

基于结构感知花瓣拓扑图的细粒度牡丹识别

陈寒睿^{1*#}, 鞠奥宇^{1*}, 郑宏艺¹, 李军保¹, 张 玉², 李 真³, 刘扬帆¹

¹洛阳理工学院电子信息学院, 河南 洛阳

²洛阳理工学院材料学院, 河南 洛阳

³洛阳理工学院人工智能学院, 河南 洛阳

收稿日期: 2026年6月28日; 录用日期: 2026年7月20日; 发布日期: 2026年7月30日

摘 要

目的: 针对牡丹图像中重瓣花冠遮挡、花瓣边界粘连和局部纹理相近等问题, 构建基于花瓣拓扑关系的识别模型PTRGNet。方法: 训练阶段采用SAM候选掩码与人工校正获得实例级花瓣标注; 测试阶段由训练后的实例分割器自动输出花瓣候选。模型以花瓣实例为节点, 融合空间形态特征和ROI语义特征, 并将距离、方向角、重叠率、外接框IoU和面积差作为边属性输入边感知GAT, 再与视觉分支门控融合。结果: 在Peony-8上, PTRGNet取得67.8% Top-1准确率和67.1% F1分数; 在Oxford Flowers-102上, 弱伪拓扑设置取得89.8% Top-1和98.6% Top-5。结论: 花瓣拓扑可为牡丹识别提供颜色、纹理之外的结构证据, 但模型性能仍受实例分割质量影响。

关键词

牡丹识别, 花瓣拓扑, 实例分割, 图注意力网络, 结构感知分类

Fine-Grained Peony Recognition Based on a Structure-Aware Petal Topology Graph

Hanrui Chen^{1*#}, Aoyu Ju^{1*}, Hongyi Zheng¹, Junbao Li¹, Yu Zhang², Zhen Li³, Yangfan Liu¹

¹School of Electronic Information, Luoyang University of Technology, Luoyang Henan

²School of Materials Science and Technology, Luoyang University of Technology, Luoyang Henan

³School of Artificial Intelligence, Luoyang University of Technology, Luoyang Henan

Received: June 28, 2026; accepted: July 20, 2026; published: July 30, 2026

Abstract

Objective: Peony recognition is often disturbed by dense petals, weak boundaries and similar local

*共同第一作者。

#通讯作者。

文章引用: 陈寒睿, 鞠奥宇, 郑宏艺, 李军保, 张玉, 李真, 刘扬帆. 基于结构感知花瓣拓扑图的细粒度牡丹识别[J]. 图像与信号处理, 2026, 15(3): 401-414. DOI: 10.12677/jisp.2026.153036

textures. This work builds a Petal Topology Relation Graph Network (PTRGNet) to introduce structural cues into fine-grained peony classification. Methods: SAM is first used as an annotation aid to produce candidate petal masks, which are then corrected manually for instance-level supervision. At test time, petal masks are generated by the trained instance segmenter without manual input. PTRGNet represents each petal as a graph node, combines spatial morphology with ROI semantic features, and describes pairwise relations using distance, direction, mask overlap, bounding-box IoU and relative area difference. These edge attributes are fed into an edge-aware graph attention module and fused with the visual branch through a gated head. Results: On Peony-8, PTRGNet reached 67.8% Top-1 accuracy, 66.9% precision, 67.3% recall and 67.1% F1 score under predicted masks, exceeding EfficientNet-B0 by 8.1 percentage points. Human-refined masks gave 70.2% accuracy, while zero-shot SAM masks gave 63.5%, showing the effect of mask noise. On Oxford Flowers-102, the weak pseudo-topology setting achieved 89.8% Top-1 and 98.6% Top-5 accuracy. Conclusion: Petal topology supplies useful structural evidence for peony recognition, although stronger validation still depends on reliable automatic masks and larger instance-level flower datasets.

Keywords

Peony Recognition, Petal Topology, Instance Segmentation, Graph Attention Network, Structure-Aware Classification

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

细粒度视觉分类的难点在于类间差异细微、类内变化明显。Peony-8 类别数不大,但牡丹花冠常见重瓣、遮挡和内外层花瓣混杂,单靠颜色或纹理容易受背景与局部相似性干扰。因此,本文关注花瓣层次、相对位置和重叠关系能否为牡丹识别提供更稳定的结构线索[1][2]。

从植物形态描述看,牡丹识别不只依赖花色,花型、瓣形、层次和花冠结构同样重要[3]。卷积网络和视觉 Transformer 能学习轮廓与纹理,但在同色花瓣交叠或边界模糊时,容易受阴影、背景和偶然色块影响。将花瓣实例及其关系显式建模,可把形态经验转化为可计算输入。

近年来,知识蒸馏、视觉提示学习、多区域表示和视觉语言模型已用于花卉及细粒度识别[4][5]。因此,评价结构建模方法不能只依赖弱基线。本文在统一划分和训练条件下比较轻量模型、卷积骨干、Transformer 骨干与 PTRGNet,并同时报告精度与复杂度。

花瓣结构建模还面临实例标注问题。若测试阶段仍需人工点选或修正,模型就难以自动部署。本文仅在训练标注阶段使用 SAM 生成候选区域;验证集和测试集主流程均由训练后的实例分割器自动产生花瓣掩码,再进入拓扑图模块。

1.1. 任务定义

在该设置中,Peony-8 被用作有限类别牡丹结构识别数据集。输入是一幅花冠图像,输出是牡丹类别;模型在预测时同时利用全局视觉特征、花瓣实例特征和花瓣之间的关系特征。Oxford Flowers-102 不含花瓣实例真值,因而主要承担公开花卉数据上的泛化参照作用,并用来观察弱伪拓扑在无实例标注条件下的表现。

