

基于超图神经网络的双模态特征融合藏药材植株识别算法

陈健福¹, 孙 燕^{1*}, 裴九场²

1. 青海民族大学, 青海 西宁 810007

2. 青海藏医院, 青海 西宁 810003

摘要: **目的** 准确识别藏药材, 实现藏药材智能化挖掘及管理。**方法** 提出基于超图的双模态特征融合藏药材植株识别算法 HerbiFusionNet 模型。首先, 利用改进的 ResNet152-CA 模型提取藏药材图像的空间特征, 将基于 Transformer 架构的 BERT 模型提取藏药材文本的语义特征, 实现 2 种模态特征的互补与融合; 其次, 计算融合后特征向量的相似性, 构建超图网络; 最后, 通过超图神经网络捕获藏药材植株复杂关联关系, 获得藏药材准确的分类。**结果** 相比于单一模态 ResNet-152-CA 模型, 引入融合双模态特征并基于超图神经网络的 HerbiFusionNe 模型, 藏药材识别准确率为 96.28%, 其准确率增加了 4.40%。提出的 HerbiFusionNet 模型实证了融合图像和文本的双模态特征利用超图结构挖掘藏药材数据内复杂关系的有效性。**结论** HerbiFusionNet 模型提升了藏药材识别的准确率, 能有效捕捉藏药材图像与文本之间的高阶关系, 展现了超图神经网络在处理藏药植株复杂数据结构中的优势, 为后续深入挖掘“症状-方剂-药材”关系及安全使用奠定了标准化基础, 推动了藏药研究和应用的发展。

关键词: 藏药材; 超图; 双模态特征融合; 卷积神经网络; HerbiFusionNet 模型

中图分类号: TP18; R282.5 **文献标志码:** A **文章编号:** 0253-2670(2025)04-1310-08

DOI: 10.7501/j.issn.0253-2670.2025.04.019

Bimodal feature fusion algorithm of Tibetan medicinal materials recognition based on hypergraph neural network

CHEN Jianfu¹, SUN Yan¹, PEI Jiuchang²

1. Qinghai Minzu University, Xining 810007, China

2. Qinghai Tibetan Hospital, Xining 810003, China

Abstract: Objective To identify Tibetan medicinal materials accurately and realize the intelligent mining and management of Tibetan medicinal materials. **Methods** HerbiFusionNet, a hypergraph-based dual-modal feature fusion model for Tibetan medicinal plants recognition, was proposed. Firstly, the improved ResNet152-CA model was used to extract the spatial features of Tibetan medicinal materials images, and the Transformer architecture based BERT model was used to extract the semantic features of Tibetan medicinal materials texts to realize the complementarity and fusion of the two modal features. Then, the similarity of the fused feature vectors was calculated to construct a hypergraph network. Finally, the complex association relationship of Tibetan medicinal plants was captured by hypergraph neural network, and the accurate classification of Tibetan medicinal materials was obtained. **Results** The experimental results show that, compared with the single-modal ResNet-152-CA model, the HerbiFusionNe model based on hypergraph neural network introduced by fusing dual-modal features has an accuracy rate of 96.28%, which is increased by 4.40%. The HerbiFusionNet model proposed in this study demonstrates the effectiveness of using hypergraph structure to mine complex relationships in Tibetan medicinal materials data by fusing dual-modal features of image and text. **Conclusion** The HerbiFusionNet model improves the accuracy of Tibetan medicinal materials recognition, can effectively capture the high-order relationship between Tibetan medicinal materials images and texts, and shows the advantages of hypergraph neural network in dealing with the complex data structure of Tibetan medicinal plants. It lays a standardized foundation for further exploration of the “symptoms-prescription-

收稿日期: 2024-10-02

基金项目: 青海省海南州可持续发展议程创新示范区科技创新平台项目 (2024-HN-P04); 中国高校产学研创新基金异构智能计算专项 (2024HY004)

作者简介: 陈健福 (1999—), 男, 硕士, 从事中医药数据分析研究。E-mail: 2909708181@qq.com

*通信作者: 孙 燕 (1973—), 女, 教授, 硕士生导师, 研究方向为复杂网络及其应用。E-mail: sy0623@163.com

medicinal materials” relationship and safe use, and promotes the development of Tibetan medicine research and application.

