研究方向。正如前文所述,不论是SEnet还是CBAM的空间注意力模型都只关注了自身的注意力机制,并未结合深度、宽度和基数这三个重要因素,因此在分析空间注意力模型结构后认为其可以与宽度这一因素相结合能够更加合理地实现在空间上的权重分配。

经过Inception网络多尺度融合的启发,本文认为可以对空间注意力模型增加其视野范围,从而更充分地学习空间之间的结构信息。但是由于 7×7 卷积参数量过于庞大,因此在并联过程中会大大降低运算速度。本文从文献[21]中获得启发,在运用多卷积核的同时采用空洞卷积,通过对 3×3 卷积设置不同的膨胀系数,使其扩增为 5×5 和 7×7 等尺寸的空洞卷积核,使得网络可以学习到不同尺寸感受野的特征信息,以更好权衡不同大小特征的权重;并且采用空洞卷积可以使参数量大幅度减少,在使用多卷积核的同时保证计算时间减少。本文尝试了多种多卷积核的组合架构,通过对 3×3 、 5×5 和 7×7 尺寸空洞卷积核采取不同的组合形式,进行实验对比。经过大量实验反复论证得出新的空间注意力机制模型:多卷积注意力模型(Multi-convolutional Block Attention Module, MBAM),其通过将这三种空洞卷积核并联的方式组合,可以更有效地提升小样本的分类效果;并且此三种空洞卷积核参

数量之和只有 CBAM 空间注意力模型中 7×7 尺寸卷积核参数量的一半,从而使计算时间大幅减少。MBAM 结构如图 5 所示。

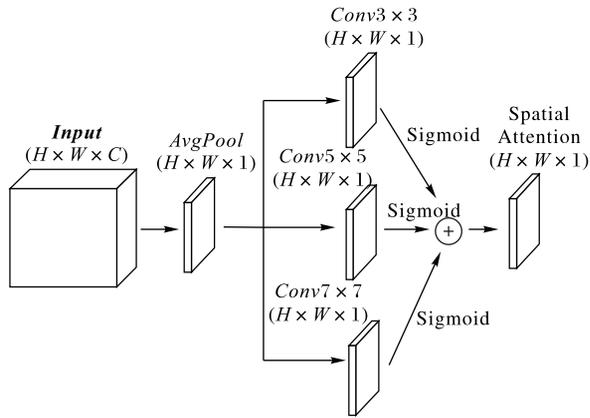


图 5 多卷积注意力模型

Fig. 5 Multi-convolutional block attention model

MBAM 输入为 $Input \in \mathbb{R}^{H \times W \times C}$,对特征图在通道上采取全局平均池化操作,得到空间信息 $\mathbb{R}^{H \times W \times 1}$,AvgPool 代表全局平均池化操作符。空间信息分别经过 3 个 3×3 尺寸的空洞卷积,其中设置 3 个卷积核的扩张系数(dilations)分别为 1、2、3,三种卷积核操作符为 U_1, U_2, U_3 ,其输出结果为 W_1, W_2, W_3 。最后通过 3 个卷积核的并联得到空间上的权重。MBAM 计算过程如式(9)、(10)所示。

$$W_i = U_i(AvgPool(Input)); i = 1, 2, 3 \quad (9)$$

$$SA = \sigma(W_1) + \sigma(W_2) + \sigma(W_3) \quad (10)$$

式(10)中:SA 为空间注意力权重,σ 代表 Sigmoid 激活函数。

3 实验结果与分析

针对多通道颜色空间和改进的注意力机制模型,本文分别通过实验验证二者的有效性,并在最后将两者结合得到最终的实验结果。

3.1 网络结构

本文采用 TensorFlow 架构来实现 ResNet-v2-101 网络模型的搭建,共有 101 层结构,BS 设置为 64,学习率以 0.1 为起始,采用指数衰减的方式在训练过程中降低学习率,其中衰减指数为 0.98。实验设备配置为 Xeon Gold 5120 CPU、64 GB 内存、GV100 GPU。中药材粉末显微特征样本一共 8 774 张图片,每张图像大小为 224×224,其中包括原始训练集 7 986 张(81%),验证集 798 张(10%),测试集 788 张(9%)。原始训练集在经图像预处理(随机裁剪和多角度旋转)扩充后,训练集数量达到 21 564 张。

