

本文引用格式: 周苏娟,李嘉涛,何啟森,等.基于通道注意力机制的中药饮片图像识别方法[J].自动化与信息工程,2023,44(5):63-67;73.

ZHOU Sujuan, LI Jiatao, HE Qisen, et al. Image recognition method for Chinese herbal pieces based on channel attention mechanism[J]. Automation & Information Engineering, 2023,44(5):63-67;73.

## 基于通道注意力机制的中药饮片图像识别方法\*

周苏娟<sup>1,2</sup> 李嘉涛<sup>2</sup> 何啟森<sup>2</sup> 孟江<sup>3</sup> 刘波<sup>1</sup>

(1.广东工业大学自动化学院, 广东 广州 510006

2.广东药科大学医药信息工程学院, 广东 广州 510006

3.广东药科大学中药学院, 广东 广州 510006)

**摘要:** 针对中药饮片识别采用人工方式存在的耗费人力物力、主观性强, 容易造成偏差等问题, 提出基于通道注意力机制的中药饮片图像识别方法。首先, 构建中药饮片图像数据库; 然后, 采用改进的 AlexNet 模型对莪术、姜、橘核和牡丹皮的饮片图像进行识别; 最后, 针对同一饮片不同炮制品的图像特征差异不明显问题, 引入通道注意力机制。对比实验结果表明: 基于通道注意力机制的 AlexNet 模型比 AlexNet 模型的平均精确度提高了 2.18%, 识别准确率提高了 2.05%; 且降低了参数量及 FLOPs。

**关键词:** 中药饮片; 图像识别; 通道注意力机制; AlexNet 模型

中图分类号: TP391.4

文献标志码: A

文章编号: 1674-2605(2023)05-0010-06

DOI: 10.3969/j.issn.1674-2605.2023.05.010

## Image Recognition Method for Chinese Herbal Pieces Based on Channel Attention Mechanism

ZHOU Sujuan<sup>1,2</sup> LI Jiatao<sup>2</sup> HE Qisen<sup>2</sup> MENG Jiang<sup>3</sup> LIU Bo<sup>1</sup>

(1.School of Automation, Guangdong University of Technology, Guangzhou 510006, China

2.College of Medical Information Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China

3.College of Traditional Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China)

**Abstract:** A channel attention mechanism based on image recognition method for Chinese herbal pieces is proposed to address the issues of human and material resources consumption, strong subjectivity, and bias in manual recognition of Chinese herbal pieces. Firstly, construct an image database of Chinese herbal pieces; Then, an improved AlexNet model was used to recognize the sliced images of Zedoary Turmeric, Ginger, Orange Kernel, and Moutan Cortex; Finally, a channel attention mechanism is introduced to address the issue of insignificant differences in the characteristics of different processed products of the same slice. The comparative experimental results show that the AlexNet model based on channel attention mechanism has an accuracy improvement of 2.18% and a classification accuracy improvement of 2.05% compared to the AlexNet model; And it reduces the number of parameters and FLOPs.

**Keywords:** Chinese herbal pieces; mage recognition; channel attention mechanism; AlexNet model

### 0 引言

中医中药是我国医药行业的重要组成部分。在中药蓬勃发展的过程中, 中药饮片的识别技术占据了重要的地位。目前, 中药饮片的识别主要根据专家的知

识和经验或通过图像比对做出判断<sup>[1]</sup>, 不仅耗费大量的人力物力, 还具有主观性, 容易造成偏差。随着人工智能技术的发展, 利用机器学习来实现图像识别逐步应用于中医药领域<sup>[2-3]</sup>, 不仅提高了中药饮片识别的

\* 基金项目: 国家自然科学基金资助项目 (82173973, 81473352)