Keywords: Tibetan medicinal materials; hypergraph; bimodal feature fusion; convolutional neural networks; HerbiFusionNet model

藏药起源于青藏高原的传统医药体系，高度依赖于药用植物的使用。这些药材来自该地区的高海拔和多样化的生态系统，几个世纪以来一直被用来治疗各种疾病^[1-2]。许多传统藏草药已被临床证明具有强大的药理特性，包括抗炎、镇痛、抗菌和抗氧化作用。然而，使用藏药材却有着 2 个挑战：一是毒性，一些藏药材使用不当可能会产生不良影响，所以在使用草药前需咨询合格的藏医；二是缺乏对藏药材种植、加工和质量控制的标准化规范。藏药材识别依赖于经验丰富的专业藏医学技术人员，通过观察、触摸、品尝和闻嗅等传统方法进行鉴别。人工识别极易出现错误，准确性高度依赖于技术人员的经验积累。众所周知，培养中藏医方向的专业人才需要长期的训练。目前行业内没有成体系的藏药材识别分类的标准发布，现状制约了藏药材医疗事业的现代化、智能化的进步和发展。

随着日新月异科学技术的发展，基于深度学习的神经网络广泛应用于军事、农业和医学各个领域^[3]。利用计算机图像处理及视觉技术自动识别和分类藏药材，不仅减轻了人力成本、减少了人工误判，可提高藏药材的分类识别的效率和精度。然而，在复杂背景下识别草本藏药材仍然有很多困难：（1）藏药材的形态多样性，许多草药表现出相似的特征，仅依靠视觉特征难以区分不同种类。（2）缺乏全面的藏药材图像数据集，现有的数据集在规模、种类和质量上都存在局限，阻碍了鲁棒的图像识别模型的开发。（3）现实场景中遇到的复杂光照和背景条件会影响图像质量和识别精度，光照强度、角度和背景杂乱的差异会导致噪声和伪影，干扰图像的训练结果。应对这些挑战，研究人员探索了各种技术，有深度学习、图像特征提取和多模态分析。

本研究提出了基于超图神经网络的双模态特征融合的藏药植株识别算法，分别提取图像模态的空间特征和文本模态的语义特征后，再融合图像和文本特征，融合后的特征向量作为藏药植株的特征属性，用超图神经网络对双模态特征进行学习，为识别藏药材提供了快速方法，挖掘了藏药材图像与文本的复杂关联关系。

1 基于神经网络识别药材的相关研究

目前基于卷积神经网络（convolutional neural

network, CNN）^[4]已经在药材识别中得到了广泛应用。李立鹏等^[5]采用 ResNet-101 网络和迁移学习技术，对苜蓿、芦笋、蓝马鞭草、车前草等 62 种野生植物进行识别，识别准确率 85.6%，结果表明残差神经网络模型对野生植物识别的有效性。高宏元等^[6]对分解 CNN 模型 Inception-V3 进行微调，识别了 293 种天然草地植物，准确率 89.41%，实验表明分解 CNN 模型在多样化植物识别中的能力。王一等^[7]提出一种基于多通道颜色空间与注意力机制的改进 CNN 方法，针对 34 种中药材的 8 774 张导管图像，识别准确率比原始 Res-Net-50 提升 4.1%，多通道注意力改进卷积对小样本分类准确性能力。温瑶等^[8]提出了一种对称卷积核多通道残差块，对 110 种的藏药材的准确率为 98.22%，结合对称多通道残差框架准确率。周丽媛等^[9]研究了基于多特征融合和 Alex-Net 的药材切片鉴别方法，特征融合结合神经网络 Alex-Net 对 32 种复杂背景下藏药材切片的 3 610 张图像识别准确率为 91.68%。以上研究可以看出基于 CNN 算法在药材识别中识别能力并得到了广泛应用。

但 CNN 也存在局限性：（1）泛化能力不足。CNN 模型在特定场景下的识别率较高，但其泛化能力较弱，难以有效应对不同场景或数据分布的变化。（2）对数据量有依赖性。CNN 模型的高准确率通常依赖于大量的训练数据，数据不足时，模型性能可能显著下降。（3）计算资源消耗大。CNN 模型在训练和部署时需要消耗大量计算资源，限制了其在内存和处理能力有限的设备上的应用。（4）处理序列数据困难。CNN 在处理结构化数据（如图像）时表现良好，但对于序列数据（如文本）的处理能力不足。这限制了其不同模态学习中的应用，无法有效挖掘文本与图像之间的复杂关联。