本文所采用的 ResNet-v2-101 网络结构如表 1 所示。实验中所有提到的注意力机制模型都依据图 2(b)所示,添加到对应每个卷积层的 CA 和 SA 中。

3.2 多通道颜色空间

通过比较不同颜色空间对中药材粉末显微特征图像分类准确率的影响。经对比可知,XYZ、HCrCb 颜色空间的准确率低于 RGB 空间,因此将不作为重点考虑;HSV、YCrCb 和 YUV 空间在 9 种颜色空间中分类效果相对较好,认为此三个空间中所包含的中药材显微特征信息对网络的学习提供重要作用。实验结果如表 2 所示。

表 1 ResNet-v2-101 网络结构

Tab. 1 ResNet-v2-101 network structure

网络层名	输出尺寸	操作, 通道
Input	224×224	None, C
Conv1	112×112	7×7, 64 3×3 MaxPool, 64
Conv2	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \\ CA, 256 \\ SA, 256 \end{bmatrix} \times 3$
Conv3	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \\ CA, 512 \\ SA, 512 \end{bmatrix} \times 4$
Conv4	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \\ CA, 1024 \\ SA, 1024 \end{bmatrix} \times 23$
Conv5	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \\ CA, 2048 \\ SA, 2048 \end{bmatrix} \times 3$
Output	1×1	AvgPool, 2048 34-d Fully Connected, 34 Softmax, 34

表 2 三通道颜色空间准确率

Tab. 2 Three-channel color space accuracy

单位:%

颜色空间	准确率	颜色空间	准确率
RGB	89.8	YUV	91.4
HSV	91.4	XYZ	85.3
YCrCb	91.2	HLS	91.1
Lab	90.7	HCrCb	80.3

通过将 RGB 颜色空间与其他颜色空间并联的方式将图像输入分类网络,对实验结果进行比较。经实验证明,RGB-HSV 的 6 通道颜色空间作为网络输入可以得到最优的分类识别效果,其准确率由 RGB 颜色空间的 89.8% 提高到 91.6%。实验结果如表 3 所示。根据颜色空间的原理可以得知,与其他颜色空间相比 HSV 颜色空间的亮度、色调和饱和度是分离开的三种变量,能够更加直观地表达色彩的变化。相较于 RGB 颜色空间通过三原色混合得到的不同颜色,HSV 颜色空间可以更好地表达亮度、色调和色度上的变化差异,因此其作为 RGB 颜色空间上对通道信息的补充,RGB-HSV 的六通道颜色空间可以为中药材图像提供更多重要的特征信息,提升网络对小样本数据的准确率。

3.3 注意力机制模型

3.3.1 通道注意力模型

TAM 是通过将全局平均池化和全局最大池化串联,再经过全连接层得到的通道权重分配,在结构中存在着压缩比 d 这一变量,经本文实验论证表明,压缩比 d 参数设置为 4 对小样本数据的分类识别效果最佳。实验结果如表 4 所示。

表 3 六通道颜色空间准确率
Tab. 3 Six-channel color space accuracy

颜色空间	准确率	颜色空间	准确率
RGB-HSV	91.6	RGB-XYZ	90.5
RGB-YCrCb	91.1	RGB-HLS	91.0
RGB-Lab	90.9	RGB-HCrCb	90.9
RGB-YUV	89.7		

表 4 RGB-HSV+TAM 不同压缩比的准确率对比
Tab. 4 Accuracy comparison of different compression ratios of RGB-HSV+TAM