自动化程度及稳定性，还确保了统一的识别结果<sup>[4]</sup>，推动了中医药客观化、标准化的研究。

近年来，深度学习<sup>[5]</sup>在图像识别领域取得了较大的突破。王健庆等<sup>[6]</sup>利用 Google Net 深度学习模型对常见的 100 种中药饮片数据库进行训练及识别，达到了 92% 的平均识别率。胡继礼等<sup>[7]</sup>借助深度迁移学习理论，在 Inception-V3 模型<sup>[8]</sup>上执行自适应和 finetune 操作，对中药饮片图像进行分析处理，平均错误率接近人工组，但其最大值与最小值的极差优于人工组。然而，对于相同品种的中药饮片，如姜有干姜、炮姜、姜炭等不同炮制品，其图像特征的差异不明显，且性味、功效各不相同<sup>[9]</sup>，如采用以上方法进行识别，效果并不理想。

为此，本文提出基于通道注意力机制的中药饮片图像识别方法，通过引入基于注意力机制的 AlexNet 模型，提高对同一种中药不同炮制品的识别准确率。

## 1 深度学习

深度学习是机器学习的热门研究领域之一，相比传统的神经网络，它从输入层到输出层经历的层次数目更多<sup>[10]</sup>。近年来，深度学习在语音和图像识别方面取得了突飞猛进的进展。

### 1.1 卷积神经网络

卷积神经网络(convolutional neural network, CNN)是深度学习的代表算法之一。世界上第一个卷积神经网络是 LECUN<sup>[11]</sup>在 1998 年提出的 LeNet。Google Net<sup>[8]</sup>、ResNet<sup>[12]</sup>和 AlexNet<sup>[13]</sup>等模型都是在此基础上发展起来的。AlexNet 模型由 5 个卷积层和 3 个全连接层组成，采用 ReLU 作为激活函数，解决了深度网络中 Sigmoid 的梯度弥散问题；利用随机失活方法在训练过程中随机忽略一些神经元，避免了过拟合问题。

### 1.2 注意力机制

近年来，注意力机制在图像理解、语音识别等领域得到一定的应用，其可以辅助模型分析输入信息对当前任务的作用，通过评判不同输入信息的重要程度得到信息不同部分的权重，使任务模型更为准确<sup>[14]</sup>。注意力模型通常分为通道注意力模型、空间注意力模

型、通道和空间混合注意力模型。对于二维图像的 CNN 来说，一个维度对应图像的尺度空间，另一个维度对应通道。本文利用通道注意力机制建立中药饮片的图像识别模型。

## 2 基于通道注意力机制的中药饮片图像识别

目前，中药饮片图像识别领域暂无公开的标准数据库用于深度学习研究。本文采用的中药饮片图像均由广东药科大学中药数字化质量评价重点实验室及国家中医药管理局/广东省高校中药质量工程技术研究中心课题组提供。

### 2.1 数据集准备

中药饮片种类繁多，据最新的《中华本草》记载，中药共有 8 980 种，常见的中药饮片有一千余种。本文以 4 种中药饮片（莪术、姜、橘核、牡丹皮）和 2 种中药饮片不同程度的炮制品（干姜、姜炭、炮姜，牡丹皮轻炭、牡丹皮标炭、牡丹皮重炭）图像为研究对象，如图 1 和图 2 所示。