这里关注的不是“大量类别检索”本身,而是花冠内部结构是否真的参与了类别判别。Peony-8 中不

少类别在颜色、采集背景和单片花瓣纹理上存在重叠, 普通分类器容易抓住局部色块或背景相关性。结构感知建模把视角从单片纹理转向多片花瓣的排列、遮挡和尺度关系, 判别依据也随之更接近植物形态观察。

上述边界限定了实验解释: Peony-8 用于分析有限类别牡丹识别中的结构线索, Oxford Flowers-102 用于观察粗糙伪拓扑在公共花卉数据上的变化, 二者不能替代更大规模的实例级花卉基准。

1.2. 主要贡献

提出 PTRGNet, 将花瓣实例组织为图节点, 并将距离、方向角、重叠率、外接框 IoU 和面积差作为边属性送入关系学习过程。

形成 SAM 候选掩码与人工校正结合的花瓣实例标注流程, 对重瓣样本中的欠分割、碎片伪影和遮挡层次进行一致处理。

构造边感知图注意力层, 使几何边属性参与注意力计算, 而不是预先用手工公式给边分配固定权重。

在相同划分、相同增强策略和相同训练范围内比较轻量模型、卷积骨干、Transformer 骨干和 PTRGNet, 并同时报告识别性能与计算开销。

2. 相关工作

2.1. 细粒度分类与花卉识别

细粒度分类经历了从整图特征到部件定位、多区域注意力和 Transformer token 交互的演进。RA-CNN 递归定位判别区域, TransFG 利用 token 关系刻画局部差异[6] [7]。这些研究说明, 花卉识别需要关注局部区域和部件关系, 而不能只与普通分类器比较。

国内花卉图像识别研究也开始引入知识蒸馏、改进 ViT 和视觉提示学习等技术, 以缓解小样本条件下的泛化问题[4] [5]。PTRGNet 的切入点不同: 它并不继续加深视觉骨干, 而是把花瓣实例之间的结构关系单独建模, 再与视觉特征融合。

小规模数据集对训练设置很敏感。预训练、增强和学习率都会影响强骨干表现。为避免弱基线偏差, 所有基线采用相同划分、输入分辨率和增强策略; PTRGNet 默认视觉骨干为 MobileNetV2+CBAM, 表中差异主要反映结构分支的影响。

2.2. 实例分割与 SAM 辅助标注

花瓣拓扑图要求实例节点可靠。标准 U-Net 输出的是语义区域, 难以自然区分同类重叠花瓣; Mask R-CNN 一类实例分割模型则可同时给出目标边界、类别和实例掩码[8] [9]。SAM 能快速产生候选区域, 但在重瓣牡丹图像中, 内层花瓣常被合并成一片, 边缘处也会出现碎片化掩码[10]。基于这一现象, SAM 仅作为标注辅助工具使用, 不直接替代人工核验后的实例真值。

具体做法是: 训练集先由 SAM 生成候选掩码, 随后人工完成拆分、合并和边界修正, 由此得到花瓣实例监督; 模型部署或测试时, 只输入原始图像, 训练后的分割器自动产生花瓣实例, 拓扑建图模块再据此生成节点和边。主结果由这条自动流程得到, 人工精修掩码只用于分析上限。

2.3. 图神经网络与视觉语言前沿

图神经网络适合描述数量不固定的对象集合及其关系, 图卷积、图注意力和图表示学习的基本范式已有较系统的综述[11] [12]。在视觉任务中, 区域、部件或目标之间的关系常被显式组织成图。与此同时, FineCLIP 等视觉语言模型把区域语义对齐引入细粒度理解, 图基础模型中的可迁移图词汇也为跨任务结

构推理提供了新的思路[13] [14]。

视觉场景图生成(Scene Graph Generation, SGG)也将图像组织为对象节点和关系边, 典型方法通过信息传递联合建模目标、关系和上下文[15]。不同的是, SGG 面向通用对象 - 谓词关系, PTRGNet 面向同一花冠内部花瓣的几何拓扑, 边属性为距离、方向角、重叠率、外接框 IoU 和面积差, 不预测自然语言谓词。

3. 方法

3.1. 整体架构

PTRGNet 由视觉特征提取、花瓣实例分割与建图、边感知 GAT 以及门控融合分类头组成。视觉分支给出整朵花冠的外观表示和 ROI 语义特征; 实例分割分支产生花瓣候选; 建图模块计算节点属性和边属性; GAT 聚合相邻花瓣的信息; 分类头最后根据视觉向量和图向量的融合结果输出类别。

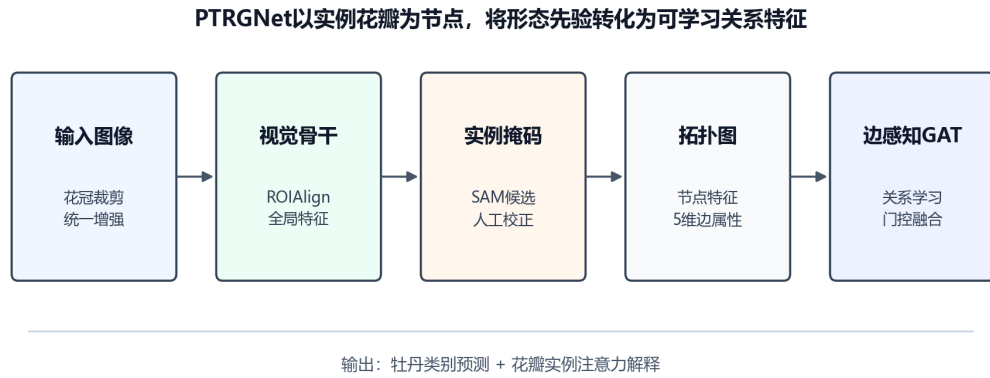


Figure 1. Overall architecture of PTRGNet
图 1. PTRGNet 整体架构

3.2. 视觉分支

设输入图像为 x , 视觉骨干输出特征图 $F = \text{Backbone}(x)$ 。PTRGNet 默认采用 MobileNetV2+CBAM, 以控制参数数量和时延; ResNet50、EfficientNet-B0 和 ViT-B/16 作为独立视觉基线。对每个花瓣 ROI, 采用 RoIAlign 或平均池化得到语义特征。

$$F_{\text{att}} = A(F) \odot F \quad (1)$$

式中, A 为注意力模块或 Transformer token 产生的响应图, $\odot$ 表示逐元素乘法。该视觉注意力仅作为特征增强组件, 核心贡献仍在花瓣实例拓扑建模。

