

基于 Swin Transform 的亳州花茶成分识别

董鹏 盛魁 马健

亳州职业技术学院 信息工程系

DOI:10.12238/acair.v3i2.13500

[摘要] 近年来,深度学习技术在图像识别领域展现出强大的应用潜力,其中热门的Swin Transformer模块通过移动窗口来层级式的提取图像中特征信息,从而使得图像信息特征更加细致化。亳州花茶是一种以亳州地区特产花草、药料为原料制作而成的茶类保健饮品,其主要成分包括当地特有花朵和中草药搭配而成,但是由于花茶识别具有识别复杂,相似度高,特征难区分等痛点,不能更加细致的识别花茶成分。本文通过传统的卷积神经网络和Swin Transformer模块相结合提出一种新型的CNN-SwinTransformer架构来捕捉花茶数据的局部特征,揭示花茶识别过程中的内在结构,提高分类精准度,对亳州花茶成分进行准确识别,为其品质评估和生产过程的优化提供科学依据。

[关键词] 人工智能; 深度学习; Transform; 图像识别

中图分类号: TP18 **文献标识码:** A

Identification of Bozhou scented tea ingredients based on Swin Transform

Peng Dong Kui Sheng Jian Ma

Department of Information Engineering, Bozhou Vocational and Technical College

[Abstract] In recent years, deep learning technology has shown great application potential in the field of image recognition. Among them, the popular Swin Transformer module extracts feature information from images in a hierarchical manner by moving windows, making the image information features more detailed. Bozhou scented tea is a tea health drink made from the specialty flowers and medicinal materials in Bozhou. Its main ingredients include local unique flowers and Chinese herbal medicines. However, due to the pain points of complex recognition, high similarity, and difficult feature distinction in scented tea recognition, it is not possible to identify the scented tea ingredients more carefully. This paper proposes a new CNN-SwinTransformer architecture by combining traditional convolutional neural networks and Swin Transformer modules to capture the local features of scented tea data, reveal the internal structure in the scented tea recognition process, improve classification accuracy, accurately identify the ingredients of Bozhou scented tea, and provide a scientific basis for its quality evaluation and production process optimization.

[Key words] artificial intelligence; deep learning; Transform; image recognition

引言

亳州具有世界药都之称,近年来花茶逐渐因具有保健,养颜等功效,在社会中大受欢迎。花茶是一种或几种中草药成分而组成的新型饮品,但是具体成分的辨别较为复杂,传统的人工挑选和鉴别,较为费时费力,企业在加工期间,使用新型算法来减少成本,提高效率。

随着时代发展,深度学习技术在图像识别领域逐渐成为热门,通过深度学习所搭建的不同的模型,不仅减少了人工的时间成本,又提高了图像识别的准确率,该技术的应用,为企业进行新型的数字赋能,在自然语言处理中,Transform的发展作为主力,该模型因具有可以连接上下文语义等特点,所以大受欢迎。图像识别领域中也逐渐引入了这一前沿技术,由于花茶成分繁杂,

难以辨认,传统的Transform无法满足现对图片识别的要求,而Swin Transform模型利用不同滑动窗,在训练模型时,花茶成分中中草药形状大小各不相同,对数据进行不同尺寸的特征提取,从而把握更多细节特征,提高分类准确率。

1 发展现状

1.1 深度学习发展

随着数据量越来越大,以及算力能力的提升,作为机器学习的一个重要分支,深度学习已经被应用到各个领域,如计算机视觉、自然语言处理、生物信息学等方面。例潘1等人通过使用深度学习对其中医古籍进行自然语言处理研究,鲍2等人也探究了深度学习在基因组学中的进展

随着研究的深入,研究人员已针对特定的任务构建了众多

的深度学习的框架,如深度神经网络、卷积神经网络、循环神经网络等。这些学习框架的提出有力的推动深度学习的应用和发展,并已被应用到物体识别、产品推送、边缘检测等应用场景。最近,深度学习也被成功地应用到中草药识别中,并且推出了更多不同的变体模型,以用来解决针对性问题。例郭3等人通过搭建的深度学习的ERNIE模型对古代方剂进行智能推荐,相对于传统机器学习方法,深度学习能够从网络中学习更高层次的特征,从而获得了更好的学习性能,提高分类准确度。因此,基于深度学习的网络模型在中草药识别中展现了更大潜力,近期也有不少研究人员致力于对于中草药的识别,例杨4等人通过利用深度学习准确的完成中草药植物图像的识别任务。

1.2 Transformer模型发展

深度学习领域中,卷积神经网络等模型曾长期主导。然而,随着Transformer模型的提出,其在自然语言处理等任务上的出色表现引起了广泛关注,逐渐成为深度学习领域的新兴代表。Transformer是一种深度学习模型架构,主要用于自然语言处理任务,如机器翻译、自动翻译等问题。张5等人通过该模型对多语种同时进行翻译,取得不错效果。