压缩比 d	准确率	压缩比 d	准确率
2	92.9	8	93.0
4	93.8		

其次, 本文将对比实验 2017 年 ImageNet 竞赛冠军 SEnet 结构和 2018 年 ECCV 的 CBAM 通道注意力模型来验证 TAM 的有效性, 不同通道注意力模型结构如图 6 所示, 其中 SEnet 结构的 Pooling 可采用 AvgPool 或 MaxPool 的方式进行全局池化操作。经实验可得出结论, 本文所设计的 TAM 结构可以更好地结合全局平均池化和全局最大池化的优点, 实现了在通道注意力上小样本准确率的提升。实验结果如表 5 所示。

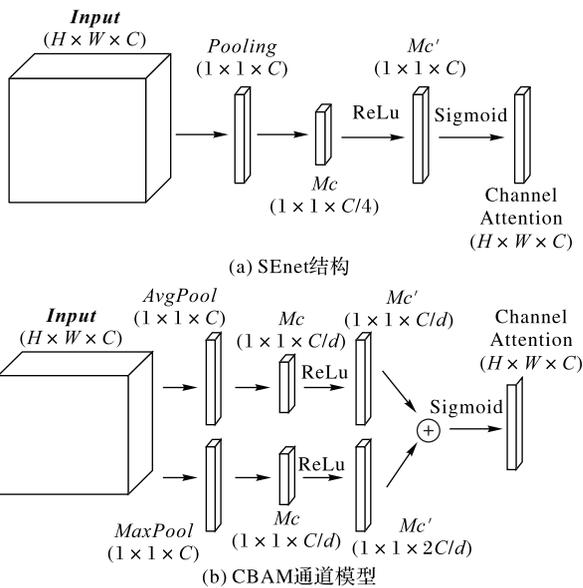


图 6 不同通道注意力模型
Fig. 6 Different channel attention models

表 5 通道注意力模型准确率比较
Tab. 5 Accuracy comparison of channel attention models

方法	准确率
RGB-HSV	91.6
RGB-HSV+SEnet(AvgPool)	91.8
RGB-HSV+SEnet(MaxPool)	93.3
RGB-HSV+CBAM 通道模型	92.3
RGB-HSV+TAM	93.8

3.3.2 空间注意力模型

为验证所提出空间注意力模型的有效性, 本文将对不同的全局平均池化空间注意力模型结构进行实验对比。首先本文采用单个空洞卷积的方式, 将 3×3 卷积核的 dilations 设置为 1、2、3、4, 膨胀后卷积核变为 3×3 、 5×5 、 7×7 和 9×9 , 比较单个空

洞卷积核的识别效果, 单空洞卷积核的空间注意力模型如图 7 所示, 其中 Conv 为空洞卷积所在位置。经实验对比后发现 3×3 、 5×5 和 7×7 卷积核的效果最好。实验结果如表 6 所示。

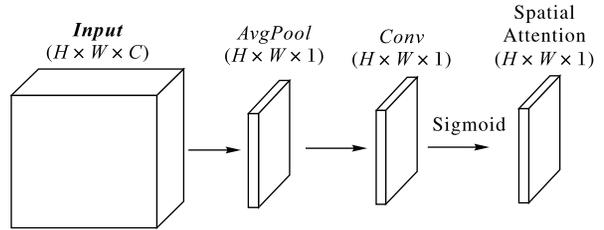


图 7 单空洞卷积核的空间注意力模型
Fig. 7 Spatial attention model of single-cavity convolution kernel

表 6 不同膨胀系数的准确率比较

卷积核	扩张系数	准确率/%
3×3	1	91.8
5×5	2	91.8
7×7	3	91.8
9×9	4	90.1

经过以上实验对比, 多卷积核的结构将依据 3×3 、 5×5 和 7×7 这三种卷积核进行设计, 通过对空洞卷积核采取两路并联和三路并联的组合方式来实现实验对比。其中两路并联的结构如图 8 所示, 通过对 Conv1 和 Conv2 采用不同的空洞卷积核进行组合。实验结果如表 7 所示(表 7 中“+”代表卷积核的并联), 从而得到最优的结果为 MBAM 结构, 通过空间上权重分配实现对小样本数据分类准确率的提升。