为挖掘不同模态数据的复杂关联关系，基于图论的普通图和超图的网络模型逐渐成为研究热点。超图神经网络（hypergraph neural network, HGNN）^[10]因其是捕捉高阶关系和复杂特征依赖的模型，得到了学者的关注。目前针对藏药材基于超图神经网络模型在藏药材的研究较少，国内外顶刊发表的藏药材分类研究成果几乎没有。

本研究基于 HGNN^[11]将藏药材的图像和相关

文本多模态特征融合到超图结构中,捕捉藏药材多模态特征间的高阶关联关系。相比于 CNN 模型,超图 CNN 对于多模态维度高、复杂关系数据,从数学底层逻辑具有优势,为藏药材的深度挖掘提供了新的技术路径。

2 藏药材数据集

藏药材植株数据集所选的 18 种藏药材基于青海省认定的“十八青药”标准,18 种藏药材分布在青海的各个地市县自然环境中,包括冬虫夏草、枸杞、唐古特大黄、青贝母、秦艽、羌活、锁阳、沙棘、獐牙菜、黄芪、红景天、甘松、当归、水母雪莲、铁棒锤、川赤芍、西南手参、绿绒蒿。

2.1 数据来源

本研究构建的每种藏药材植物双模态数据集包含高质量的图像和详细的文本描述,采用了 3 种数据收集方法,以确保数据样本的高质量 and 多样性。

2.1.1 实地采集图像 本研究与青海省藏医院合作,依据《中药资源普查》的 10 年调研结果,实地采集每种藏药材各 100 张高质量现场图像并制作标本,现场图像包括近景、远景等图像,共 1 800 张。采集标本的图文信息实例如图 1 所示。



图 1 第 4 次全国普查雪莲标本实例

Fig. 1 Sample of *Saussurea sinensis* from the 4th national census

2.1.2 网上爬取图像 采用 Python 爬取网上图像数据,数据来源于百度图片,确保所爬取的内容均为公开可获取的信息。针对每种藏药材,爬取 400 张图片,经过筛选和清洗后共获得 7 200 张高质量图像。最终,合并数据集,形成包含 18 种藏药材的总计 9 000 张图像。

2.1.3 专业人员描述文本 相应的文本数据参考了《中药资源普查》的结果和《四部医典》中的藏药植株描述,再通过青海省藏医药专业人员对每张图像进行相应的藏药植株形态特征和生长环境的详细描述,从而形成 9 000 条高质量的文本数据。

2.2 数据预处理

2.2.1 图像数据预处理 首先将图像统一转换为 JPG 格式,并压缩至不超过 500 kb。每类藏药材的图像单独存放在对应文件夹中,并统一数据标签。采用留出法将数据按 80%和 20%划分为训练集和测试集,并进行多次划分和训练评估,取其平均值作为最终结果,多次划分的目的相当于利用留出法产生验证集,以提高模型的泛化能力。为提高模型的鲁棒性和泛化能力,实施随机旋转、缩放、翻转、裁剪等数据增强,以及亮度、对比度、饱和度等颜色属性的随机调整^[12]。

2.2.2 文本数据的预处理 包括中文文本清洗、分词和向量化。清洗步骤:去除无关字符(如停用词和标点符号),减少噪声对模型分类效果的影响。分词将文本切割成有意义的词组,采用基于词典的字符串匹配或基于统计的机器学习分词,灵活处理中文语义。文本采用常用的分词工具 jieba 处理中文数据。这些操作为后续模型训练提供了多样化的数据基础。藏药材的图文信息示例如表 1 所示。

表 1 藏药材的图文信息示例

Table 1 Examples of graphic information on Tibetan medicinal materials

藏药名	藏药图像	文本描述
大黄		粗壮木质根状茎, 叶片长宽接近, 基部近心行花小, 花紫红色
雪莲		粗根状茎, 直立茎叶片椭圆形, 茎端密集球形总花序
獐牙菜		直立茎, 圆锥状复聚伞花序几乎占据整个植株, 蒴果椭圆状披针形
秦艽		匍匐茎, 基生叶和茎生叶, 顶生伞形总状花, 花蓝色

3 藏药材识别网络模型

3.1 图像模态

图像模态是将残差网络 ResNet152^[13] 的每个残差块中引入 Coord-Attention (坐标注意力) 模块^[14], 从而提出了改进的网络模型 ResNet1-52-CA 网络, 作为提取藏药材植株图像的特征模块, 如图 2 所示。