3.3. 花瓣实例分割与标注来源

拓扑图依赖实例级花瓣掩码 $M = \{m_1, m_2, \dots, m_n\}$ 。Peony-8 的训练掩码由“SAM 辅助预标注 + 人工校正”产生。标注时先裁剪花朵前景, 再用 Dense Grid Point Prompts 和花冠级 Box Prompts 调用 SAM, 之后由标注者处理实例拆分、候选合并、边界修正和遮挡层次记录。

$$M_0 = \text{SAM}(x; P_{\text{grid}}, B_{\text{flower}}) \quad (2)$$

$$M_{\text{gt}} = \text{HumanRefine}(M_0; \text{split}, \text{merge}, \text{boundary}, \text{layer}) \quad (3)$$

标注记录显示, 单瓣和半重瓣样本中 SAM 候选掩码大多能覆盖主要花瓣, 召回率约为 85%, 人工主

要修整边缘。重度重瓣样本更易欠分割, 需要人工拆分。总体人工校正时间约为纯手工实例标注的 30%, 该数值仅表示标注效率。

测试阶段不把人工精修掩码作为主输入。人工掩码用于训练实例分割器; 测试时由分割器预测 M_{pred} , 再根据 M_{pred} 构建拓扑图。为了单独观察掩码质量的影响, 实验设置了三种输入: 人工精修掩码(Oracle mask), 训练后分割器预测掩码(Predicted mask), 以及未经人工修正的 SAM 零样本掩码(SAM zero-shot mask)。

三种掩码条件对应不同含义。Oracle mask 给出节点质量较理想时的参考上限; Predicted mask 对应实际自动识别流程; SAM zero-shot mask 则反映外部分割模型直接迁移到牡丹重瓣图像时的噪声影响。除专门的掩码鲁棒性实验外, 主表中的 PTRGNet 均采用 Predicted mask。

3.4. 拓扑图构建

获得实例掩码和视觉特征后, 构建图 $G=(V, E, X, Z)$ 。 V 中每个节点对应一个花瓣实例; X 保存空间形态特征和 ROI 语义特征; Z 描述两片花瓣之间的几何关系。

$$\phi_{\text{spatial}}(i) = [x_c, y_c, \text{area}, \text{aspect}, \text{circularity}, \text{orientation}, \text{layer}] \quad (4)$$

$$h_i = \text{MLP}([\phi_{\text{spatial}}(i) \parallel f_{\text{semantic}}(i)]) \in R^{256} \quad (5)$$

$$z_{ij} = [d_{\text{norm}}(i, j), \theta_{\text{norm}}(i, j), \text{overlap}(i, j), \text{IoU}(b_i, b_j), |A_i - A_j| / \max(A_i, A_j)] \in R^5 \quad (6)$$

其中, $d_{\text{norm}}(i, j)$ 为两个质心之间的欧氏距离, 并按图像对角线归一化; $\theta_{\text{norm}}(i, j)$ 为节点连线方向角归一化到 0 至 1 后的数值; $\text{IoU}(b_i, b_j)$ 为两个外接框的交并比; $\text{overlap}(i, j) = |m_i \cap m_j| / \min(|m_i|, |m_j|)$ 表示较小掩码被另一掩码覆盖的比例, 用来刻画遮挡和强重叠。

z_{ij} 保持为 5 维边属性, 不预先压缩为人工权重。完全枚举花瓣对的复杂度为 $O(|V|^2)$; 实现中用 k-NN 或 radius graph 将边数控制在 $O(k|V|)$ 。距离、角度、IoU 和面积差均通过 PyTorch broadcasting 在 GPU 上批量计算。

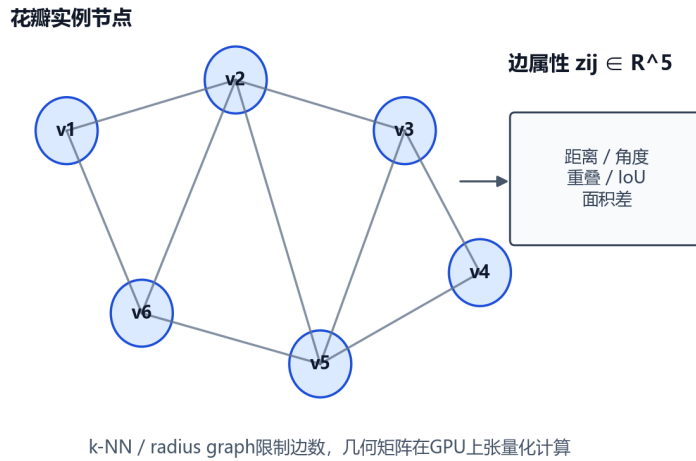


Figure 2. Petal topology graph construction

图 2. 花瓣拓扑图构建示意

算法 1. 花瓣拓扑图构建(Algorithm 1. Petal topology graph construction).

