

基于计算机视觉的冬虫夏草图像细粒度品质鉴定研究

祁晋东¹ 周艳霞¹ 边巴旺堆¹

QI Jindong ZHOU Yanxia BIANBA Wangdui

摘要

针对人工鉴定冬虫夏草品质效率低下的问题,基于视觉深度学习技术提出了一种新型的冬虫夏草细粒度视觉分类网络,旨在辅助人工高效辨别三类不同品质的虫草。所提出的网络采用了特征融合技术和注意力引导的图像增强策略。前者能够有效生成更具判别力的多尺度特征,后者通过将注意力图像裁剪增强模块和注意力图像融合增强模块相结合的方式,实现对图像的精准增强,有效提升了模型的泛化性能和鲁棒性。实验结果表明,设计的网络在 CUB-200-2011 细粒度图像分类数据集展现了优越的竞争力,同时在 CSCS-3 冬虫夏草图像细粒度品质数据集上取得了高达 96.8% 的分类准确率。

关键词

冬虫夏草; 细粒度视觉分类; 特征融合; 注意力引导; 图像增强策略

doi: 10.3969/j.issn.1672-9528.2024.03.004

0 引言

冬虫夏草是一种名贵的藏药材,具有广泛的药用价值和市场需求。传统上对冬虫夏草品质的鉴定主要根据其外观特征、色泽和虫体形态。然而,这些传统鉴别方法不仅要投入大量人力,而且容易受到主观因素的干扰,从而限制了其可靠性和效率。随着科技的发展,迫切需要建立一种高效准确的标准化鉴别方法,以满足日益增长的冬虫夏草产业需求,促进其可持续发展。

为了有效鉴别不同品质的冬虫夏草,本文采用了计算机视觉图像分类领域的细粒度视觉分类(fine-grained visual classification, FGVC)方法。当前,细粒度视觉分类面临的主要挑战在于同一类物体间的特征差异较小,会导致冬虫夏草的分类效果不佳。因此,本文的研究重点是如何关注冬虫夏草图像语义区域的微小差异。

近年研究表明,以卷积神经网络(convolutional neural network, CNN)为代表的视觉深度学习在图像分类领域表现卓越。虽然 CNN 能够捕捉物体的显著性特征,但在细粒度视觉分类中。本文需要生成多个具有判别力的特征,因此,在冬虫夏草的 FGVC 任务中,必须设计轻量高效的神经网络结构,以取得卓越的分类性能。本研究基于 FGVC,以快速实现对冬虫夏草图像的细粒度品质鉴定,为其提供了一种新方法和新途径。

1 相关工作

在本节中,本文将回顾细粒度图像分类网络和图像数据增强策略,然后进行 FGVC 网络的设计和改进,以提高冬虫夏草细粒度分类的准确性和鲁棒性。

1.1 细粒度视觉分类

基于部位的 R-CNN^[1]方法,将物体局部部位与全局特征结合起来,引导网络更多关注细粒度特征。PS-CNN^[2]通过堆叠多个 CNN 模型,每个模型专注于检测和表示物体的不同部分,最终综合这些信息以提高细粒度分类性能。SPDA-CNN^[3]在提高物体细粒度性能方面,通过统一语义检测和细粒度识别,取得了显著的效果。

随着细粒度视觉分类方法的不断发展,通过引入注意力机制,仅依赖物体的标签,就可使模型更加关注细粒度图像中更有贡献度的语义特征信息。

RA-CNN^[4]通过结合多个尺度的图像输入与注意力机制,实现了更精确的图像分类。通过由粗糙到精细的三个尺度上的分类过程,RA-CNN 能够逐步聚焦关键区域,提取更细致的特征。这种模型设计不仅提高了分类准确性,也为解决细粒度图像分类问题提供了新的思路和方法。此外,WS-DAN^[5]通过引入注意力机制生成多幅注意力图,从而引导图像增强,自动精准聚焦图像中关键区域,完成了显著特征的放大和遮盖操作,确保模型能同时兼顾“显性”和“隐性”特征,从而极大增强模型的泛化能力。