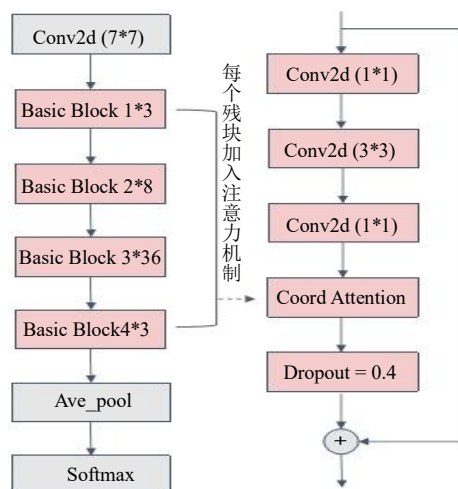


图 2 残差块中引入注意力机制

Fig. 2 Attention mechanism introduced into residual block

图 2 左边为 ResNet-152 的结构, 包含 4 个主要的残差模块, 分别包含 3、8、36、3 个残差块。Coord Attention 是一种增强空间特征表达的注意力机制, 捕获特征图的空间依赖并保留精确的位置定位信息。对输入特征图的高 (H)、宽 (W) 和通道数 (C) 进行编码, 采用 ($H, 1$) 和 ($1, W$) 的平均池化沿水平和垂直方向进行编码, 生成方向感知特征图。这种编码方法保留了空间方向的位置信息, 并捕获了特征的长期相关性。图 2 右边为 Coord Attention 模块嵌入每个残差块的网络结构。

3.2 文本模态

文本模态基于 Transformer^[15] 架构的 BERT (bidirectional encoder representations from transformers) 模型^[16] 来提取藏药材的文本特征, BERT 是一种预训练文本深度学习模型, 优势在于双向编码能力、捕捉句子中的全局上下文信息、文本理解的准确性, 利用自注意力机制 BERT 能高效捕捉文本中词汇之间的复杂关系, 可理解文本的上下文依赖, 模型结构如图 3 所示。

3.3 超图网络构建

3.3.1 藏药材超图构建 (1) 经 ResNet152-CA 捕

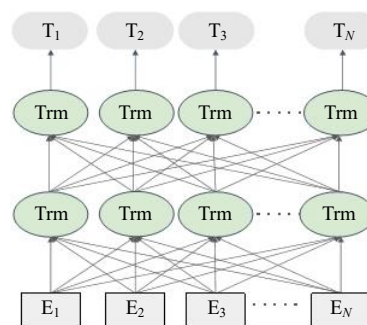


图 3 BERT 模型的架构

Fig. 3 Architecture of BERT model

获的图像模态特征及 BERT 模型获得的文本模态特征, 得到的特征维度分别为 2048 维和 768 维。(2) 采用拼接 CAT 方法^[17] 融合特征, 在通道数上保留了原始图像和文本特征信息, 为下游学习到了完整的特征表示。融合后的特征向量维度为 1×2816 。(3) 特征融合后, 超图的超节点为特征融合的藏药材, 超边为基于曼哈顿距离^[18] 相似度最高的 10 个节点形成, 构建超图网络^[19]。(4) 通过超图神经网络进行训练, 超图卷积层通过在超边邻域内聚合节点特征生成更具表征能力的节点特征, 得到更好的分类准确率。本研究提出的基于超图神经网络的双模态特征融合藏药材识别算法 HerbiFusionNet 流程如图 4 所示。

3.3.2 超图概念 图论中的超图可以捕捉图中节点之间更复杂的高阶关系。普通图中, 每条边连接 2 个节点 v_i, v_j , 边只能表征二元关系。超图的每条超边包含多个超节点, 超边记为 e_j , 超节点记为 v_i , 超边 e_j 包含超节点 v_i , 自然地表征超节点的高阶关系。如图 5 所示, 超图结构为 $e_1 = \{v_1, v_3\}$ 、 $e_2 = \{v_2, v_3\}$ 、 $e_3 = \{v_3, v_4, v_5\}$, 因为 v_1, v_2, v_4, v_5 节点仅参与了 1 条超边, 节点超度均为 1, 因为 v_3 节点参与了 3 条超边, 超节点 v_3 超度为 3。

本研究使用留出法对数据集进行划分, 但留出法需要训练集和测试集的数据分布相同, 不能引入额外的偏差导致对最终模型的训练和评价产生影响。因此, 本研究分析了藏药材测试集及训练集超图超节点度的概率分布。如图 6 所示, 超图超节点度分布具有幂律分布特性, 超网络中存在少数节点有高的超节点度, 大多数节点的超度较低。这种特性与复杂网络的结构相符。同时也说明训练集和测试集数据分布是一致的。