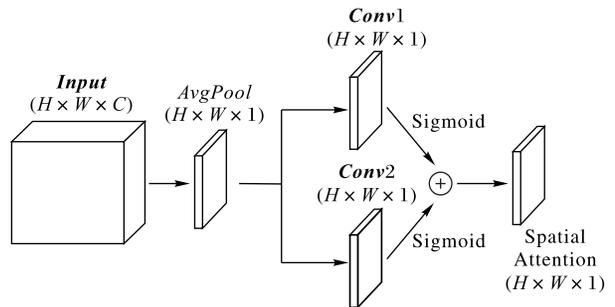


图 8 两路并联的空间注意力模型
Fig. 8 Two-way parallel spatial attention model

表 7 多卷积核准确率比较
Tab. 7 Accuracy comparison of multiple convolution kernels

卷积核	准确率	卷积核	准确率
$3 \times 3 + 5 \times 5$	91.5	$3 \times 3 + 7 \times 7$	92.0
$5 \times 5 + 7 \times 7$	91.4	$3 \times 3 + 5 \times 5 + 7 \times 7$	92.6

最终本文所提出 MBAM 结构与 2018 年 ECCV 所提出的 CBAM 空间注意力模型以及分别采用全局平均池化和全局最大池化的空间注意力模型进行比较, 不同的空间注意力机制模型如图 9 所示, 其中 Pooling 可采用 AvgPool 或 MaxPool 的方式进行全局池化操作。通过实验可以得出结论, 本文所提出的 MBAM 结构具有更好的空间权重分配机制, 能够更好地达到小样本分类准确率提升的目的。实验结果如表 8 所示。

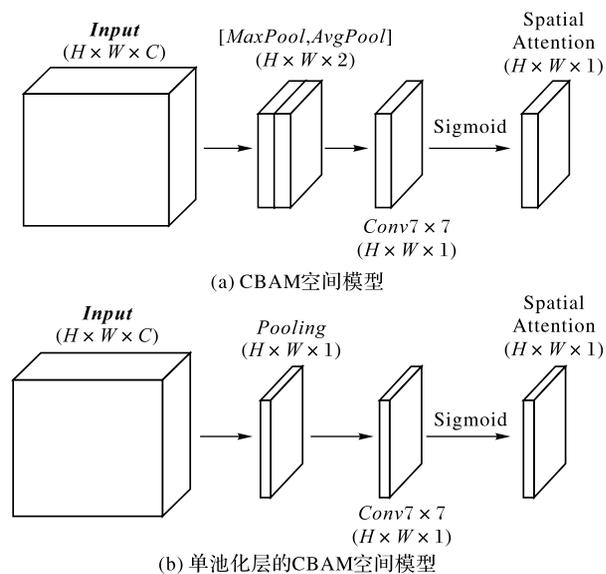


图 9 不同的空间注意力模型

Fig. 9 Different spatial attention models

表 8 空间注意力模型识别结果比较 单位: %

Tab. 8 Comparison of spatial attention model identification results unit: %

方法	准确率
RGB-HSV	91.6
RGB-HSV + CBAM 空间模型	68.5
RGB-HSV + CBAM (AvgPool)	88.8
RGB-HSV + CBAM (MaxPool)	79.8
RGB-HSV + MBAM	92.6

3.3.3 通道和空间注意力模型

本文将通道和空间注意力模型相结合,验证当两者共同使用时对小样本数据的分类识别效果。实验结果如表 9 所示。经实验对比可知,本文所设计多通道颜色空间和注意力机制模型不仅在通道和空间上具有很好的权重分配效果,且当二者相结合后,网络对小样本数据特征信息的权重分配效果也有一定的提升。

表 9 通道和空间注意力机制模型识别结果 单位: %

Tab. 9 Identification results of channel and spatial attention mechanism models unit: %