Require: image x , instance masks $M = \{m_1, m_2, \dots, m_n\}$, feature map F , neighbor size k , radius r

Ensure: topology graph $G = (V, E, X, Z)$

1: $V \leftarrow \emptyset, E \leftarrow \emptyset, X \leftarrow \emptyset, Z \leftarrow \emptyset$

2: for each $m_i \in M$ do

3: $b_i, c_i, A_i, s_i \leftarrow \text{ExtractGeometry}(m_i)$ // box, centroid, area and shape descriptors

4: $f_{\text{semantic}}(i) \leftarrow \text{RoIAlign}(F, b_i)$ or $\text{MaskPool}(F, m_i)$

5: $h_i \leftarrow \text{MLP}([\Phi_{\text{spatial}}(i) \parallel f_{\text{semantic}}(i)])$

6: $V \leftarrow V \cup \{v_i\}; X \leftarrow X \cup \{h_i\}$

7: end for

8: for $i \leftarrow 1$ to $|V|$ do

9: $N_i \leftarrow \text{KNN}(c_i, k) \cup \text{Radius}(c_i, r)$ // candidate neighbors

10: for each $j \in N_i$ and $j \neq i$ do

11: $z_{ij} \leftarrow \text{EdgeFeature}(m_i, m_j, b_i, b_j, A_i, A_j)$

12: $E \leftarrow E \cup \{(i, j)\}; Z \leftarrow Z \cup \{z_{ij}\}$

13: end for

14: end for

15: return $G = (V, E, X, Z)$

3.5. 边感知图注意力网络

设节点隐藏维度为 d 。 $h_i, h_j \in R^d$ ，边属性 $z_{ij} \in R^5$ ，节点投影矩阵 $W \in R^{d \times d}$ ，边投影矩阵 $W_e \in R^{d \times 5}$ ，注意力向量 $a \in R^{3d}$ 。 W_e 将 5 维边属性映射到 d 维空间后，与 Wh_i 、 Wh_j 拼接，故 $[Wh_i \parallel Wh_j \parallel W_e z_{ij}]$ 的维度为 $3d$ 。

$$\alpha_{ij} = \text{softmax}_j \left(\text{LeakyReLU} \left(a^T [Wh_i \parallel Wh_j \parallel W_e z_{ij}] \right) \right) \quad (7)$$

$$H_{(l+1)i} = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} Wh_j \right) \quad (8)$$

GAT 共堆叠三层，每层后加入残差连接、层归一化和 dropout。图级读出同时取平均池化和最大池化，使不同图像中花瓣数量变化时仍能得到固定长度表示。

$$h_G = \text{concat}(\text{mean_pool}(H(3)), \text{max_pool}(H(3))) \quad (9)$$

3.6. 门控融合与训练目标

视觉分支输出 f_{CNN} ，图分支输出 f_{GNN} 。二者并非简单拼接，而是先由门控单元估计当前样本更依赖哪一类证据，再送入最终分类头。

$$g = \text{softmax}(W_g \cdot \text{concat}(f_{\text{CNN}}, f_{\text{GNN}}) + b_g) \quad (10)$$

$$f_{\text{fused}} = g_1 \cdot \text{Proj}(f_{\text{CNN}}) + g_2 \cdot \text{Proj}(f_{\text{GNN}}) \quad (11)$$

$$\text{logits} = \text{MLP}(f_{\text{fused}}) \quad (12)$$

总损失由分类损失、视觉辅助损失、图分支辅助损失和分割损失组成：

$$L = L_{\text{cls}} + \lambda_1 L_{\text{CNN}} + \lambda_2 L_{\text{GNN}} + \lambda_3 L_{\text{seg}}, L_{\text{seg}} = L_{\text{mask}} + L_{\text{box}} + L_{\text{cls_inst}} \quad (13)$$

训练分三阶段进行：先训练实例分割器；再冻结视觉骨干，训练图分支和融合头；最后解冻视觉骨

干高层联合微调。第二阶段 L_{CNN} 只作用于视觉分类头, 分类梯度到第三阶段才回传到高层视觉特征。

门控融合主要处理样本之间的证据差异。花冠层次清楚、花瓣边界稳定时, 图分支提供的邻接和遮挡关系更有价值; 严重遮挡或分割失败时, 整图颜色、纹理和外轮廓仍然重要。 g_1 和 g_2 的学习使模型不必事先假定拓扑分支永远占优, 而是在样本层面调整两类信息的权重。

训练中, 分割任务和分类任务相对独立。实例分割器学习类别无关的花瓣区域和边界, 不直接接收牡丹类别标签; 分类头再利用视觉、图和融合特征判断类别。Oxford Flowers-102 无实例标注, 只计算分类损失, 弱伪拓扑由自动候选区域生成。

4. 实验设置

4.1. 数据集与划分

Peony-8 包含 8 个牡丹类别、5000 张图像, 按 70%/15%/15% 分层划分为训练集 3500 张、验证集 750 张和测试集 750 张。同一植株、同一采集批次以及近重复图像被放入同一子集, 避免相似图像跨集合造成指标虚高。Oxford Flowers-102 包含 102 个花卉类别、8189 张图像, 采用官方 1020/1020/6149 的训练、验证和测试划分。

Peony-8 的实例标注以花瓣为单位。标注者先确定花冠前景, 再编号并修正花瓣候选, 同时记录遮挡层次。对内层边界难以稳定辨认的重瓣图像, 以可辨识的最小连续花瓣区域作为节点。

数据划分时检查采集批次和近重复关系。同一花朵连续拍摄、裁剪高度重合或仅亮度略变的样本均归入同一子集, 避免相似图像跨集合抬高指标。

4.2. 基线与实现细节

Peony-8 上的比较对象包括 ResNet50、EfficientNet-B0、ViT-B/16、MobileNetV2+CBAM 和 PTRGNet。PTRGNet 的图分支与门控融合建立在 MobileNetV2+CBAM 上; 其他模型作为视觉基线, 用来观察仅凭图像级特征时的识别表现。Oxford Flowers-102 上复现 ResNet50、EfficientNet-B0、ViT-B/16、PTRGNet 视觉分支和 PTRGNet 弱伪拓扑分支。所有模型保持相同划分、输入分辨率和增强策略。