同时, 绘制了藏药材测试集构建的超图的双对数拟合曲线, 如图 7 所示。表明双对数拟合为直线,

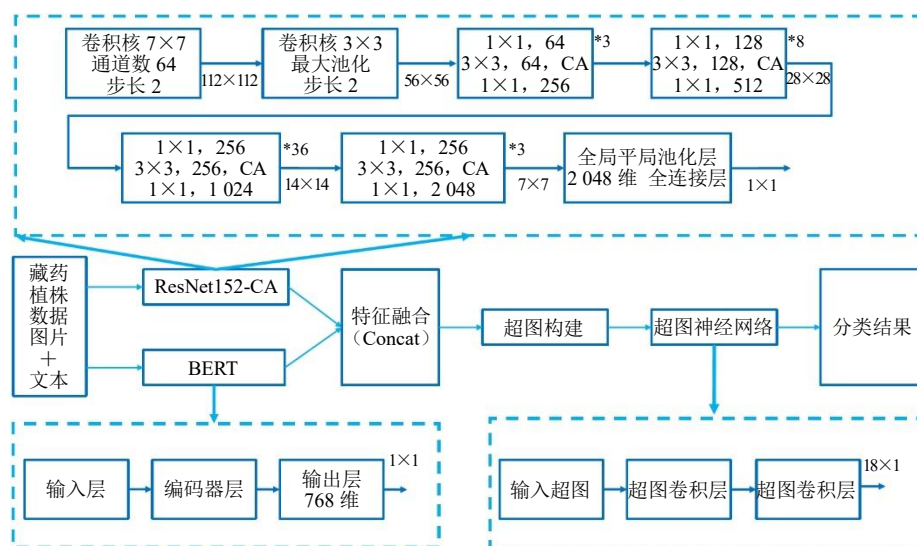


图 4 HerbiFusionNet 算法流程

Fig. 4 Flow of HerbiFusionNet algorithm

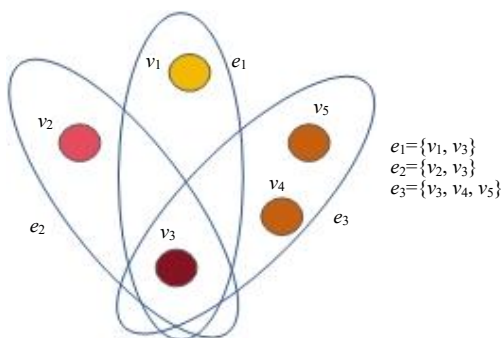


图 5 超图节点度分布示意图

Fig. 5 Distribution diagram of hypergraph node degree

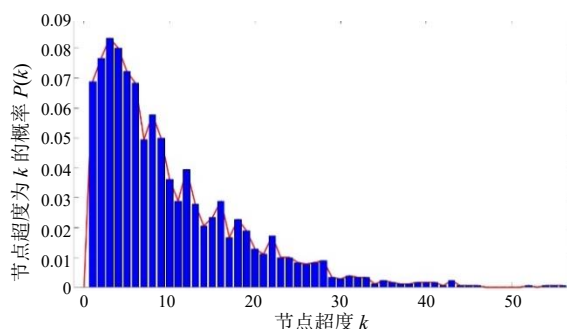


图 6 藏药材测试集超图网络图中节点度的概率分布

Fig. 6 Probability distribution of node degree in hypergraph network of Tibetan medicinal materials test set

实证了藏药材超图超节点度遵循后尾规律,表明少数节点在网络中扮演关键角色。超图的结构指标分析为后续研究和应用奠定了基础。

3.4 超图神经网络

HGNN 模型的输入之一是超图的关联矩阵 H ,

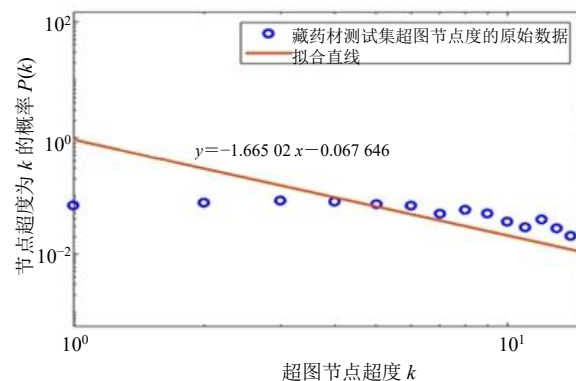


图 7 藏药材测试集构建超图的双对数拟合曲线

Fig. 7 Double logarithmic fitting curve of hypergraph constructed by test set of Tibetan medicinal materials