方法	准确率
RGB-HSV	91.6
RGB-HSV+TAM	93.8
RGB-HSV+MBAM	92.6
RGB-HSV+TAM+MBAM	93.9

3.4 多通道颜色空间和注意力机制模型

最后,为证明多通道颜色空间和注意力机制模型对小样本分类网络的有效性,将通过实验分别论证;并将两者结合验证二者在共同作用下,对中药材粉末显微特征图像准确率的提升。通过对比四组实验的训练损失折线图,可以观察到本文所提出的多通道颜色空间与注意力机制的组合可以使得训练 loss 更快收敛,达到对小样本数据在有限的 Epoch 中更快收敛的目的。训练损失折线如图 10 所示。实验结果如表 10 所示。

综上,针对小样本难以分类的问题,本文所提出对多通道颜色空间与注意力机制模型的网络改进方法,准确率分别提升 1.8 个百分点和 3.1 个百分点,将二者结合后准确率提升 4.1 个百分点;且训练 loss 值由 7.704×10^{-4} 下降到 4.732×10^{-4} , 下降 2.972×10^{-4} 。

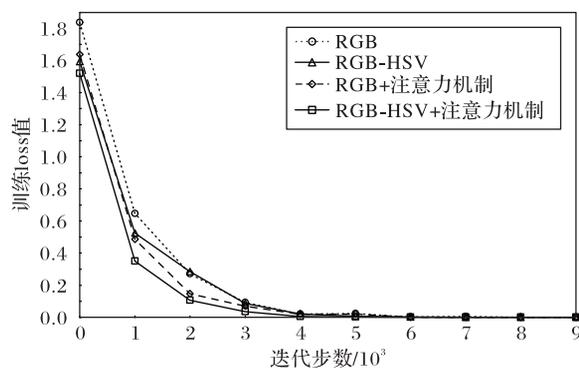


图 10 训练损失折线

Fig. 10 Line chart of training loss

表 10 多通道颜色空间和注意力机制模型识别结果

Tab. 10 Identification results of multi-channel color space and attention mechanism models

方法	测试准确率/%	训练 loss 值(1E-4)
RGB	89.8	7.704
RGB-HSV	91.6	7.641
RGB+TAM+MBAM	92.9	5.809
RGB-HSV+TAM+MBAM	93.9	4.732

4 结语

针对中药材粉末显微特征图像类别多、样本数量稀少、难以分类等问题,本文提出了一种基于小样本的网络改进方法——多通道颜色空间和注意力机制模型。实验结果表明,多通道颜色空间通过向网络中加入 HSV 颜色空间来实现针对中药材粉末显微特征相关信息的补充,弥补了小样本数据量不足的问题;注意力机制模型则通过深化通道模型的内在联系和空间模型的多尺度思想,改进 CBAM 结构,实现了 CBAM 计算效率的提升,并且更好地分配网络中通道和空间注意力的权重,使网络通过自主学习的方式,将注意力聚焦于关键的特征信息。将本文所提出的两种结构融合后,准确率得到了显著的提升,证明了两种方法对小样本分类识别有着大幅度的提升。

本文提出的方法是否可以应用于其他一些难以获取的小样本(如安全领域或医疗领域数据集)的分类识别还有待考证。并且不同通道特征信息的重要程度不一,是否能够实现将不同颜色空间的单通道进行拼接组合,使得组合后的三通道颜色空间得到六通道颜色空间相同的识别效果,从而实现减少多通道颜色空间的图像预处理时间将是接下来的研究方向。

参考文献 (References)

[1] 徐国均,徐璐珊,刘柏英,等. 中药材粉末显微鉴定[M]. 北京:人民卫生出版社,1986:14-608. (XU G J, XU L S, LIU B Y, et al. Microscopic identification of Powdered Chinese Medicinal Materials [M]. Beijing: People's Medical Publishing House, 1986: 14-608.)