训练增强包括随机裁剪、水平翻转、颜色扰动和随机擦除。分类部分采用交叉熵损失, 并依据验证集早停。各基线使用相同预训练初始化和训练轮次范围, 学习率通过验证集调节。ViT-B/16 数值仅代表本划分下的复现表现。

掩码来源单独列为实验因素。Oracle mask 使用人工精修实例掩码, 只作为结构上限参考; Predicted mask 使用训练后的实例分割器自动输出, 对应实际测试流程; SAM zero-shot mask 使用未经人工修正的 SAM 候选掩码, 用来观察域偏移和分割噪声。Oxford Flowers-102 没有花瓣实例真值, 弱伪拓扑由显著性前景、SLIC/K-Means 区域、GrabCut 和 SAM 候选掩码共同生成, 同时保留纯视觉分支作为直接参照。

关于实例分割器泛化, Peony-8 可利用人工掩码计算质量并比较 Oracle、Predicted 和 SAM zero-shot 条件。Oxford Flowers-102 缺少花瓣实例真值, 不能报告 Dice、AP 等有监督分割指标; 其结构分支仅反映弱伪掩码条件下的分类变化。

Oxford Flowers-102 的弱伪拓扑采用保守生成方式: 先用显著性检测分离花朵主体, 再在前景内用 SLIC 或 K-Means 形成局部区域, 并以 GrabCut 和 SAM 候选补充边界。随后按面积、长宽比、圆度、前景重叠率和边界位置过滤噪声节点。

由于 Oxford 的弱伪拓扑没有实例真值监督, 相关结果只被用来观察低成本结构线索的迁移现象。表 5 同时列出 PTRGNet 视觉分支和弱伪拓扑分支, 也是为了避免把视觉骨干本身的贡献误归到图分支。只

有两者直接相减后的部分, 才可视为伪拓扑带来的增量。

所有分类模型均在验证集选择最佳权重, 并在同一测试集报告结果。PTRGNet 先训练分割器, 再训练图分支与融合头, 最后联合微调, 使图分支先获得较稳定的节点输入。

评价指标包括 Top-1 准确率、精确率、召回率、F1 分数、Top-5 准确率、参数量、FLOPs 和单图推理时间。Peony-8 侧重 Top-1、精确率、召回率和 F1 分数; Oxford Flowers-102 同时给出 Top-1 和 Top-5; 复杂度统计统一输入分辨率、推理精度和硬件环境。

5. 结果

5.1. Peony-8 分类结果

表 1 给出 Peony-8 测试集的分类结果。PTRGNet 的 Top-1 准确率为 67.8%, 精确率、召回率和 F1 分数分别为 66.9%、67.3%和 67.1%。通用基线中, EfficientNet-B0 的表现最好, 准确率为 59.7%、F1 分数为 59.0%。按绝对差值计算, PTRGNet 比 EfficientNet-B0 高 8.1 个百分点, 比 MobileNetV2+CBAM 高 9.6 个百分点。

Table 1. Classification results on Peony-8

表 1. Peony-8 分类结果

方法	准确率/%	精确率/%	召回率/%	F1/%
ResNet50	56.3	55.8	54.9	55.3
EfficientNet-B0	59.7	58.9	59.2	59.0
ViT-B/16	58.9	58.2	58.6	58.4
MobileNetV2+CBAM	58.2	57.6	57.8	57.7
PTRGNet	67.8	66.9	67.3	67.1

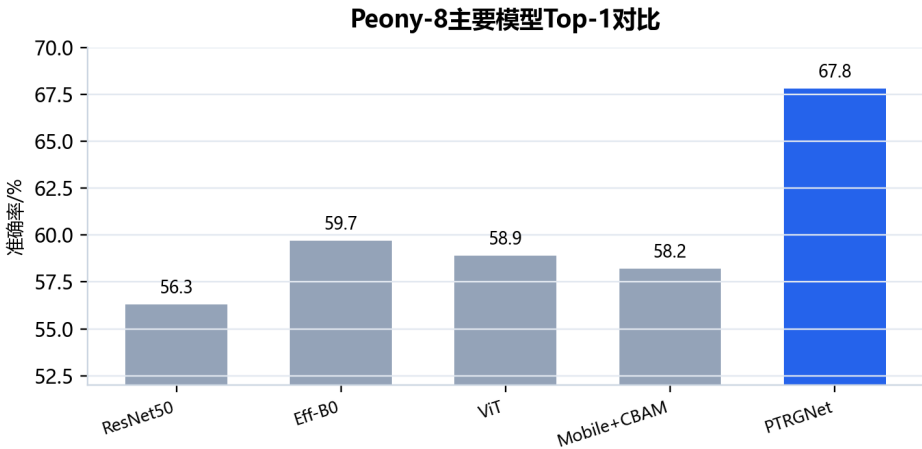


Figure 3. Top-1 comparison on Peony-8

图 3. Peony-8 主要模型 Top-1 对比

5.2. 消融实验

表 2 展示了各模块逐步加入后的变化。只保留视觉分支时, 准确率为 58.2%; 加入拓扑分支后升至 63.4%; 再引入门控融合后为 65.1%; 完整 PTRGNet 达到 67.8%。从这些数值看, 花瓣关系和融合策略都参与了最终分类。

Table 2. Ablation results of PTRGNet
表 2. PTRGNet 消融结果

模型配置	准确率/%	相对仅 CNN 提升
仅 CNN 视觉分支	58.2	-
+拓扑分支	63.4	+5.2 百分点
+门控融合	65.1	+6.9 百分点
完整 PTRGNet	67.8	+9.6 百分点