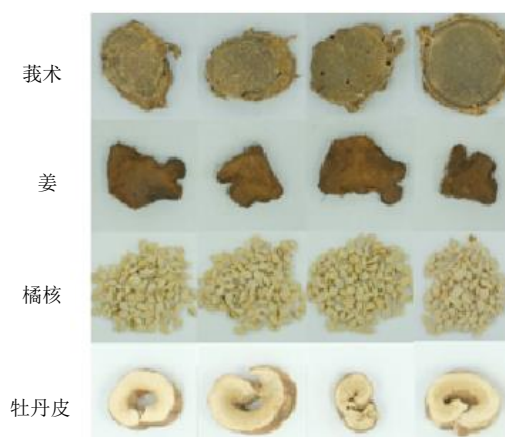


图 1 不同中药饮片图像

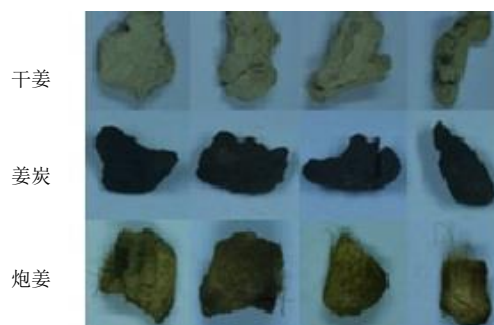




图2 不同中药饮片炮制品图像

## 2.2 数据预处理

原始数据集中的图像大小为  $4608 \times 3456 \times 3$  像素，利用 OpenCV 中的 `resize` 将其调整为  $224 \times 224 \times 3$  像素。采用随机旋转、平移、翻转、缩放等方式进行图像增强。增强后的数据集经过旋转、镜像、缩放等操作，使图像数量增倍，减少了模型过拟合的可能性。

## 2.3 数据集

经过预处理的中药饮片数据集含有 6400 幅图像，按 6:3:1 的比例划分训练集、验证集、测试集。其中，训练集有 3840 幅图像，验证集有 1920 幅图像，测试集有 640 幅图像。

## 2.4 模型构建

基于通道注意力机制的中药饮片图像识别方法以 AlexNet 模型为基础，通过改进其结构并增添新的机制，使其尾部结构的参数量与计算量大大减少。

首先，在 AlexNet 模型的第二、三、四、五次卷积后加入批量归一化操作，以加快网络收敛速度，为模型提供正则化效果，并在激活操作后加入注意力机制 (squeeze-and-excitation, SE) 模块。

然后，因为 AlexNet 模型的 3 个全连接层参数量多且计算量大 (参数量和每秒浮点运算次数 (floating point operations, FLOPs) 均达 56.68 MB)，所以删除 AlexNet 模型的 3 个全连接层，增添 1 个  $1 \times 1$  卷积层，利用 960 个  $1 \times 1 \times 256$  的卷积核得到  $5 \times 5 \times 960$  的特征映射；利用 Hardswish 激活函数做批量归一化处理。

最后，利用全局平均池化操作获取每个通道的全局信息，得到  $1 \times 1 \times 960$  的特征映射，将其展开为 960 个神经元，加入到有 1280 个神经元的全连接层，利用 Hardswish 激活函数，加入 Dropout 正则化，以减缓过拟合，输出层采用 softmax 回归。

