



# 计算机科学

COMPUTER SCIENCE

## 基于多粒度特征融合的叶片分类与分级方法

刘松岳, 王欢

引用本文

刘松岳, 王欢. 基于多粒度特征融合的叶片分类与分级方法[J]. 计算机科学, 2023, 50(3): 216-222.

LIU Songyue, WANG Huan. [Leaf Classification and Ranking Method Based on Multi-granularity Feature Fusion](#) [J]. Computer Science, 2023, 50(3): 216-222.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于知识图谱的层次粒化推荐方法](#)

Hierarchical Granulation Recommendation Method Based on Knowledge Graph

计算机科学, 2022, 49(8): 64-69. <https://doi.org/10.11896/jsjcx.210600111>

### [基于移动设备的个人云服务及资源 Mashup](#)

Towards Personal Cloud Service and Resource Mashup Based on Mobile Devices

计算机科学, 2016, 43(6): 160-166. <https://doi.org/10.11896/j.issn.1002-137X.2016.06.033>

### [并行计算模型对比分析](#)

计算机科学, 2005, 32(12): 142-145.

### [Web应用界面跨浏览器兼容性的自动检测方法](#)

Automatic Detection Method of Cross-browser Web Application

计算机科学, 2015, 42(Z11): 444-449.

### [一种基于CFDs规则的修复序列快速判定方法](#)

Rapid Decision Method for Repairing Sequence Based on CFDs

计算机科学, 2018, 45(3): 311-316. <https://doi.org/10.11896/j.issn.1002-137X.2018.03.051>

# 基于多粒度特征融合的叶片分类与分级方法

刘松岳 王欢

南京理工大学计算机科学与工程学院 南京 210094

(slyusy811@163.com)

**摘要** 长期以来,已有很多工作致力于研究植物叶片分类,虽然它们在公开数据集上表现较好,但实际应用并不理想,且难以应用于更复杂的问题,如叶片分级,即要求在对叶片进行分类的基础上,再对同一类的叶片进行更细级别(质量等级)的划分。为此,提出了一种新的植物叶片分类以及分级模型,该模型关注叶片的多粒度信息,并将粗粒度与细粒度进行有效融合。该模型包含粗粒度和细粒度两个分支,由粒度混合损失将两个分支联系起来,促使模型逐步学习由粗到细的粒度表征。采用了多步骤训练方式,每一步提取不同层级的特征,实现浅层特征与深层特征的融合。此外,还提出了几何通道注意力模块,该模块由空间变换和双线性注意力池化组成,使模型可以关注图像中更具区分度的局部区域,提取出的特征更具区分性。所提方法在 Flavia leaf 和 Swedish leaf 两个公开的叶片分类数据集上分别达到了 99.8% 和 99.7% 的分类准确率,且在所构建的烟叶分级数据集上达到了 71.9% 的分级准确率,均超过了目前最优的方法。

**关键词:** 叶片分类;叶片分级;多粒度融合;空间变换网络;双线性注意力池化

中图分类号 TP391.41

## Leaf Classification and Ranking Method Based on Multi-granularity Feature Fusion

LIU Songyue and WANG Huan

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

**Abstract** Much work has long been devoted to plant leaf classification, but these methods cannot be applied well in real applications, though they may achieve good results in public datasets. Moreover, they are hardly employed to more complex problems, e. g. leaf ranking, which requires the classification of leaves first and then ranking leaves of the same class. This paper proposes a new model for plant leaf classification as well as leaf ranking, which focuses on the granularity information of leaves and integrates multi-level granularity from coarse to fine. Specifically, the model contains two branches, coarse-grained and fine-grained, which are linked by a coarse-fine hybrid loss, prompting the model to progressively learn a coarse-to-fine representation. A multi-step training approach is used, with different levels of features extracted at each step, therefore enabling the fusion of shallow features with deep features. In addition, a geometric channel attention module, which consists of a spatial transformation and a bilinear attention pooling module, is proposed to allow our model to focus on more discriminative local regions in the image and extract more discriminative features. Our method achieves 99.8% and 99.7% classification accuracy on two publicly available leaf classification datasets, Flavia leaf and Swedish leaf, respectively, and 71.9% classification accuracy on our constructed tobacco leaf ranking dataset, both outperform the state-of-the-art methods.

**Keywords** Leaf classification, Leaf ranking, Multi-granularity fusion, Spatial transformation network, Bilinear attention pooling

## 1 引言

植物在维持地球生态系统平衡方面发挥着重要作用,准确而高效率的植物鉴定在保护物种多样性、维持人与自然和谐关系、医学药物研究等领域都有十分重要的意义。因此,自动植物物种识别的研究一直以来备受关注。在植物学中,叶片是用于植物分类的重要特征,因此相关研究大多聚焦于植物叶片的识别。目前有许多基于深度学习的植物叶片识别

方法,这些方法大多采用通用的分类卷积神经网络。虽然利用足够多的训练样本可以训练出效果很好的分类器,但当这些模型被直接用于处理更复杂的任务时,例如叶片分级,将无法输出令人满意的结果。

叶片分级本质上是一个分类任务,但旨在对同一类叶片下的不同等级的叶片实例进行分类。例如同一类的茶叶、烟叶会根据叶片的颜色、气味、油分等再次划分等级。由于需要分级的叶片属于同一类,因此它们往往在外观上高度相似。

到稿日期:2021-11-19 返修日期:2022-05-31

基金项目:国家自然科学基金(61703209)

This work was supported by the National Natural Science Foundation of China(61703209).