随着时代的发展,该模型不仅在自然语言处理领域取得成功,在图像识别领域也逐渐崭露头角。该模型由于引入了自注意力机制来处理输入序列中任意位置的信息,从而在序列处理任务中取得了很好的效果。朱6等人通过基于图像分类对该模型进行优化研究,同时Vision Transformer模型将图像分割为图块,并利用Transformer进行处理,取得了令人瞩目的性能,挑战了传统的卷积神经网络。

1.3 Transformer模型改进

传统的Transformer本应用于自然语言处理领域,是为了研究长距离的相关信息是否具有互补性质,应用于图像中进行边缘特征的表达但是会导致信息大量丢失以及出现翻译不准确或关联性不强等问题。针对Transformer模型的一些不足,研究者提出了许多改进和变种模型,例如加入局部注意力机制、引入多头注意力机制、增加层级结构等,以提升模型的性能和效率。而近期出现的Swin Transformer模型,是针对图像处理任务提出的一种改进型Transformer模型。该模型引入了跨层窗口交互机制和局部注意力机制,显著提高了对大规模图像的建模能力,成为图像领域的研究热点。Swin Transformer应用广泛,例如张7等人通过利用Swin Transformer对其手写体汉字识别进行研究与应用,生活中张8等人利用Swin Transformer的天气图像进行准确预测,利用Swin Transformer模型进行图像训练,在花茶成分识别中,同样可以更好的提取花茶成分的潜在信息,从而使得分类准确度进一步提高。

2 方法

本文首先搭建卷积神经网络称为Conv层,进行初步特征提取,然后进一步利用Swin Transformer模块深层次提取特征,最后利用Softmax函数实现花茶成分的识别。

总体框架图如下所示:首先输入不同种类花茶成分的数据

集,然后通过卷积层(ConvBlock)进行特征提取,卷积块作为编码器和解码器的每一层,可以是花茶中的复杂成分,通过卷积层可以提取更多的细节特征,通过卷积模块把输入的图片进行卷积处理,对输入数据利用卷积块在窗口内对应的元素进行相乘再相加,得出的计算结果代替窗口中心的像素亮度值,以此类推,从左到右,从上到下,从而提取更多的图片特征信息,便于下一步的输入。

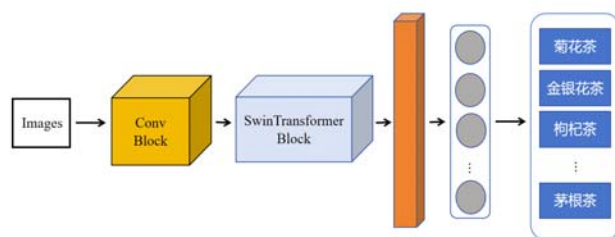


图1 总体框架图

在传统的Transformer模块中,首先对数据进行嵌入操作,将嵌入后结果位置输入全局注意力机制模块中,其中该机制计算如下表示,其中Q代表生成询问,K代表索引,V代表回答,d代表特征值信息,F代表标准化后的Softmax函数,A代表标准化权重,M代表多头注意力核心模块。

$$M = F \left(Q \times \sqrt{\frac{K}{d}} \right)$$

$$A = M \times V$$

初步提取特征后,为了弥补传统Transformer模块中基于全局注意力的信息丢失等问题的出现,接下来把卷积过后的图片特征输入到Swin Transformer模块中进一步对特征进行提取。Swin Transformer模块具体框架图如下所示,由于Swin Transformer是通过设置窗口来实现的局部注意力的提高,使得会带来更多图像的特征细节,探寻潜在的表征提取,同时也比传统Transformer模块全局注意力中计算量更小。

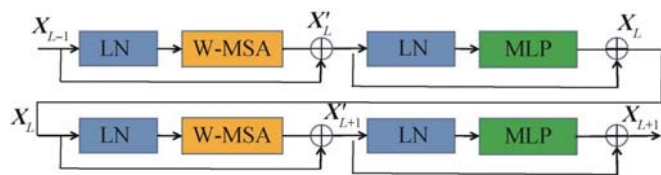


图2 Swin Transformer框架图

Swin Transformer具体结构是由归一化层,残差短连接以及多层感知机所构成,相比于全局注意力机制,利用不同窗口多层感知机来更好的进行不同花茶成分细节上的,提取不相邻图片的信息残差,其具体计算公式如下:首先Q代表query矩阵,K代表key矩阵,V代表value矩阵,B中的值代表偏置矩阵,F代表Softmax函数用于最后分类,S代表分类权重。

$$S = F \left(\frac{QK^T}{\sqrt{d}} + B \right) V$$

如图2所示, 具体来说, 利用移动不同窗口进行对特征进行分类权重, 更好的打乱特征进行交互, 探索潜在特征信息。其中, 对于该模型, 首先对经过卷积后数据进行嵌入, 后续将嵌入的结果位置信息输入到该网络中, 从而得到, 通过输入信息到注意力层中来获取多维度前后文信息, 计算公式如下:

$$X'_L = M[N(X_L)] + X_{L-1}$$

$$X'_L = M[N(X_L)] + X_L$$

总的来说, 引入利用Swin Transformer模块结构来进行寻找不同网络的间接联系以及寻找不同卷积下的相关性。大大提高了同类别花茶种类的关联性, 同时也更加细致的区分出不同花茶成分的显著差异性。