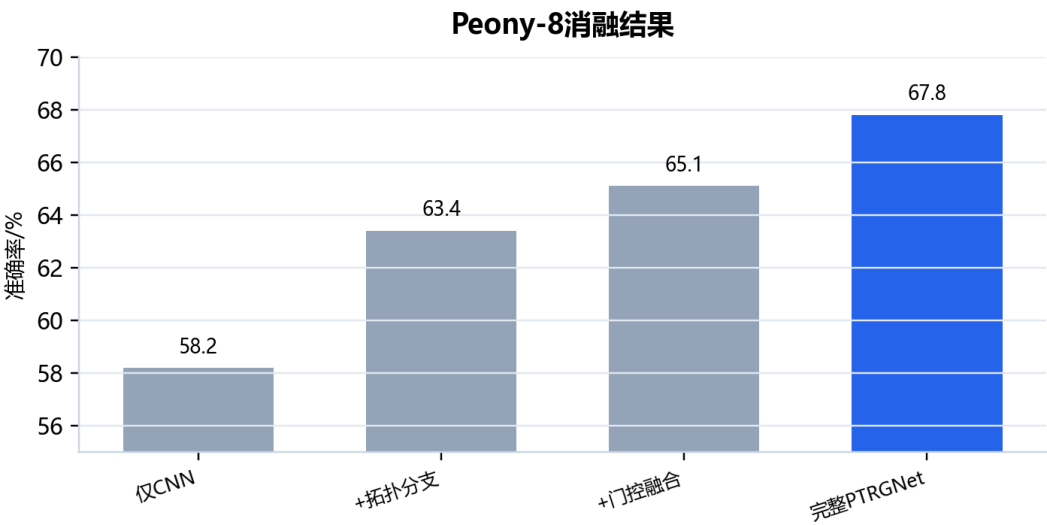


Figure 4. Ablation results of PTRGNet
图 4. PTRGNet 消融结果

5.3. 图编码器比较

在节点特征和掩码输入相同的情况下，GAT 分支取得 67.8%准确率，并在约 22 个 epoch 后趋于稳定；GCN 和 GraphSAGE 分别为 65.2%和 64.8%。花瓣之间的邻接关系并不均匀，某些边对分类更关键，GAT 的可学习注意力更适合处理这种差异。

Table 3. Comparison of graph branch encoders
表 3. 图分支结构比较

GNN 结构	准确率/%	训练稳定性	收敛轮次
GCN	65.2	良好	25
GraphSAGE	64.8	中等	30
边感知 GAT	67.8	优秀	22

5.4. 掩码来源鲁棒性

表 4 比较了三种测试阶段掩码来源。人工精修掩码对应 70.2% Top-1 准确率，可看作节点质量较理想时的参考值；训练后分割器自动预测掩码对应 67.8%；直接使用 SAM 零样本掩码时降至 63.5%。掩码质量下降会影响图分支，但自动预测掩码条件下的完整模型仍高于仅视觉分支的 58.2%。

Table 4. Peony-8 results under different mask sources
表 4. 不同掩码来源下的 Peony-8 结果

掩码来源	Top-1 准确率/%	F1/%	说明
人工精修掩码	70.2	69.5	结构上限分析
实例分割器预测掩码	67.8	67.1	默认自动流程
SAM 零样本掩码	63.5	62.7	噪声鲁棒性分析
无掩码视觉分支	58.2	57.7	MobileNetV2+CBAM

5.5. Oxford Flowers-102 泛化结果

Oxford Flowers-102 的结果见表 5。PTRGNet 视觉分支的 Top-1 准确率为 88.6%，加入弱伪拓扑后为 89.8%，增加 1.2 个百分点。由于该数据集没有花瓣实例真值，实验未计算 Dice、AP 等有监督分割指标；这一差值更适合被理解为弱伪标签条件下的结构补充现象，而非高质量花瓣拓扑或分割器跨域泛化已经得到充分证明。

Table 5. Reproduced results on Oxford Flowers-102
表 5. Oxford Flowers-102 复现实验结果

方法	拓扑输入	Top-1 准确率/%	Top-5 准确率/%
ResNet50	无	84.2	96.8
EfficientNet-B0	无	85.9	97.3
ViT-B/16	无	86.1	97.5
PTRGNet 视觉分支	无	88.6	98.1
PTRGNet 弱伪拓扑	伪掩码	89.8	98.6

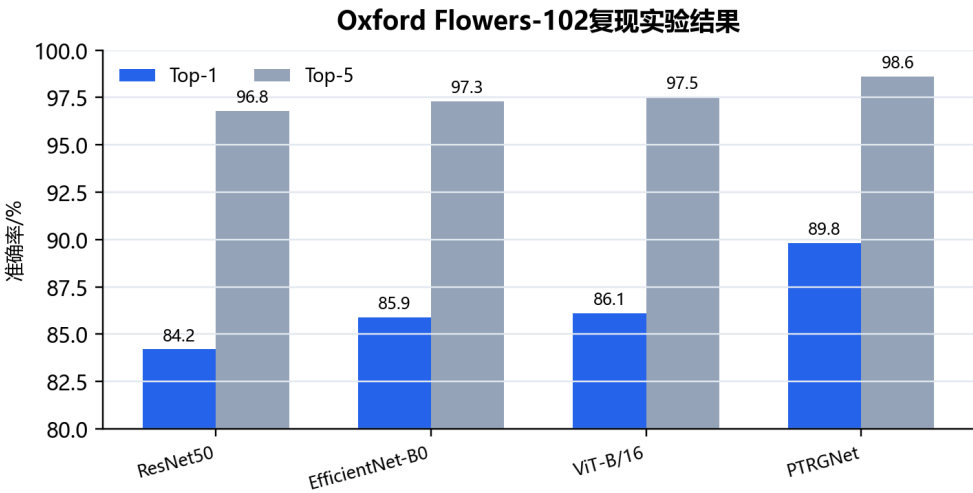


Figure 5. Reproduced results on Oxford Flowers-102
图 5. Oxford Flowers-102 复现实验结果