通信作者:王欢(wanghuanphd@njust.edu.cn)

显然,叶片分类和分级都需要更多的粒度信息,而分级比分类需要考虑更多的粒度。然而,很少有方法关注叶片分类中多粒度的使用,尤其是叶片分级任务。我们认为,提取更细粒度的特征,可以有效提高叶片分类的准确率,且细粒度特征对于叶片分级任务来说是必要的,甚至需要多粒度的融合,才能在分级任务中取得不错的效果。

鉴于此,本文充分考虑使用多种粒度的优势,提出了一种用于叶片分类和分级的模型架构,其可以有效地进行多粒度的融合。受到 PAN<sup>[1]</sup> 以及 PMG<sup>[2]</sup> 的启发,所提模型由粗粒度和细粒度两个分支组成,且网络训练上采用多步骤的渐进训练方式。通过这样的设计,网络可以将不同粒度的特征进行结合,逐步关注具有区分度的叶片区域。相比 PAN 和 PMG,本文模型同时在网络结构和训练方法上实现了多粒度的融合,能够提取更精细的特征。

此外,模型的细粒度分支中还包含了一个几何通道注意力模块,用于对提取出的特征做进一步处理。该模块让网络聚焦于图像的局部细微区域,提取出的特征更利于叶片分级这样的细粒度分类任务。最后,我们构建了一个烟叶分级数据集,用于验证本文模型在叶片分级任务上的有效性。

综上,本文的主要贡献如下:

- (1) 提出了同时用于叶片分类与分级的多粒度融合网络,该网络融合粗细粒度的特征,可以有效区分高度相似的叶片。
- (2) 构建了用于叶片分级的烟叶数据集,以促进细粒度和多粒度识别等相关技术在叶片分类和分级任务中的深入研究。数据集以及代码近期将在 Github<sup>1)</sup> 上发布。
- (3) 对于叶片分类和分级两种任务,所提模型分别在公开的植物叶片数据集和构建的叶片分级数据集上都达到了目前的最高水平。

本文第 2 节介绍了相关工作;第 3 节对所提方法进行了详细说明;第 4 节给出了本文方法与其他方法性能比较和消融研究的实验结果;最后总结全文。

## 2 相关工作

本节先对已有的叶片分类与分级方法做简要概述,再对细粒度和多粒度分类的相关工作进行阐述。由于传统的分类分级方法主要是提取手工设计的特征,与本文方法有较大区别,因此本文只对基于深度学习的方法做简要概述。

### 2.1 叶片分类

近年来,鉴于神经网络强大的特征表征能力,不断有方法将深度学习用于叶片分类。Shah 等<sup>[3]</sup> 提出了一种双路径 CNN,其拥有两个分支,分别用于提取形状特征和纹理特征。这样提取出的特征就包含了具有全局信息的形状特征和局部信息的纹理特征。Lee 等<sup>[4]</sup> 同样采用了双分支的结构,但与 dual-path CNN 不同的是,他们讨论了提取出的形状、纹理这两个特征的不同融合方法。此外, Lee 等<sup>[5]</sup> 还提出了一种混合的多器官植物识别网络 HGO-CNN,该网络同样包含两个分支,分别提取全局特征和局部特征。并且其还将输入图像缩放成 3 个不同的尺寸,分别输入进网络,促使网络发掘更

细微的特征。Quach 等<sup>[6]</sup> 将传统方法与深度学习相结合,使用神经网络将提取出的特征转换为更好的表示,并使用 SVM 来进行叶片的分类。Han 等<sup>[7]</sup> 也将传统特征融入深度学习方法中,提出了多特征融合的叶片识别方法,他们提取叶片的 LBP 以及 Gabor 特征,并将二者融合输入进卷积神经网络进行叶片的识别。

### 2.2 叶片分级

利用深度学习来解决叶片分级的研究相对较少,且大多数工作都采用结构较为简单的卷积神经网络的问题。例如, Dasari 等<sup>[8]</sup> 设计了对烟叶进行分级的卷积神经网络,该网络框架除输入输出层之外,主要包含 3 层卷积层、3 层池化层以及 1 层全连接层。Marzan 等<sup>[9]</sup> 利用 Haar 级联分类器来检测烟叶图像,然后利用卷积神经网络对烟叶图像进行模型训练和测试。这些网络虽然取得了不错的分级效果,但我们认为这是训练测试数据的分布较为理想,且待识别的等级数较少的缘故。当数据分布复杂时,这种简单结构的网络不足以提取有足够区分度的细粒度和多粒度特征。

### 2.3 细粒度和多粒度分类方法

细粒度分类旨在识别给定类别下的子类,这要求网络提取出更具区分性的局部细微特征。Lin 等<sup>[10]</sup> 提出了一种双线性结构,通过两个独立的 CNN 计算成对特征交互,以捕获图像的局部信息。Xu 等<sup>[11]</sup> 提出了一种多注意力金字塔网络 APN,它将多尺度、多粒度的特征进行融合,被用于中草药的分级。这也让我们思考,是否可以利用多粒度融合机制来进行分级。而近期,一些工作也将多粒度融合用在细粒度分类中。例如, Du 等<sup>[2]</sup> 采用一种渐进式训练策略 PMG 和拼图生成机制来训练网络。他们利用拼图生成器来生成包含不同粒度级别信息的图像,而用 PMG 的训练策略使网络逐步学习从小粒度到大粒度的特征,实现了多粒度的融合。Wang 等<sup>[1]</sup> 研究细粒度领域的域适应问题,为图像设计了一种层次标签结构,且提出了一种双分支网络 PAN 来逐步挖掘从粗到细的特征,也实现了多粒度的融合。