如果超边 e_j 包含超节点 v_i , 则关联矩阵相应的位置记为 1, 否则记为 0。超图的关联矩阵如公式 (1) 所示。

$$H = \begin{cases} 0 & v_i \notin e_j \\ 1 & v_i \in e_j \end{cases} \quad (1)$$

对超图神经网络的卷积层进行特征提取,利用超边的邻域内对节点特征进行聚合来实现。与普通图卷积不同,超图卷积算法^[20]考虑了超图的高阶关系,超节点聚合特征,捕捉超节点与超边之间的复杂拓扑关系。根据公式 (2) 构建超图卷积神经网络第 1 层结构。

$$Y^{l+1} = \sigma \left(D_v^{-\frac{1}{2}} H W D_e^{-\frac{1}{2}} H^T D_v^{-\frac{1}{2}} \right) Y^l \phi^l \quad (2)$$

Y^l 为超图上节点在第 l 层的输入特征, Y^{l+1} 为超图上节点在第 $l+1$ 层的输入特征, σ 是非线性激活函数, H 记为超图的

关联矩阵, W 为超边的权值, D_e 和 D_v 分别为超边度和超节点度的对角矩阵

4 结果与分析

4.1 实验设置

硬件配置为 Intel Core i9-13900K 处理器和 NVIDIA RTX 4090 显卡 (24 GB VRAM)。使用 Python 3.12 和 PyTorch 2.3.0, 计算任务通过 CUDA 加速完成。

4.2 参数设置

学习率 0.000 1, Batch-size 为 32, 优化算法为 Adam, 迭代次数为 50, Dropout 率设置为 0.4, 以防止过拟合。

4.3 基于 CNN 卷积神经网络算法

基于改进的 CNN 算法 VGG-16、GoogleNet、ResNet-18、ResNet-50、DenseNet-121、ResNet-152-CA 识别藏药材图像, 利用准确率和参数量进行性能评估, 准确率是指模型正确预测的样本数占总样本数的比例, 参数量是指模型中需要训练的参数的总数, 它反映了模型的复杂度, 如表 2 所示。模型均使用交叉熵作为损失函数, 配置 Adam 优化器。ResNet-152-AC 准确率为 91.95%, 参数量较 DenseNet-121 低; Google-Net 准确率为 89.67%; 虽然 VGG-16 准确率较高, ResNet-50 参数量较大, 但需求的计算资源高。

4.4 基于多模态特征融合分类算法

藏药材植物在复杂背景环境下, 单一 CNN 神经网络识别准确率不高, 影响了分类准确性。本研究基于图像和文本的双模态特征融合方法用准确率和 F1 值 2 个指标进行比较, F1 值是精确率和召回率的调和平均值, 用于衡量模型对正例样本的预测能力, F1 值越高, 说明模型在预测正例时的精确度和召回率都较好。以交叉熵作为损失函数, Adam

优化器模型的收敛速度和训练效果。特征融合模态的对比结果如表 3 所示。由表 3 可见, ResNet-152-CA+BERT 算法准确率为 95.28%, F1 值为 95.30%; DenseNet-121+BERT 的准确率为 93.67%, F1 值为 93.78%; GoogleNet+BERT 和 ResNet-50+BERT 算法在准确率上分别为 90.11%和 90.39%; VGG-16+BERT 算法有 87.89%的准确率。考虑准确率和 F1 值的平衡, ResNet-152-CA+BERT 算法最均衡。

表 3 基于多模态特征融合算法比较
Table 3 Comparison of algorithms based on multimodal feature fusion

模型	准确率/%	F1 值/%
VGG-16+BERT	87.89	87.94
Google-Net+BERT	90.11	90.13
ResNet-18+BERT	90.11	90.14
ResNet-50+BERT	90.39	90.41
DenseNet-121+BERT	93.67	93.78
ResNet-152-CA+BERT	95.28	95.30

4.5 基于超图神经网络特征融合算法

改进的超图神经网络的实验结果如表 4 所示。引入 HGNN, 模型的准确率和 F1 值提升, 证实了 HGNN 捕获藏药材复杂关系的有效性。HerbiFusionNet 算法准确率为 96.28%, F1 值为 96.35%; 算法 DenseNet-121+BERT+HGNN 准确率为 95.72%, F1 值为 96.13%; 算法 GoogleNet+BERT+HGNN 和 ResNet-50+BERT+HGNN 的准确率分别为 94.67%和 94.94%。相比之下, 虽然 VGG-16+BERT+HGNN 准确率为 93.89%, 其较大的参数需求计算资源更大。由此可见, HerbiFusionNet 模型在准确率和 F1 值方面最高。