最后通过SoftMax函数作为激活函数, 将一个数值向量归一化为一个概率分布向量, 选取权重高的确定为最终分类目标, 从而得到正确的分类任务。

3 实验

本实验采集了不同花茶成分的中药图片作为训练数据, 其中包括了百合茶, 菊花茶, 大麦茶以及茅根茶等不同中药成分花茶, 通过把数据集输入到该模型中进行训练, 最后通过测试集测试所识别花茶成分的准确率。

在图片分类任务中, 除了准确度作为评价指标之一, 还利用不同的评价指标来作为衡量该模型性能, 以及利用不同的模型来进行对比。在评价指标中最主要的包括以下四个变量, 真正例(True Positives, TP), 真负例(TN(True Negatives, TN), 假负例(FN(False Negatives, FN) 以及假正例(False Positives, FP), 分别代表真正类, 真负类, 假负类以及假正类, 而本文中主要评价指标包括准确度, 精确率以及召回率, 其中准确率作为最为基础和原始的评价指标。

其中准确率的定义是预测正确的结果占总样本的百分比, 计算公式如下:

$$\text{准确率} = \frac{TP + TN}{TP + FN + FP + TN}$$

其中精准率含义是在所有被预测为正的样本中实际为正样本的概率, 其计算公式如下:

$$\text{精确率} = \frac{TP}{TP + FP}$$

其中召回率含义是在实际正样本中被预测为正样本的概率, 其计算公式如下:

$$\text{召回率} = \frac{TP}{TP + FN}$$

通过以上三个指标来测试该模型实际性能。

为了进一步测试该模型的实际效果, 本文采用了不同的对比实验来进行测试, 首先没有加入Swin Transformer模块, 只利用卷积神经网络来进行实验, 称为CNN模型对以上数据集进行测试; 利用通过引入全局注意力的传统Transformer模块来进行实验, 利用不同对比方法, 三种不同模型实验结果以及不同评价指

标如下表所示。

表1 不同模型分类性能表

模型	准确率(%)	精确率(%)	召回率(%)
CNN	85.6	72.3	86.3
Transformer	87.2	74.1	88.2
CNN+Swin Transformer	89.3	75.6	90.4

由上表可以清楚的看出, 本文所提出的通过加入Swin Transformer模块, 识别花茶中中药成分的准确度以及精确度还有召回率会比其他两种不同的网络模型识别度高, 这充分说明本文所提出的模型, 在花茶成分数据集上具有良好的分类性能, 从而有效的说明引入Swin Transformer模块, 能够充分利用特征的局部信息进行重点关注, 以及前后信息之间的特征传递, 从而对成分的有效识别。

4 结论

本文搭建了一种基于Swin Transformer用于检测亳州花茶成分识别与检测, 相较于传统方法, 通过对花茶成分复杂数据的学习和理解, 我们的模型在花茶成分的识别准确度和速度上均取得了显著提升。这为生产过程的自动化和质量控制提供了可行的解决方案。Swin Transformer模型的引入不仅提高了亳州花茶的品质检测水平, 为花茶识别提供了一种强大而灵活的工具, 降本增效, 可以更好地应对复杂、多样化的花茶图像数据。同时对企业来说, 不仅可以增强产品竞争力和市场地位, 还有助于提升品牌形象、拓展产品创新, 促进产业链的可持续发展, 对企业发展具有重要意义, 也对其他产业具有一定的参考价值 and 意义。

[基金项目]

基于深度学习的亳州花茶识别的研究(2023BYK2311)。

[参考文献]

- [1]潘雪峰. 基于深度学习的中医古籍自然语言处理研究[D]. 长春中医药大学, 2023. 771.
- [2]鲍艳春, 石彩霞, 张传强, 等. 深度学习在基因组学中的研究进展[J/OL]. 遗传, 1-15.
- [3]郭凤英, 高晨曦, 王一凡, 等. 基于深度学习和医案相似度的古代方剂智能推荐方法研究[J/OL]. 中华中医药学刊, 1-12.
- [4]杨静, 郝润平, 张迎迎. 基于深度学习的中草药植物视觉识别方法研究[J]. 中国体视学与图像分析, 2023, 28(2): 203-211.
- [5]张通. 基于SwinTransformer和CNN的手写体汉字识别算法研究与应用[D]. 青岛科技大学, 2023.
- [6]朱宗楠. 基于图像分类的Transformer模型优化研究[D]. 中原工学院, 2023.
- [7]张卓然, 张倩, 宋智. 基于残差SwinTransformer的天气图像识别技术研究[J]. 成都信息工程大学学报, 2023, 38(6): 637-642.
- [8]张中泽, 齐洁. 基于Transformer模型的多语种同时翻译[C]. 东华大学信息科学与技术学院, 2021: 6.

作者简介:

董鹏(1996--), 男, 安徽亳州人, 硕士, 主要教学大数据方向。