1.2 图像增强

图像数据增强是一种常见的正则化手段,可以促使模型在训练过程中接触更多的“变异”样本,使其更好地适应未知的数据分布。在计算机视觉领域,尤其是细粒度分类任务

1. 西藏大学信息科学技术学院 西藏拉萨 850000

[基金项目] 西藏大学研究生高水平人才培养计划项目 -2021-GSP-S120, 西藏大学校级教改项目 -XZDXJXYJ202212, 西藏自治区教育科学基金项目 -XZEDTP220099

中,数据增强在提高模型性能和鲁棒性方面扮演着关键角色。

在图像数据增强方面,主要涉及单样本图像增强和多样本图像增强。在单样本数据增强中,主要是对当前的数据样本进行缩放、旋转、对称和翻转为主的几何变化以及色彩变化等操作,旨在提升卷积神经网络对图像平移的鲁棒性,进而增强模型的稳健性。此外,多样本图像增强则是通过把从多个图像中提取的特征、风格和结构等融合到原始数据中,从而产生新的样本。Mixup^[6]是最先提出的图像混叠增广方案,其原理简单且方便实现。通过对两幅图像进行混合以生成新的训练样本,以促进模型的泛化性能。而 Cutmix^[7]与 Mixup 直接对两幅图进行相加不同,它是从一幅图中随机裁剪出一个感兴趣区域 (region of interest, ROI), 然后覆盖当前图像中对应的区域。

此外,还有一些以生成对抗网络 GANs^[8]和条件生成对抗网络 Conditional GANs^[9]为代表的无监督图像增强方法,其本质是利用模型学习样本图像的分布,然后随机生成与原图像分布相一致的图像。

为了有效应对冬虫夏草细粒度分类的挑战,本文创新性地提出了一个多通道的细粒度视觉分类网络结构,能丰富特征信息的表达,有效提升模型的性能。该模型主要包含两个重要模块,其一是本文引入了注意力机制引导的图像增强模块,该模块能够将模型学习到的关键特征融入训练过程中,从而提升模型的鲁棒性;其二便是本文开发的特征融合模块,该模块能够筛选出贡献度高的特征,并进行多尺度的局部和全局特征融合,合成细粒度分类任务中更具判别性的特征,显著增强模型的特征表达能力。

2 网络结构

2.1 注意力图像增强模块

注意力机制 (attention mechanism) 起源于对人类视觉的研究,它模拟了人类在处理大量信息时选择性关注的特点,并将之运用在视觉神经网络中。注意力机制的关键在于将关注点自动聚焦在信息量丰富的特征表达上,同时抑制不太有用的表达。

在细粒度视觉分类任务中,当物体尺寸特别小时,如果采用常规的方式进行图像增强,不仅会引入大量的背景噪声,而且对细粒度分类任务贡献度小。如果基于物体的注意力图,就可以对其进行针对性的增强。本文的增强主要分为注意力机制引导的图像裁剪增强和图像融合增强两部分。

2.1.1 注意力机制引导的图像裁剪增强模块

冬虫夏草是一种名贵的藏药材,具有广泛的药用价值和市场需求。传统上对冬虫夏草品质的鉴定主要根据其外观特征、色泽和虫体形态。然而,这些传统鉴别方法不仅要投入大量人力,而且容易受到主观因素的干扰,从而限制了其可靠性和效率。随着科技的发展,迫切需要建立一种高效准确的标准化鉴别方法,以满足日益增长的冬虫夏草产业需求,促进其可持续发展。

ACM 的具体操作流程如下。首先将原始图像提取的注意力图 $A \in \mathbf{R}^{1 \times H \times W}$ 的值进行 0~1 之间的归一化操作,详细的操作公式为:

$$A^* = \frac{A - \min(A)}{\max(A) - \min(A)} \quad (1)$$

然后按照阈值 θ 对注意力图 A^* 进行筛选,将大于此阈值 θ 的像素值设为 1,否则就置为 0,并将注意力图中阈值为 1 的图像部分上采样到原图像大小。最后,得到注意力引导的裁剪图像。