表 2 基于 CNN 图像模态算法比较
Table 2 Comparison of image modal algorithms based on CNN

模型	参数量	准确率/%
VGG-16	134.33	85.56
GoogleNet	56.18	89.67
ResNet-18	11.18	90.22
ResNet-50	21.29	89.28
DenseNet-121	69.72	92.89
ResNet-152-CA	59.14	91.95

表 4 基于超图的特征融合模态算法比较
Table 4 Comparison of feature fusion modal algorithms based on hypergraph

模型	准确率/%	F1 值/%
VGG-16+BERT+HGNN	93.89	94.12
GoogleNet+BERT+HGNN	94.67	94.72
ResNet-18+BERT+HGNN	94.11	94.12
ResNet-50+BERT+HGNN	94.94	94.99
DenseNet-121+BERT+HGNN	95.72	96.13
ResNet-152-CA + BERT + HGNN (HerbiFusionNet)	96.28	96.35

4.6 超图构建的方法类型验证

本研究比较了欧氏距离相似度、余弦相似度和哈曼顿距离相似度 3 种方法构建超图性能。步骤如下：（1）确定骨干网络统一为 ResNet-152-CA + BERT；（2）消融实验评估超图构建的 3 种方法的分类性能。由表 5 可见，欧氏距离和哈曼顿距离的计算复杂度相似，基于哈曼顿距离构建的超图性能最高。余弦相似度的计算复杂度虽然最低，但准确率低，参数量主要由样本数量 n 和特征维度 d 决定。由哈曼顿距离构建的超图分类准确率为 96.28%，相比于欧式距离和余弦相似度构建超图分别提升了 5.39% 和 11.25%。结果表明，由哈曼顿距离来构建超边，使得超图能够有效地反映样本之间的关联关系，并为后续的超图神经网络训练提供基础。

表 5 超图构建的方法类型验证
Table 5 Verification of method type for hypergraph construction

构建超图	计算复杂度	准确率/%
欧氏距离	$O(n^2 \cdot d)$	90.89
余弦相似度	$O(n^2)$	85.03
哈曼顿距离	$O(n^2 \cdot d)$	96.28

4.7 超图神经网络有效性验证

为了验证 HGNN 的有效性，本研究将 HGNN 结构改为普通的图卷积神经网络（graph convolutional networks, GCN），并保持其他设置不变。HGNN 有效性验证结果如表 6 所示，超图结构在捕获不同模态数据相关性方面的效果明显优于普通图结构，其准确率和 F1 指标分别提升了 4.56% 和 4.37%。这表明，HGNN 中的超图卷积操作能够捕获更为细致的深层向量，增强了模态之间的融合度，从而显著提升了模型的准确率。

4.8 消融实验

本研究的消融实验基于藏药材的双模态数据和 HGNN，评估了不同模型对分类性能的影响。结果如表 7 所示，单一图像模态的所有模型平均准确率为 89.93%，F1 值为 88.98%。引入文本模态后，

表 6 超图神经网络有效性验证
Table 6 Validation verification of hypergraph neural networks

模型	准确率/%	F1 值/%
HerbiFusion + GCN	91.72	91.98
HerbiFusion + HGNN	96.28	96.35

表 7 消融实验结果

Table 7 Results of ablation experiment

图像	文本	HGNN	准确率/%	F1/%
√	×	×	89.93	88.98
√	√	×	91.24	91.70
√	√	√	94.94	95.20

所有模型的准确率提升至 91.24%，F1 值提升至 91.70%，表明双模态融合显著增强了分类效果。进一步将 HGNN 应用于融合特征后，所有模型的平均准确率提升至 94.94%，F1 值提升至 95.20%。表明 HGNN 挖掘了多模态特征中的复杂关系。消融实验实证了超图神经网络处理高阶关联关系及捕获多模态融合特征的有效性。