- [2] 中华人民共和国卫生部药典委员会. 中华人民共和国药典[M]. 北京:人民卫生出版社, 1977: 4-308. (Chinese Pharmacopoeia Commission. Chinese Pharmacopoeia[M]. Beijing: People's Medical Publishing House, 1977: 4-308.)
- [3] 中华人民共和国卫生部药典委员会. 中华人民共和国药典中药粉末显微鉴别彩色图集[M]. 广州:广东科技出版社, 1999: 3-374. (Chinese Pharmacopoeia Commission. Colored Collection of Microscopic Identification of Chinese Material Medica Powder from Chinese Pharmacopoeia [M]. Guangzhou: Guangdong Science and Technology Press, 1999: 3-374.)
- [4] 赵中振. 中药显微鉴别图鉴[M]. 沈阳:辽宁科学技术出版社, 2005: 1-323. (ZHAO Z Z. An Illustrated Microscopic Identification of Chinese Material Medica[M]. Shenyang: Liaoning Science and Technology Publishing House, 2005: 1-323.)
- [5] 赵中振,陈虎彪. 中药显微鉴定图典[M]. 福州:福建科学技术出版社, 2016: 17-335. (ZHAO Z Z, CHEN H B. Chinese Medicinal Microscopic Identification [M]. Fuzhou: Fujian Science and Technology Publishing House, 2016: 17-335.)
- [6] 王一丁,石铎,李耀利,等. 基于SqueezeNet深度网络的中药材粉末显微特征图像识别研究[J]. 电子显微学报, 2019, 38(2): 130-138. (WANG Y D, SHI D, LI Y L, et al. Studies on identification of microscopic images of Chinese medicinal materials powder based on SqueezeNet deep network[J]. Journal of Chinese Electron Microscopy Society, 2019, 38(2): 130-138.)
- [7] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [8] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2019-05-20]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [10] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1-9.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [12] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9908. Cham: Springer, 2016: 630-645.
- [13] 韩宏,王鹏,唐振民,等. 彩色图像中复杂背景的多人脸检测[J]. 南京理工大学学报(自然科学版), 2001, 25(6): 597-601. (HAN H, WANG P, TANG Z M, et al. Multiple human face detection in color image with complex background [J]. Journal of Nanjing University of Science and Technology (Natural Science Edition), 2001, 25(6): 597-601.)
- [14] 杨红玲,宣士斌,莫愿斌,等. 基于YCbCr颜色空间手势分割[J]. 广西民族大学学报(自然科学版), 2017, 23(3): 61-66. (YANG H L, XUAN S B, MO Y B, et al. Based on the gesture segmentation YCbCr color space[J]. Journal of Guangxi University for Nationalities (Natural Science Edition), 2017, 23(3): 61-66.)
- [15] 陈昌红,刘彬,张浩. 基于多通道和卷积神经网络的极光分类[J]. 计算机技术与发展, 2018, 28(12): 200-204. (CHNE C H, LIU B, ZHANG H. Aurora images classification based on multi-channel fusion and convolutional neural network [J]. Computer Technology and Development, 2018, 28(12): 200-204.)
- [16] 江凤兵. 基于RGB-H-CbCr新颜色空间的肤色检测算法研究[J]. 科技广场, 2011(9): 34-38. (JIANG F B. The study of skin detection algorithms based on the RGB-H-CbCr new color space[J]. Science Mosaic, 2011(9): 34-38.)
- [17] CHOLLET F. Xception: deep Learning with depthwise separable convolutions [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1800-1807.
- [18] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017:5987-5995.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11211. Cham: Springer, 2018: 3-19.
- [21] LI X, WANG W HU X, et al. Selective kernel networks [EB/OL]. [2019-08-20]. <https://arxiv.org/pdf/1903.06586.pdf>.

This work is partially supported by the Specific Fund of Traditional Chinese Medicine Industry (201407003), the Key Project of Increase and Decrease of Expenditure at the Central Government Level (2060302).

WANG Yiding, born in 1967, Ph. D., professor. His research interests include biometrics recognition, machine vision.

HAO Chenyu, born in 1996, M. S. candidate. His research interests include computer vision, deep learning.

LI Yaoli, born in 1977, Ph. D., lecturer. Her research interests include identification and quality control of traditional Chinese medicines.

CAI Shaoqing, born in 1960, Ph. D., professor. His research interests include identification and quality evaluation of crude drugs.

YUAN Yuan, born in 1978, Ph. D., research fellow. Her research interests include identification of traditional Chinese medicines, molecular pharmacognosy.