上述方法中, PMG 和 PAN 采用了不同的方法来实现多粒度的融合, PMG 采用渐进的多步训练策略,在训练方法上实现多粒度融合。而 PAN 则是由粗细粒度两个分支组成,在网络结构上实现多粒度融合。但是 PAN 的分支仅是普通的卷积神经网络,无法挖掘细微的特征,而 PMG 的网络结构也不能很好地利用叶片等级与类别间的层级关系。本文针对这些问题,提出了多粒度特征融合的叶片分类与分级方法。

## 3 多粒度特征融合的叶片分类分级方法

### 3.1 问题定义

叶片分类任务中,给定数据集  $\mathcal{Q}_a = \{(x_1, y_1), \dots, (x_A, y_A)\}$ , 其中  $y_i$  为样本  $x_i$  的标签,  $y_i \in \{y^n \mid n=1, \dots, N_i\}$ ,  $N_i$  为叶片类别数。而与叶片分类任务不同,在叶片分级任务中,数据集由三元组构成  $\mathcal{Q}_b = \{(x_1, y_1^l, y_1^r), \dots, (x_A, y_A^l, y_A^r)\}$ , 其中,  $y_j^l \in \{y_j^n \mid n=1, \dots, N_j^l\}$  和  $y_j^r \in \{y_j^n \mid n=1, \dots, N_j^r\}$  分别为样本  $x_i$  的等级标签和

<sup>1)</sup> <https://github.com/wanghuanphd/leaf-classification-and-ranking>

组标签,  $N_f$  和  $N_c$  分别为总的级别数和组数。以烟叶分级为例, 根据烟叶生长部位可以将烟叶分为上部组(B)、中部组(C)和下部组(X), 其次根据成熟度、油分等品质因素将每个大组的烟叶划分为3~4级。若一片烟叶等级为B2F, 则表示其为上部组的第二级, 那么其组标签为B, 等级标签为B2F。一个组至少包含一个级别, 因此有  $N_f \geq N_c$ 。

相比叶片分类, 分级问题中样本多了一级标签, 而分级的关键也在于如何更好地利用粗粒度标签和细粒度标签间的层级关系。为了使分类任务和分级任务相统一, 本文按照叶片外观对实验中的叶片分类数据集进行分组, 外观相似的叶片为一组。有了人为添加的组标签, 叶片分类任务就能够按照分级任务处理。因此, 下文对这两种任务不加区分, 统一用分级任务进行描述, 下文所述的级别对应到分类任务中就是叶片的类别。

本节先介绍叶片分级的网络架构、粒度融合损失函数, 之后说明多步骤渐进训练方法, 最后阐述所提出的几何通道注意力模块。

### 3.2 网络架构

图1给出了叶片分级的网络架构, 其包含粗粒度和细粒度两个分支。图1的最上方是粗粒度分支, 它以叶片的组标签为监督信息, 促使网络学习叶片分组。其由一个特征提取器  $G$  和一个  $N_c$  分类器  $C$  组成。  $G$  为通用的主干特征提取网络, 例如 ResNet<sup>[12]</sup>,  $C$  为包含批量归一化以及 Elu 激活函数的全连接分类模块。图1的中间为细粒度分支, 其由多个阶段组成, 而每个阶段包含一系列的 ResNet 残差模块。图1的下方给出了阶段2的结构。此外, 细粒度分支还包含了几何通道注意力模块, 用于对提取出的特征进行进一步处理, 使特征更具区分性。最后, 细粒度分支的每个阶段同样包含了分类模块, 用于输出预测叶片级别的概率向量。

训练时, 网络粗粒度和细粒度分支的信息将通过下文所述的粒度融合损失函数融合在一起。通过多阶段、多步骤的渐进学习方式, 网络将先学习粗粒度信息, 而后逐步由粗粒度过渡至细粒度, 挖掘细粒度的信息来进行组内的叶片分级。