5.6. 复杂度分析

表 6 汇总复杂度。PTRGNet 参数量为 8.7 M，FLOPs 为 1.2 G，单图推理时间为 52 ms。它比

MobileNetV2+CBAM 和 EfficientNet-B0 更耗时，但仍低于 ViT-B/16。换言之，实例分割和图推理确实增加了代价，不过当前规模仍处在可接受范围内。

Table 6. Model complexity and inference latency
表 6. 模型复杂度与推理时间结果

方法	参数量/M	FLOPs/G	推理时间/ms
ResNet50	25.6	4.1	45
EfficientNet-B0	5.3	0.4	32
ViT-B/16	86.4	17.6	78
MobileNetV2+CBAM	3.4	0.3	28
PTRGNet	8.7	1.2	52

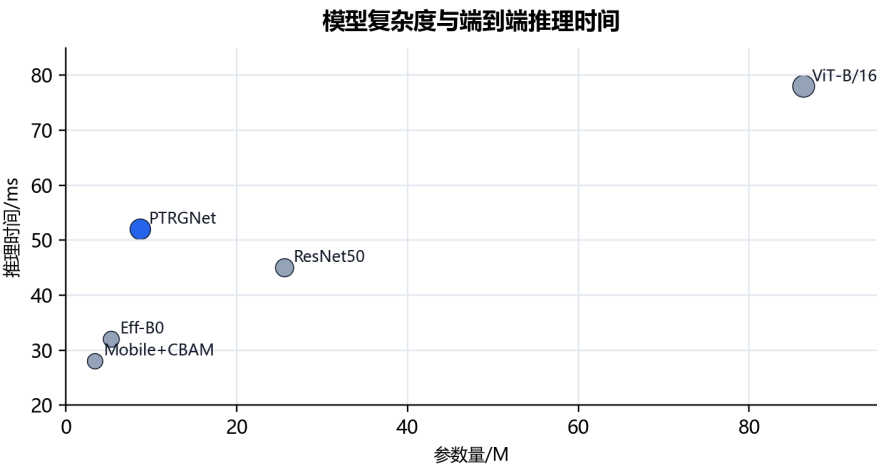


Figure 6. Complexity and inference latency comparison
图 6. 复杂度与推理时间对比

5.7. 结果分析

消融结果表明，PTRGNet 的增益来自花瓣 ROI 与边感知 GAT 的共同作用。前者减少背景干扰，后者依据距离、方向和重叠关系调整邻居信息。GAT 优于 GCN 和 GraphSAGE，也说明花瓣关系不宜用完全固定的邻域聚合处理。

掩码来源实验显示，人工精修掩码给出最高准确率；自动分割器虽有局部误差，仍能保留主要花瓣层次；SAM 零样本掩码在重叠区域易欠分割，使节点数量和边属性偏离。图分支不依赖测试阶段人工交互，但依赖稳定的自动分割器。

Oxford Flowers-102 上的提升幅度较小，也在预期之内。该数据集没有花瓣实例真值，弱伪拓扑中难免混入花蕊、叶片边缘或粘连花瓣。因而，Oxford 结果更适合作为迁移观察：在公共花卉数据上，粗结构关系可能有帮助，但其结论强度弱于 Peony-8 的实例级设置。

6. 讨论

6.1. 方法有效性

这些结果显示，把花瓣实例作为结构单元有助于牡丹图像识别。全局视觉特征主要关注颜色、纹理

和轮廓, 拓扑分支则补入花瓣相对位置、重叠关系和层次分布。两类信息对应的视觉线索不同, 与植物形态识别中“看花冠结构”的经验较为一致。边属性以特征形式参与注意力计算, 也避免了手工固定边权与 GAT 学习机制之间的矛盾。

把消融结果和复杂度结果放在一起看, PTRGNet 的提升并非单纯来自参数量增加。它的参数规模远小于 ViT-B/16, 但在 Peony-8 上优于多个视觉基线, 结构关系和区域语义的互补应是主要来源。

从模型设计看, 图网络并不是视觉骨干的替代品。视觉骨干先提供花瓣区域的语义表示, 图分支再描述这些区域之间的相对关系。牡丹花冠具有明显的层次结构, 单片花瓣的纹理有时不足以区分类别, 多个花瓣之间的排列和遮挡反而更接近植物学中的形态依据。边感知 GAT 的作用, 正是把这种关系信息传递给分类头。

固定边权通常要人为决定距离、角度、重叠和面积差各占多大比例, 这在不同花型和不同拍摄条件下很难统一。PTRGNet 保留这些几何量作为边属性, 由注意力机制根据样本和类别监督自行调整邻居贡献。该做法保留了几何含义, 也减少了经验参数带来的不确定性。

6.2. 测试阶段掩码可靠性

测试阶段掩码来源决定了方法能否自动运行。主结果采用训练后的分割器预测掩码, Oracle mask 仅作理想节点质量参考, SAM zero-shot mask 用于观察外部分割模型直接迁移的噪声影响。三组结果表明, 实例分割器是 PTRGNet 的关键组成部分。

在部署环境中, 预测掩码误差大致有三类: 欠分割、过分割和边界漂移。欠分割会把多个内层花瓣合并成一个节点, 导致图中节点偏少; 过分割会产生碎片节点, 使局部边数增加; 边界漂移会改变面积、重叠率和 ROI 语义特征。三种掩码来源并列报告后, 分割误差如何传导到图分类结果会更清楚。

6.3. 错误分析

错误样本主要有三类: 内层花瓣极度拥挤, 分割器易合并节点; 花瓣边缘与背景或同色花瓣对比度低, 边界漂移影响几何和 ROI 特征; 类别之间花型和颜色接近, 单片花瓣纹理不足以稳定区分。拓扑图能缓解这些困难, 但不能完全抵消掩码误差。