改进后的 AlexNet 模型结构如图 3 所示。

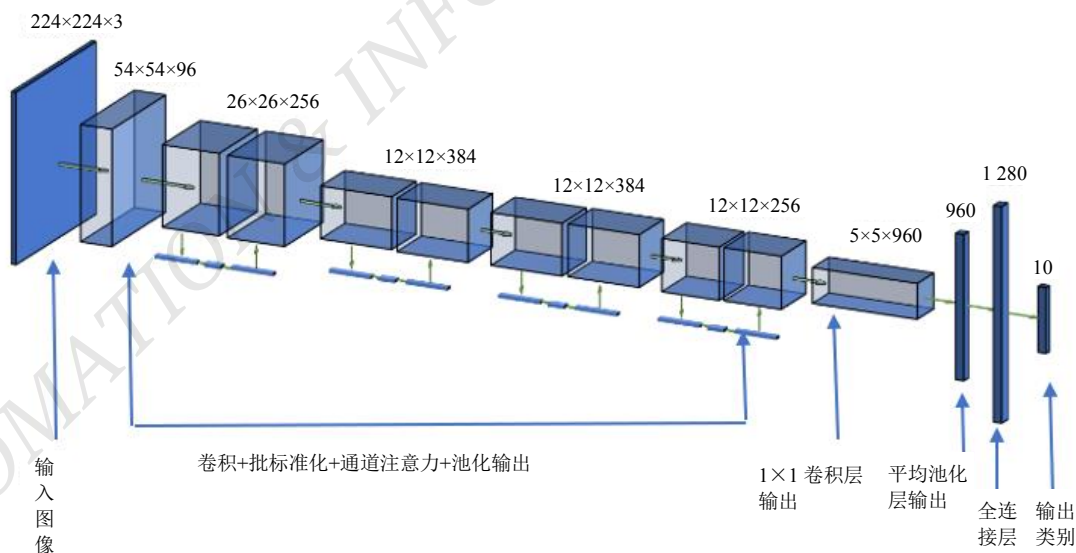


图3 改进后的 AlexNet 模型结构

借助轻量级的核心库 `fvcore` 求得 AlexNet 模型参数量从 56.68 MB 降至 2.76 MB，FLOPs 从 56.68 MB 降至 8.68 MB，参数量与计算量均大为减少。

## 3 实验

本实验采用 Pytorch 深度学习框架。硬件环境为：

处理器 Intel(R) Core(TM) i5-8300H CPU @ 2.30 GHz 四核; 内存 8.00 GB; 显卡 NVIDIA GeForce GTX 1050 Ti, 显存 4.00 GB。

为验证同一种中药饮片不同炮制品的识别效果, 将 AlexNet 模型与基于通道注意力机制的 AlexNet 模型进行对比实验, 识别结果如表 1 和表 2 所示。其中, Precision、Recall、Specificity、F1\_Score 分别为评估识别性能的精确度、召回率、特异度和 F1\_Score 指标。识别准确率 (Accuracy) 的计算公式为

$$A_{\text{accuracy}} = (TP + TN) / (TP + FP + TN + FN)$$

式中:  $TP$  为模型正确识别为正例的样本数,  $TN$  为模型正确识别为反例的样本数,  $FP$  为模型错误识别为正例的样本数,  $FN$  为模型错误识别为反例的样本数。

表 1 AlexNet 模型的识别结果

| 中药饮片  | $P_{\text{precision}}$ | $R_{\text{recall}}$ | $S_{\text{specificity}}$ | $F1$   |
|-------|------------------------|---------------------|--------------------------|--------|
| 莪术    | 98.92%                 | 99.73%              | 99.85%                   | 99.32% |
| 姜     | 98.93%                 | 95.88%              | 99.85%                   | 97.38% |
| 橘核    | 100.00%                | 99.71%              | 100.00%                  | 99.86% |
| 牡丹皮   | 99.15%                 | 99.15%              | 99.89%                   | 99.15% |
| 干姜    | 98.53%                 | 99.63%              | 99.85%                   | 99.08% |
| 姜炭    | 91.72%                 | 99.25%              | 99.13%                   | 95.34% |
| 炮姜    | 98.61%                 | 88.38%              | 99.89%                   | 93.22% |
| 牡丹皮标炭 | 90.57%                 | 94.12%              | 99.10%                   | 92.31% |
| 牡丹皮轻炭 | 95.11%                 | 94.40%              | 99.53%                   | 94.76% |
| 牡丹皮重炭 | 97.78%                 | 98.51%              | 99.78%                   | 98.14% |
| 平均值   | 96.93%                 | 96.88%              | 99.69%                   | 96.86% |

识别准确率: 97.16%

由表 1 可知, AlexNet 模型对中药饮片的识别准确率为 97.16%, 但同一种中药不同炮制品的识别效果欠佳,  $F1\_Score$  分数最低的是牡丹皮标炭, 仅为 92.31%。

表 2 基于通道注意力机制的 AlexNet 模型识别结果

| 中药饮片 | $P_{\text{precision}}$ | $R_{\text{recall}}$ | $S_{\text{specificity}}$ | $F1$    |
|------|------------------------|---------------------|--------------------------|---------|
| 莪术   | 99.73%                 | 100.00%             | 99.96%                   | 99.86%  |
| 姜    | 99.48%                 | 99.48%              | 99.92%                   | 99.48%  |
| 橘核   | 100.00%                | 100.00%             | 100.00%                  | 100.00% |
| 牡丹皮  | 99.72%                 | 99.72%              | 99.96%                   | 99.72%  |