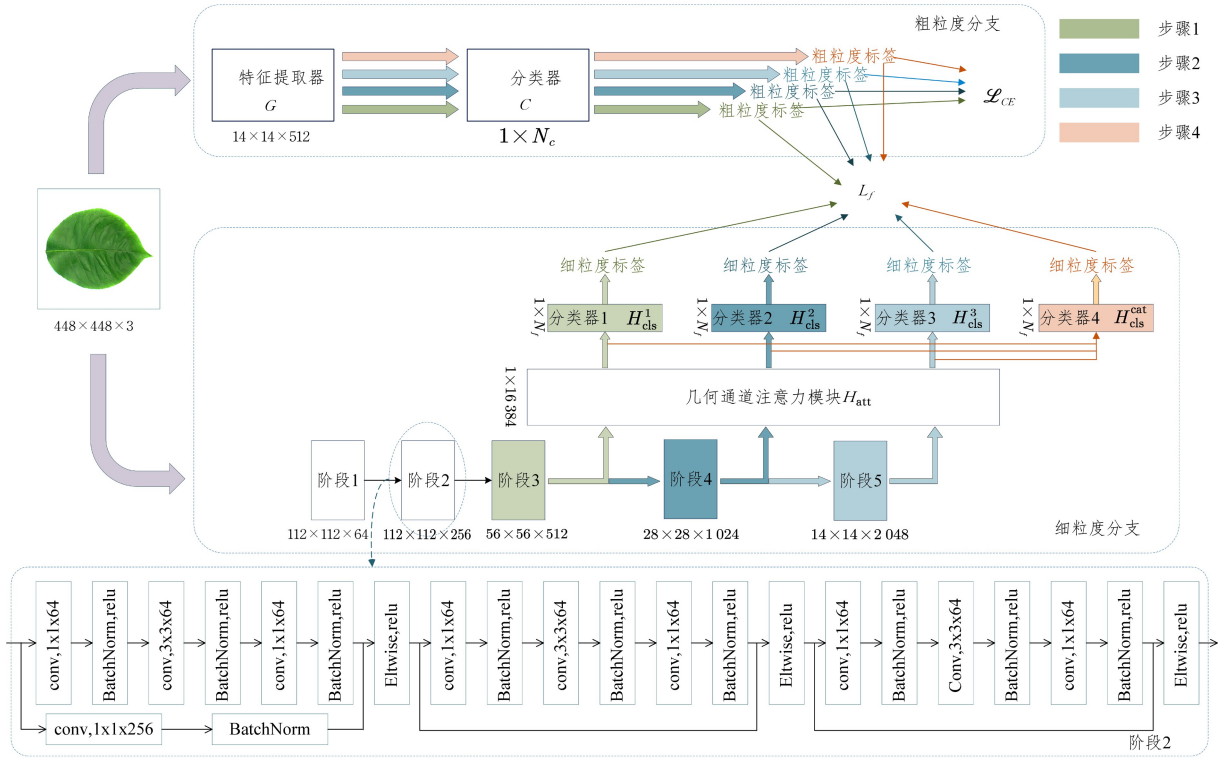


图1 多粒度特征融合网络架构(电子版为彩图)

Fig. 1 Architecture of Multi-granularity feature fusion network

### 3.3 多粒度融合损失函数

网络由粗粒度分支和细粒度分支两部分组成, 其中粗粒度分支对叶片的粗粒度标签进行预测。对于输入的图像  $x$ , 粗粒度分支输出  $x$  的粗粒度概率分布  $\hat{y}_c = C(G(x))$ , 其通过最小化交叉熵目标函数  $\mathcal{L}_{CE}(\hat{y}_c, y_c)$  进行训练, 其中,  $y_c$  为粗粒度标签的真实值,  $\mathcal{L}_{CE}$  为交叉熵损失函数。

细粒度分支用于预测叶片的等级, 即细粒度标签。  $F$  表示细粒度分支的主干特征提取网络,  $F$  包含  $K$  个阶段(每个阶段由一些卷积、池化等运算构成), 考虑后  $S$  个阶段, 每个阶段可以输出特征  $f^l \in \mathbb{R}^{H_l \times W_l \times C_l}$ , 其中  $H_l, W_l, C_l$  分别表示第  $l$  阶段的特征图的高、宽和通道数量, 其中  $l = (K - S +$

$1, K - S, \dots, K)$ 。在每个阶段后再引入几何通道注意力模块  $H_{att}$  (见 3.5 节), 其将  $f^l$  作为输入, 输出为一个向量  $v^l = H_{att}(f^l)$  然后, 每个注意力模块后跟随着一个包含了批量归一化和 Elu 非线性函数的全连接层的分类模块  $H_{cls}^l$ , 用于预测叶片等级的概率分布  $\hat{y}_f^l = H_{cls}^l(v^l)$ 。

这样,  $x$  经过两组分支, 可以得到一个粗粒度预测分布  $\hat{y}_c$  以及  $S$  个细粒度预测分布  $\hat{y}_f^l, l = (K - S + 1, K - S, \dots, K)$ 。对于每一个阶段对应的  $\hat{y}_f^l$ , 使用粗细粒度混合损失函数来训练网络。

$$L_f(\hat{y}_c, \hat{y}_f, y_f) = \mathcal{D}_{KL}((\epsilon y_f + (1 - \epsilon) \cdot \hat{y}_c^{ext}) \| \hat{y}_f) \quad (1)$$

其中,  $y_f$  为细粒度标签的真实值,  $\mathcal{Q}_{KL}$  为 KL 散度,  $\hat{y}_c^{ext}$  为  $\hat{y}_c$  将其维度扩展到与  $y_f$  一致后的概率分布。  $\epsilon$  调整细粒度真实标签和粗粒度预测分布的混合程度, 其随着迭代的增加, 逐渐从 0 上升到 1, 进而促使网络逐步由粗粒度转到细粒度的学习。  $\epsilon$  满足式(2):

$$\epsilon = \frac{1 - \exp(-10\rho)}{1 + \exp(-10\rho)} \quad (2)$$

其中,  $\rho$  是当前迭代轮次与总的迭代次数的比值。由于  $\hat{y}_c \in \mathbb{R}^{N_c}$ ,  $y_f \in \mathbb{R}^{N_f}$ ,  $N_f \geq N_c$ , 因此需要将  $\hat{y}_c$  的维度扩展到与  $y_f$  一致者才能相加。不失一般性, 粗粒度分支预测样本属于第  $i$  组的概率为  $\hat{y}_c(i)$ , 假设共有  $n$  个等级属于第  $i$  组, 记为  $j_1, j_2, \dots, j_n$ 。将预测的组别概率值赋值给属于该组的所有级别的相应位置上, 即  $\hat{y}_c^{ext}(j_1) = \hat{y}_c^{ext}(j_2) = \dots = \hat{y}_c^{ext}(j_n)$ , 这样即可实现维度的扩展。