5 结论

本研究提出了一种基于超图的双模态特征融合藏药材识别算法 HerbiFusionNet，该算法采取 ResNet152-CA 模型提取图像特征，BERT 模型提取文本特征。利用双模态特征融合向量构建超图网络，捕获了藏药材的复杂的关联关系。HerbiFusionNet 算法融合了图像和文本信息，基于 HGNN 的建模藏药材的高阶关系，提高了藏药材识别的准确性。本研究基于融合双模态特征的 HGNN 模型使得藏药材分类的平均准确率从 89.93% 升至 91.24%。融合双模态特征的加入超图使得平均准确率从 91.24% 升至 94.94%。提出的 HerbiFusionNet 模型准确率 96.28%，证实了 HGNN 可提高分类性能的作用。本研究的藏药材数据集采用 18 种包含 9 000 张高质量图像和详细的文本描述，为模型训练提供了数据基础。HerbiFusionNet 模型为藏药材的智能识别提供了新的技术路径，实证了特征融合的 HGNN 对藏药材识别分类性能。本研究的 HerbiFusionNet 模型不仅适用于藏药材的图像和文本分类，还具备推广至其他植物分类任务的潜力。通过合理调整模型参数和数据预处理方法，该模型能够在不同植物类别、生态环境或地区的药用植物识别中发挥重要作用。例如，在农业领域 HerbiFusionNet 可应用于农作物健康监测，结合图像和文本数据来分析作物的生长状态，从而帮助农民及时采取相应措施。今后的工作以更多模态数据优化超图构建方法以及探索更先进的特征融合技术，以提升模型的性能和适用范围。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 才让南加. 藏医本草发展史概述 [J]. 中医药导报, 2023, 29(6): 80-86.
- [2] 完么安建. 藏医学概论 [J]. 中西医结合心血管病电子杂志, 2017, 5(15): 8-9.
- [3] 郑远攀, 李广阳, 李晔. 深度学习在图像识别中的应用研究综述 [J]. 计算机工程与应用, 2019, 55(12): 20-36.
- [4] 王艳, 孙薇, 周小平. 基于深度学习的中草药植物图像识别方法研究 [J]. 中医药信息, 2020, 37(6): 21-25.
- [5] 李立鹏, 师菲蓬, 田文博, 等. 基于残差网络和迁移学习的野生植物图像识别方法 [J]. 无线电工程, 2021, 51(9): 857-863.
- [6] 高宏元, 高新华, 冯琦胜, 等. 基于深度学习的天然草地植物物种识别方法 [J]. 草业科学, 2020, 37(9): 1931-1939.
- [7] 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.
- [8] 温瑶, 祁晋东, 周艳霞, 等. 基于深度神经网络的藏药材识别分类研究 [J]. 信息与电脑, 2022, 34(24): 151-153.
- [9] 周丽媛, 高红梅, 赵启军, 等. 基于多特征融合结合深度学习模型的药材切片鉴别 [J]. 世界科学技术—中医药现代化, 2024, 26(1): 211-217.
- [10] Feng Y F, You H X, Zhang Z Z, *et al.* Hypergraph neural networks [J]. *Proc AAAI Conf Artif Intell*, 2019, 33(1): 3558-3565.
- [11] Zhao W, Tan S L, Guan Z Y, *et al.* Learning to map social network users by unified manifold alignment on hypergraph [J]. *IEEE Trans Neural Netw Learn Syst*, 2018, 29(12): 5834-5846.
- [12] 朱晓慧, 钱丽萍, 傅伟. 图像数据增强技术研究综述 [J]. 软件导刊, 2021, 20(5): 230-236.
- [13] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [A] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Las Vegas: IEEE, 2016: 770-778.
- [14] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design [A] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Nashville: IEEE, 2021: 13708-13717.
- [15] Cohen S B, Collins M. Advances in neural information processing systems 25 [A] // *Neural Information Processing Systems* [M]. Lake Tahoe: Neural Information Processing Systems Foundation, 1997: 1-4.
- [16] Devlin J, Chang M W, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2018-10-11]. <https://arxiv.org/abs/1810.04805v2>.
- [17] 胡良梅, 高隽, 何柯峰. 图像融合质量评价方法的研究 [J]. 电子学报, 2004, 32(S1): 218-221.
- [18] 李彬. 基于相对曼哈顿距离的 Web 聚类算法研究 [J]. 电子商务, 2013(11): 57-58.
- [19] 李全鑫, 庞俊, 朱峰冉. 结合 Bert 与超图卷积网络的文本分类模型 [J]. 计算机工程与应用, 2023, 59(17): 107-115.
- [20] 胡秉德, 王新根, 王新宇, 等. 超图学习综述: 算法分类与应用分析 [J]. 软件学报, 2022, 33(2): 498-523.

[责任编辑 潘明佳]