续表

| 中药饮片  | $P_{\text{precision}}$ | $R_{\text{recall}}$ | $S_{\text{specificity}}$ | $F1$    |
|-------|------------------------|---------------------|--------------------------|---------|
| 干姜    | 100.00%                | 100.00%             | 100.00%                  | 100.00% |
| 姜炭    | 99.24%                 | 97.01%              | 99.93%                   | 98.11%  |
| 炮姜    | 96.76%                 | 99.17%              | 99.71%                   | 97.95%  |
| 牡丹皮标炭 | 96.92%                 | 98.82%              | 99.71%                   | 97.86%  |
| 牡丹皮轻炭 | 99.25%                 | 98.13%              | 99.93%                   | 98.68%  |
| 牡丹皮重炭 | 100.00%                | 98.88%              | 100.00%                  | 99.44%  |
| 平均值   | 99.11%                 | 99.12%              | 99.91%                   | 99.11%  |

识别准确率: 99.21%

由表 2 可知, 基于通道注意力机制的 AlexNet 模型比 AlexNet 模型的识别准确率提高了 2.05%, 炮姜与牡丹皮标炭的  $F1\_Score$  分别提高了 4.73%、5.55%, 说明炮姜和牡丹皮标炭这两种对应姜和牡丹皮的炮制品介于轻炭和重炭之间, 在外观上难以区分。

为了更加直观地显示模型的识别效果, AlexNet 模型和基于通道注意力机制的 AlexNet 模型的混淆矩阵分别如图 4 和图 5 所示。

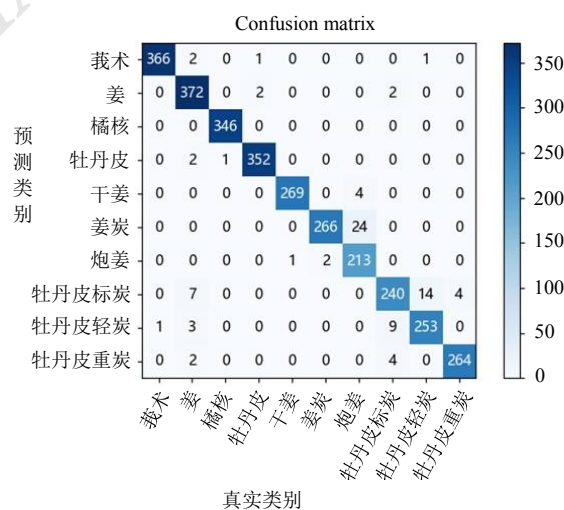


图 4 AlexNet 模型的混淆矩阵

由图 4 可知: 利用 AlexNet 模型, 240 个牡丹皮标炭被正确识别, 9 个牡丹皮标炭被错分为牡丹皮轻炭, 4 个牡丹皮标炭被错分为牡丹皮重炭; 14 个牡丹皮轻炭、4 个牡丹皮重炭和 2 个姜被错分为牡丹皮标炭; 在炮姜识别中, 24 个炮姜被错分为姜炭, 说明 AlexNet 模型难以识别不同炮制程度的牡丹皮或姜, 容易混淆同一种中药饮片的不同炮制品。