### 3.4 多步骤训练

网络采用多步骤训练的策略<sup>[2]</sup>来进行训练, 图 1 给出了阶段数为 5、步骤数为 4 的设置下模型的训练过程。为了直观地展现, 图 1 中用不同的颜色来区分不同的步骤。多步骤训练的过程如下: 每一次迭代, 一个批次的数据  $X$  将会分 4 步输入网络, 进行 4 次计算。第 1 步从阶段 1 到阶段 3, 输出为分类器  $H_{cls}^3$  的输出, 第 2 步从阶段 1 到阶段 4, 输出为分类器  $H_{cls}^4$  的输出, 以此类推。在第  $i$  步中, 细粒度分支的输出对应第  $K-S+i$  阶段的输出  $\hat{y}_f^{K-S+i}$ , 网络参数由最小化损失函数  $L$  来更新:

$$L(\hat{y}_c^i, y_c, \hat{y}_f^{K-S+i}, y_f) = \alpha \mathcal{L}_{CE}(\hat{y}_c^i, y_c) + \beta L_f(\hat{y}_c^i, \hat{y}_f^{K-S+i}, y_f) \quad (3)$$

其中,  $L$  为  $\mathcal{L}_{CE}$  和  $L_f$  的加权和,  $\alpha$  和  $\beta$  为权重参数,  $\hat{y}_c^i$  为粗粒度分支在第  $i$  步的输出。

此外, 在每轮迭代的最后一个步骤中, 把所有  $S$  个阶段的输出拼接起来, 进而得到  $\mathbf{v}^{cat} = \text{concat}(\mathbf{v}^{K-S+1}, \mathbf{v}^{K-S}, \dots, \mathbf{v}^K)$ 。同样, 其后有一个分类模块用于输出细粒度分布  $\hat{y}_f^{cat} = H_{cls}^{cat}(\mathbf{v}^{cat})$ 。  $\hat{y}_f^{cat}$  即为细粒度分支最后一步的输出。

为了验证该训练方式的有效性, 我们将 Grad-CAM<sup>[13]</sup> 应用于 3 个注意力模块中的卷积层, 图 2(a)~图 2(c) 是 3 个卷积层的热力图, 分别对应细粒度分支的第 3~5 阶段的输出。图 2(a) 中有较高响应的主要是叶片边缘和内部的一些线条。相比图 2(a), 图 2(b) 中网络关注的位置集中在叶片内, 而在图 2(c) 中, 网络尤其关注叶片上部这一特定区域, 这一区域也是图 2(b) 中没有体现的, 在网络更深层被发掘出的具有显著区分度的区域。这表明, 随着层数的增多, 模型关注的特征由粗到细, 由一般的形状特征到更具体的语义特征。这也说明了多步骤训练以及最后将多个阶段的特征联合起来的意义所在。

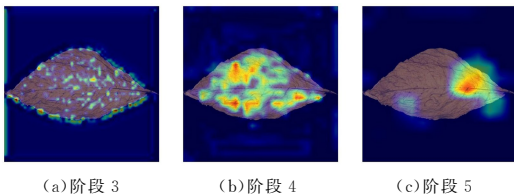
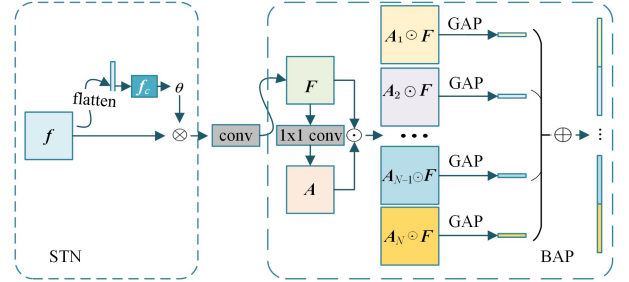


图 2 各阶段卷积层可视化

Fig. 2 Visualization of convolutional layers at each stage

### 3.5 几何通道注意力模块

几何通道注意力模块由空间变换网络(Spatial Transform Network, STN)<sup>[14]</sup>和双线性注意力池化(Bilinear Attention Pooling, BAP)<sup>[15]</sup>构成, 如图 3 所示。STN 将原始图片中的空间信息变换到另一个空间, 并保留关键信息; BAP 通过全局平均池化压缩空间维度的信息, 而保留了通道信息。几何通道注意力模块同时提供了空间注意力以及通道注意力机制, 使提取出的特征更利于分级。



注:  $\otimes$  表示仿射变换,  $\odot$  表示逐元素点积,  $\oplus$  表示向量间的拼接操作

图 3 几何通道注意力模块

Fig. 3 Geometric channel attention module

几何通道注意力模块的具体计算如下: 考虑细粒度分支中某个阶段的输出  $f \in \mathbb{R}^{H \times W \times C}$ , 将其张成向量后送入一个全连接层, 输出变换参数  $\theta \in \mathbb{R}^{2 \times 3}$ , 利用  $\theta$  构建仿射变换矩阵, 并对  $f$  进行仿射变换, 即可得到经过空间变换的特征  $f^* = \text{affine}(f, \theta)$ 。  $f^*$  经过一层卷积层后得到  $F \in \mathbb{R}^{H \times W \times M}$ 。  $A \in \mathbb{R}^{H \times W \times N}$  表示图像  $x$  的注意力图, 其通过  $F$  得出:

$$A = \text{Conv}(F) = \bigcup_{k=1}^N A_k \quad (4)$$

其中,  $\text{Conv}(\cdot)$  为  $1 \times 1$  的卷积, 之后将  $F$  与每个  $A_k$  做逐元素点积, 得到  $N$  组特征图  $F_k \in \mathbb{R}^{H \times W \times M}$ :

$$F_k = A_k \odot F, k=1, 2, \dots, N \quad (5)$$

其中,  $\odot$  表示两个张量间的逐元素点积运算。之后对每个特征图  $F_k$  通过全局平均池化(Global Average Pooling, GAP)来得到  $N$  个向量:

$$a_k = \text{GAP}(F_k), k=1, 2, \dots, N \quad (6)$$

将  $N$  个向量堆叠在一起, 得到特征矩阵  $P = (a_1, a_2, \dots, a_N)^T \in \mathbb{R}^{N \times M}$ 。最后把矩阵  $P$  张成一个向量, 记为双线性向量  $p$ , 再分别对  $p$  计算符号平方根和  $L_2$  归一化, 进而得到注意力模块输出的特征向量  $v \in \mathbb{R}^{MN \times 1}$ :

$$v' = \text{sign}(p) \sqrt{|p|} \quad (7)$$

$$v = v' / \|v'\|_2 \quad (8)$$

在得到特征向量  $v$  后, 将其输入全连接层的分类器模块, 便可得到预测级别的概率分布。而在推理时, 无需粗粒度分支, 只需细粒度分支最后步骤的联合输出  $\hat{y}_f^{cat}$  即可。因此, 除了最后步骤所需的分类器  $H_{cls}^5$ , 其余的分类器也无须使用。这样, 推理阶段网络输出的预测级别为:

$$O = \arg \max(\hat{y}_f^{cat}) \quad (9)$$

## 4 实验结果及讨论

实验采用 3 个数据集来检验本文方法的有效性。对于叶片分类问题, 我们在 Flavia leaf<sup>[16]</sup> 与 Swedish leaf<sup>[17]</sup> 上进行实验, 进一步地, 选用烟叶数据集来测试本文模型的叶片分级效果。4.1 节介绍了各数据集的具体细节; 4.2 节阐述了实验的

设置,提供了本文方法与其他一些方法的比较结果;为了说明本文方法设计的优点,4.3节给出了消融实验结果。

#### 4.1 数据集

Flavia leaf 和 Swedish leaf 是实验中进行叶片分类的两个数据集。其中,Flavia leaf 共包含 1907 张叶片图像,共有 32 类,每一类包含 50~70 张叶片图像;Swedish leaf 共包含 15 类叶片的 1125 张图像,每一类 75 张图像。

对于这两组数据集,我们把外观相似的叶片划分为一组,两个数据集划分后各有 5 组。图 4(a)、图 4(b)分别为两个数据集的部分样本,每一类展示一张叶片,同一个虚线框内的叶片属于同一组。实验中按照 6:2:2 的比例随机划分 10 组训练集、验证集以及测试集,取这 10 组的平均测试集分类准确率作为最终的分分类准确率。

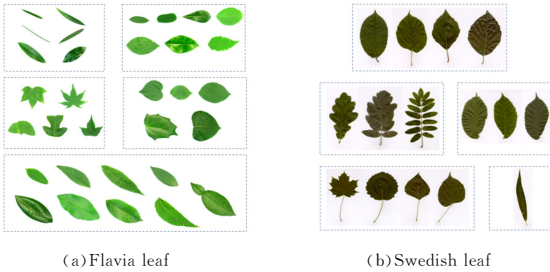


图 4 叶片分类数据集样本图示

Fig. 4 Samples of leaf classification datasets

实验中选用烟叶分级任务来检验本文方法在叶片分级上的有效性。共采集了烟草上部和中部 5 个等级的烟叶,总共 2680 张。采集的烟叶共有 B、C 两组,其中 B 组烟叶有 2 个等级,分别为 B2F、B3F,C 组烟叶有 3 个等级,分别为 C2F、C3F 和 C4F,故  $N_c=2, N_f=5$ 。烟叶样本如图 5 所示。将这 2680 张图像随机打乱后来划分数据集,包含 1488 张训练集图像,691 张验证集图像,501 张测试集图像。

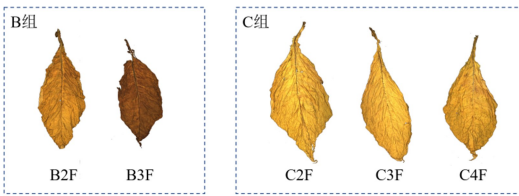


图 5 烟叶数据集样本图示

Fig. 5 Leaf samples of tobacco dataset

#### 4.2 实验设置及与其他方法的比较

ResNet18 和 ResNet50 分别为粗粒度分支和细粒度分支的主干网络。训练步骤数设为 4,几何通道注意力模块中,注意力的通道数为 32(本文在第 4.3 节对步骤数以及通道数的设置做了详细讨论,此处设置的是实验最佳值)。式(3)中,  $\alpha=\beta=0.5$ 。为了与其余分级方法进行公平比较,实验按照文献[2]的标准进行参数的设置:输入的图像首先调整为  $550 \times 550$ ,然后随机裁剪成  $448 \times 448$ ,训练阶段采用随机水平翻转作为数据增强。测试过程中,输入图像调整为  $550 \times 550$ ,之后中心裁剪为  $448 \times 448$ 。