6.4. 复杂度与效率

复杂度方面, PTRGNet 比轻量视觉分支增加了实例分割、候选边筛选和 GAT 推理, 端到端时间从 28 ms 增至 52 ms。由于几何特征采用 GPU 张量化计算, 图构建未形成明显 CPU 瓶颈。若需更高实时性, 可减小 k 值或使用更轻的实例分割器。

6.5. 应用前景

PTRGNet 的附加价值在于分类证据可回溯到花瓣实例和关系边。对于植物资源调查、品种识别和人工复核, 这种结构化输出比单一类别标签更易解释; 未来也可扩展到芍药属其他类群及多器官关系建模。

在农业信息化和花卉资源管理中, 研究者常需了解模型依据的形态线索。高注意力花瓣节点、关键邻接边和重叠区域可辅助复核, 也可作为品种展示和资源库整理中的形态描述特征。

随着交互式分割和弱监督实例分割发展, 花瓣实例标注成本有望降低。若后续建设跨地点、跨年份和跨拍摄条件的开放数据集, PTRGNet 可与视觉语言模型或图基础模型结合, 扩展到更广的植物结构识别任务。

7. 结论

基于结构感知花瓣拓扑图的 PTRGNet 面向牡丹图像识别而设计。模型以花瓣实例为节点, 保留 5 维几何边属性, 并通过边感知 GAT 学习花瓣之间的关系; 门控融合模块将全局视觉特征和拓扑特征结合, 避免单一分支支配分类结果。

在 Peony-8 上, PTRGNet 取得 67.8% Top-1 准确率和 67.1% F1 分数, 比 EfficientNet-B0 高 8.1 个百分点; 在 Oxford Flowers-102 上, 弱伪拓扑设置达到 89.8% Top-1 和 98.6% Top-5。消融结果显示, 拓扑分支、门控融合和 GAT 关系学习均对最终结果有实际贡献。

不同掩码来源的对比给出了方法边界: 人工精修掩码对应较高结构上限, 自动分割器预测掩码可支撑端到端流程, 而 SAM 零样本掩码会因域偏移和欠分割导致性能下降。Oxford Flowers-102 的弱伪拓扑增益有限, 也提示公共数据上的结构解释仍受伪掩码质量约束。

总体来看, 实例级花瓣拓扑为牡丹识别提供了颜色和纹理之外的形态证据。后续工作更适合沿着三个方向推进: 扩大实例级花卉数据集, 检验跨品种和跨季节稳定性, 以及压缩实例分割与图推理开销, 使结构感知识别更接近实际部署需求。

致 谢

感谢数据标注与实验复核人员对花瓣实例掩码质量控制的支持。感谢我们的指导老师以及各位同学对本论文及实验的付出与支持。

数据与伦理声明

数据可用性声明: Peony-8 的图像来源、去重规则、训练/验证/测试划分、花瓣实例掩码和质检记录拟随补充材料公开; 若个别图像存在版权限制, 将提供可复现采集说明或替代索引。

代码可用性声明: 训练代码、模型配置、随机种子、预训练权重说明和实验日志拟在论文录用后通过作者的开源仓库公开。

伦理声明: 本研究不涉及人类受试者或动物实验。若使用网络图像, 数据发布将遵守原始来源许可, 并移除存在版权风险的样本。

利益冲突声明: 作者声明无利益冲突。

参考文献

- [1] 罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报, 2017, 43(8): 1306-1318.
- [2] Nilsback, M. and Zisserman, A. (2008) Automated Flower Classification over a Large Number of Classes. 2008 *Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, Bhubaneswar, 16-19 December 2008, 722-729. <https://doi.org/10.1109/icvgip.2008.47>
- [3] 李嘉珏. 中国牡丹与芍药[M]. 北京: 中国林业出版社, 1999.
- [4] 陈少真, 叶武剑, 刘怡俊. 基于知识蒸馏与改进 ViT 网络的花卉图像细粒度分类[J]. 光电子·激光, 2024, 35(1): 29-40.
- [5] 廖宁, 曹敏, 严骏驰. 视觉提示学习综述[J]. 计算机学报, 2024, 47(4): 790-820.
- [6] Fu, J., Zheng, H. and Mei, T. (2017) Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 4476-4484. <https://doi.org/10.1109/cvpr.2017.476>
- [7] He, J., Chen, J., Liu, S., Kortylewski, A., Yang, C., Bai, Y., et al. (2022) TransFG: A Transformer Architecture for Fine-Grained Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 852-860. <https://doi.org/10.1609/aaai.v36i1.19967>
- [8] 项伟康, 周全, 崔景程, 莫智懿, 吴晓富, 欧卫华, 王井东, 刘文予. 基于深度学习的弱监督语义分割方法综述

- [J]. 中国图象图形学报, 2024, 29(5): 1146-1168.
- [9] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/iccv.2017.322>
- [10] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
- [11] 徐冰冰, 岑科廷, 黄俊杰, 沈华伟, 程学旗. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.
- [12] 马帅, 刘建伟, 左信. 图神经网络综述[J]. 计算机研究与发展, 2022, 59(1): 47-80.
- [13] Jing, D., He, X., Luo, Y., Fei, N., Yang, G., Wei, W., et al. (2024) FineCLIP: Self-Distilled Region-Based CLIP for Better Fine-Grained Understanding. *Advances in Neural Information Processing Systems* 37, Vancouver, 10-15 December 2024, 27896-27918. <https://doi.org/10.52202/079017-0875>
- [14] Wang, Z., Zhang, Z., Chawla, N., Zhang, C. and Ye, Y. (2024) GFT: Graph Foundation Model with Transferable Tree Vocabulary. *Advances in Neural Information Processing Systems* 37, Vancouver, 10-15 December 2024, 107403-107443. <https://doi.org/10.52202/079017-3412>
- [15] Xu, D.F., Zhu, Y.K., Choy, C.B. and Li, F.F. (2017) Scene Graph Generation by Iterative Message Passing. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 3097-3106. <https://doi.org/10.1109/cvpr.2017.330>