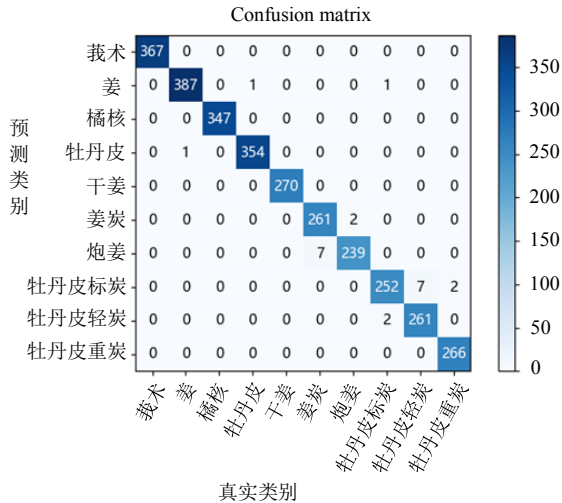


图5 基于通道注意力机制的 AlexNet 模型的混淆矩阵

由图5可知:利用基于通道注意力机制的 AlexNet 模型识别炮姜与姜炭、牡丹皮标炭与牡丹皮轻炭时,出错个数较少,炮姜被错分为姜炭的个数仅有 2 个,而利用 AlexNet 模型识别时,被错分的个数达 24 个;在牡丹皮不同炮制程度的识别上,被错分的个数也有一定程度的减少,说明经过基于通道注意力机制的 AlexNet 模型的识别效果整体优于 AlexNet 模型。

#### 4 结论

本文构建了莪术、姜、橘核和牡丹皮 4 种不同中药饮片的图像数据库及姜和牡丹皮不同炮制品的图像数据库,研究引入深度学习实现对不同中药饮片以及同一种中药饮片不同炮制品的自动识别。实验结果表明,采用 AlexNet 模型识别 4 种不同的中药饮片图像能够达到较好的效果,但对于同一中药饮片的不同炮制品,由于图像特征差异不明显,识别效果不理想。而通过基于通道注意力机制的 AlexNet 模型来识别,提高了识别准确率,同时参数量以及 FLOPs 均大幅降低。

本文的中药饮片识别只涉及到莪术、姜、橘核和牡丹皮 4 个类别,后期需要进一步扩充更多类别的中药饮片数据库,包括同一种饮片的不同炮制品。另外,在中药饮片质量评价标准方面,本文探讨的中药饮片图像识别只是涉及外观上的评价,后期可以结合饮片内在成分含量等多特征角度,进一步完善中药饮片的

综合评价。

#### 参考文献

- [1] 庄奕珊.基于深度神经网络的中药材识别[D].广州:华南理工大学,2018.
- [2] 王一丁,郝晨宇,李耀利,等.基于深度学习的小样本中药材粉末显微图像识别[J].计算机应用,2020,40(5):1301-1308.
- [3] 陈雁,邹立思.基于 BMFnet-WGAN 的中药饮片智能甄别[J].中国实验方剂学杂志,2021,27(15):107-114.
- [4] ZHOU S, MENG J, HUANG Z, et al. A method for discrimination of processed ginger based on image color feature and a support vector machine model[J]. Analytical Methods, 2016,8(10):2201-2206.
- [5] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.
- [6] 王健庆,戴恺,李子柔.基于深度学习的中药饮片图像识别研究[J].时珍国医国药,2020,31(12):2930-2933.
- [7] 胡继礼,王永康,阚红星.基于深度迁移学习的中药饮片识别研究[J].新乡学院学报,2019,36(3):62-67.
- [8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:1-9.
- [9] ZHOU S, MENG J, LIU B. Investigation into the pharmacokinetic-pharmacodynamic model of Zingiberis Rhizoma/Zingiberis Rhizoma Carbonisata and contribution to their therapeutic material basis using artificial neural networks[J]. RSC Advances, 2017,7(41):25488-25496.
- [10] 刘飞,张俊然,杨豪.基于深度学习的医学图像识别研究进展[J].中国生物医学工程学报,2018,37(1):86-94.
- [11] 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.
- [12] LI X, DING L, WANG L, et al. FPGA accelerates deep residual learning for image recognition[C]//2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2017: 837-840.
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017,60(6):84-90.
- [14] 李正涛.基于注意力机制的光谱地物分类方法研究[D].武汉:华中科技大学,2020.

(下转第 73 页)