与文献[2]相同,实验中使用随机梯度下降(SGD)作为优化器,动量为 0.9,权重衰减为 0.0005,批大小为 16,共迭代

训练 100 次。此外,为了减缓模型在初始阶段提前过拟合的现象,保持模型深层的稳定性,采用学习率预热加上余弦退火作为学习率优化策略,前 20 次迭代学习率线性增加到 0.001,后 80 次迭代学习率按照余弦函数值进行衰减。我们使用 PyTorch 在 RTX 3090 上进行所有的实验。

表 1 列出了本文方法与其他方法在 Flavia leaf 和 Swedish leaf 上的分类准确率的比较结果。除了深度学习方法,本文还选取了 3 个基于传统方法的叶片分类模型来进行比较,分别为基于形状特征的 MDM<sup>[18]</sup>和 MFD<sup>[19]</sup>,以及基于形状和纹理特征的 MTD+LBP-HF<sup>[20]</sup>。由表 1 可以看出,本文方法在两组数据集上均取得了最好的结果。在传统方法中,MTD+LBP-HF 取得了最高的准确率,这是因为它将形状和纹理特征进行结合,特征更复杂也更具有代表性。本文方法借由多步骤训练融合了不同层级的特征,充分挖掘了叶片的细粒度区域,因而取得了最好的结果。

表 1 各叶片分类方法的准确率

Table 1 Accuracy of each leaf classification

方法	Flavia leaf	Swedish leaf
MDM <sup>[18]</sup>	82.6	93.6
MFD <sup>[19]</sup>	89.5	97.6
MTD+LBP-HF <sup>[20]</sup>	99.2	98.5
VGG16 <sup>[21]</sup>	93.5	95.8
Dual-path CNN <sup>[3]</sup>	99.3	96.3
Deep-Plant <sup>[4]</sup>	98.2	97.5
HGO-CNN <sup>[5]</sup>	97.5	96.8
Leaf-CNN <sup>[6]</sup>	99.6	99.6
本文方法	99.8	99.7

表 2 列出了本文方法与其他的一些方法在烟叶分级问题上的效果比较。由于目前很少有工作在分级问题上进行研究测试,因此实验中选用了一些具有代表性的细粒度分类网络以及表 1 中的部分叶片分类网络在烟叶数据集上进行测试。

表 2 各叶片分级方法的准确率

Table 4 Accuracy of each leaf ranking method

方法类型	方法	准确率/%
通用分类网络	VGG16 <sup>[21]</sup>	55.9
	ResNet50 <sup>[12]</sup>	65.0
叶片分类网络	Dual-path CNN <sup>[3]</sup>	55.1
	Deep-Plant <sup>[4]</sup>	57.5
	HGO-CNN <sup>[5]</sup>	66.1
	Leaf-CNN <sup>[6]</sup>	55.5
	B-CNN <sup>[10]</sup>	65.1
细粒度分类网络	WSDAN <sup>[15]</sup>	64.7
	PMG <sup>[2]</sup>	67.1
	TransFG <sup>[22]</sup>	66.7
细粒度叶片分级网络	本文方法	71.9

与其他方法相比,本文方法在烟叶数据集上的分级准确率也达到了最高。细粒度分类网络的效果普遍好于叶片分类网络,这是因为叶片分级这个任务相比普通的叶片分类需要更多的粒度信息。而本文方法在已有的细粒度分支上还增加了粗粒度分支,考虑了不同粒度的融合,因而取得了最好的结果。

### 4.3 消融实验

本文方法增加额外的粗粒度分支来进行不同粒度的混合,并采用多步骤训练方式。表3列出了二者不同组合下的分级准确率,实验结果表明将二者结合可以得到最佳的结果。

表3 粗粒度分支、多步骤训练及二者组合的准确率比较

Table 3 Accuracy comparison of coarse-grained branch, multi-step training and a combination of them

粗粒度分支	多步骤训练	准确率/%
		65.7
✓		66.7
	✓	69.5
✓	✓	71.9

本文还进行了一些实验来体现注意力模块的必要性。表4列出了只使用卷积模块进行烟叶分级的准确率,以及分别加入STN, BAP后的准确率的变化情况。

表4 几何通道注意力模块中各成分不同组合下准确率的比较

Table 4 Accuracy comparison of different combinations of components of geometric channel attention module

Conv	STN	BAP	准确率/%
✓			67.1
✓	✓		70.9
✓		✓	71.1
✓	✓	✓	71.9

我们也探讨了BAP中注意力图通道数 $N$ 对最后结果的影响。表5列出了不同取值的 $N$ 对应的分级准确率情况,可以看到,32是一个较为理想的取值。

表5 不同注意力图数量对应的准确率比较

Table 5 Accuracy comparison of different numbers of attention maps

$N$	准确率/%
8	70.3
16	69.3
32	71.9
64	69.0
128	70.7

此外,对于细粒度分支训练时步骤数量的选取,本文也进行了实验。表6列出了不同的步骤数量下对应的分级准确率情况。由于当阶段数大于1时,会有额外的一步使用联合特征,因此步骤数会比阶段数多1。可以看到,随着步骤数量的增多,分级准确率一开始呈现上升的趋势,但是当步骤数量大于4时,准确率开始下降。可能是前几个阶段关注的是一些级别无关的特征,而当步骤数量大于4时,这些无关特征的加入影响了网络的学习。

表6 不同训练步骤数对应的准确率比较

Table 6 Accuracy comparison of different numbers of training steps

步骤数	对应阶段	准确率/%
1	{5}	66.7
3	{4,5}	70.7
4	{3,4,5}	71.9
5	{2,3,4,5}	69.1
6	{1,2,3,4,5}	67.5

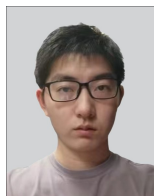
方法。首先,提出的网络包含粗粒度和细粒度两个分支,粗粒度分支对叶片进行分组,而细粒度分支的任务逐渐由分组过渡到分级,对应提取出的特征由粗到细。此外,网络使用了多步骤的策略进行训练,可以融合不同层次的特征。最后,网络还包含了一个几何通道注意力模块,它可以让网络关注图像中的局部区域。所提方法可以进行端到端的训练,除了类别标签无需其他的人工标注。不只是叶片分级,只要是满足层级结构的分类任务,网络模型只需要一些细微的调整就可以使用。实验结果表明,所提方法在两个公开的叶片分类数据集和构建的烟叶分级数据集上都取得了最佳的结果,验证了本文方法的有效性。

### 参考文献

- [1] WANG S, CHEN X, WANG Y, et al. Progressive Adversarial Networks for Fine-Grained Domain Adaptation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, IEEE, 2020: 9213-9222.
- [2] DU R, CHANG D, BHUNIA A K, et al. Fine-Grained Visual Classification via Progressive Multi-Granularity Training of Jigsaw Patches[C]// European Conference on Computer Vision. Cham: Springer, 2020: 153-168.
- [3] SHAH M P, SINGHA S, AWATE S P. Leaf classification using marginalized shape context and shape + texture dual-path deep convolutional neural network[C]// 2017 IEEE International Conference on Image Processing. New York: IEEE, 2017: 860-864.
- [4] LEE S H, CHAN C S, MAYO S J, et al. How deep learning extracts and learns leaf features for plant classification[J]. Pattern Recognition, 2017, 71: 1-13.
- [5] LEE S H, CHAN C S, MAYO S J. Multi-organ plant classification based on convolutional and recurrent neural networks[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4287-4301.
- [6] QUACH B M, CUONG D V, PHAM N, et al. Leaf Recognition Using Convolutional Neural Networks Based Features[EB/OL]. (2021-08-04) [2021-09-10]. <https://arxiv.org/abs/2108.01808>.
- [7] HAN B, ZENG S W. Plant Leaf Image Recognition Based on Multi-feature Integration and Convolutional Neural Network[J]. Computer Science, 2021, 48(6A): 113-117.
- [8] DASARI S K, PRASAD V. A novel and proposed comprehensive methodology using deep convolutional neural networks for flue cured tobacco leaves classification[J]. International Journal of Information Technology, 2019, 11(1): 107-117.
- [9] MARZAN C S, RUIZ C R. Automated tobacco grading using image processing techniques and a convolutional neural network[J]. International Journal of Machine Learning and Computing, 2019, 9(6): 807.
- [10] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear cnn models for fine-grained visual recognition[C]// Proceedings of the IEEE International Conference on Computer Vision. Washington DC:

**结束语** 本文提出了一种多粒度融合的叶片分类与分级

- IEEE, 2015:1449-1457.
- [11] XU Y, WEN G, HU Y, et al. Multiple attentional pyramid networks for Chinese herbal recognition[J]. Pattern Recognition, 2021, 110:107558.
- [12] HE K, ZHANG X, REN S, et al. Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas; IEEE, 2016: 770-778.
- [13] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization[J]. International Journal of Computer Vision, 2020, 128(2):336-359.
- [14] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems. Quebec; MIT Press, 2015:2017-2025.
- [15] HU T, QI H, HUANG Q, et al. See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification[EB/OL]. (2019-01-26) [2019-03-23]. <https://arxiv.org/abs/1901.09891>.
- [16] WU S G, BAO F S, XU E Y, et al. A leaf recognition algorithm for plant classification using probabilistic neural network[C]//2007 IEEE International Symposium on Signal Processing and Information Technology. Giza; IEEE, 2007:11-16.
- [17] SÖDERKVIST O. Computer vision classification of leaves from Swedish trees[D]. Linköping; Linköping University, 2001.
- [18] HU R, JIA W, LING H, et al. Multiscale distance matrix for fast plant leaf recognition[J]. IEEE Transactions on Image Processing, 2012, 21(11):4667-4672.
- [19] YANG C, YU Q. Multiscale fourier descriptor based on triangular features for shape retrieval[J]. Signal Process, 2019, 71: 110-119.
- [20] YANG C. Plant Leaf Recognition by Integrating Shape and Texture Features[J]. Pattern Recognition, 2021, 112(1):107809.
- [21] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL]. (2014-09-04) [2015-04-10]. <https://arxiv.org/abs/1409.1556>.
- [22] HE J, CHEN J N, LIU S, et al. TransFG: A Transformer Architecture for Fine-grained Recognition[EB/OL]. (2021-03-14) [2021-12-01]. <https://arxiv.org/abs/2103.07976>.



**LIU Songyue**, born in 1998, postgraduate. His main research interests include computer vision and deep learning.



**WANG Huan**, born in 1982, Ph.D, associate professor, is a member of China Computer Federation. His research interests include pattern recognition, image processing, infrared target detection, object tracking, robot vision navigation and artificial intelligence.

(责任编辑:喻藜)