2.1.2 注意力机制引导的图像融合增强模块

由于传统的多样本图像增强方法可能会导致不可控制的背景噪声,这种噪声会对模型的训练和分类结果产生负面影响。为了解决该问题,本文提出了一种新的图像增强方法,即注意力机制引导的图像融合增强模块 (AFM)。该模块的核心思想是利用注意力机制来指导图像融合的过程。通过学习输入图像中的重要特征和结构,使其能够自动确定关键区域,并相应地调整融合过程中的权重。

具体操作流程如下。首先对原始图像生成的注意力图 $A \in \mathbf{R}^{1 \times H \times W}$ 将其像素值大小按照降序进行排列;其次选取前 Q 个值将其设为 1,其余像素值设为 0。然后将像素值为 1 的图像部分通过上采样操作调整至原始图像的尺寸,从而得到最终的掩码 Mask^[10]。

$$\lambda = \frac{Q}{H \times W} \quad (2)$$

通过上述公式 (2), 本文可以得到系数 λ 。

$$\hat{X} = \text{mask} \odot X_1 + (1 - \text{mask}) \odot X_2 \quad (3)$$

$$\hat{y} = \lambda \cdot y_1 + (1 - \lambda) \cdot y_2 \quad (4)$$

式中: X_1 表示当前的训练图像, y_1 表示其图像的标签。 X_2 表示当前批次中的随机一幅图像。 y_2 表示随机图像的标签。 $\hat{X}$ 和 $\hat{y}$ 分别表示融合后的图像和标签。

2.2 特征融合模块

在冬虫夏草细粒度视觉分类任务中,为了有效融合冬虫夏草的局部和全局关键特征,生成更具判别性的多尺度综合特性,本文提出了一种轻量高效型的特征融合模块 (FFM)。其结构如图 1 所示。

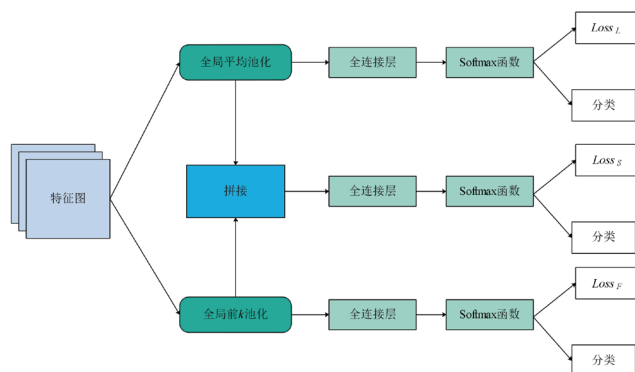


图 1 FFM 结构

FFM 同时结合了全局平均池化和全局前 k 池化技术。全局平均池化是在通道维度上取平均值,从而形成更大的感受野,然后提取更加全面详细的综合特征;全局前 k 池化则关注被卷积核激活的特定位置,学习多个关键的局部区域特征。通过结合这两种池化方法,FFM 模块能够有效地提取多个尺度下的图像特征,从而提供更加丰富的特征信息。

由于两种池化方式关注的区域信息不同,将两类特征向量进行融合^[10],可以获得局部特征和全局信息的关联,形成更具表达力的特征。因此,本文在这里采用拼接的方式进行特征融合,拼接后的特征向量 F_3 能够更好地描述图像内容,提高冬虫夏草图像细粒度分类的准确率。最后将全局平均池化特征向量 F_1 、全局前 k 池化 F_2 和拼接融合后的特征 F_3 输入到独立的全连接层,经过全连接层处理后,利用 Softmax 函数对特征进行分类,从而得到三个预测输出 y_1 、 y_2 和 y_3 ,计算这三个预测输出的均值,作为最终的预测结果。

2.3 网络整体结构

本文提出了一种新颖的冬虫夏草细粒度视觉分类网络,该网络的结构是一个包含原始输入、注意力图像裁剪增强和注意力图像融合增强的多通道网络,如图 2 所示。

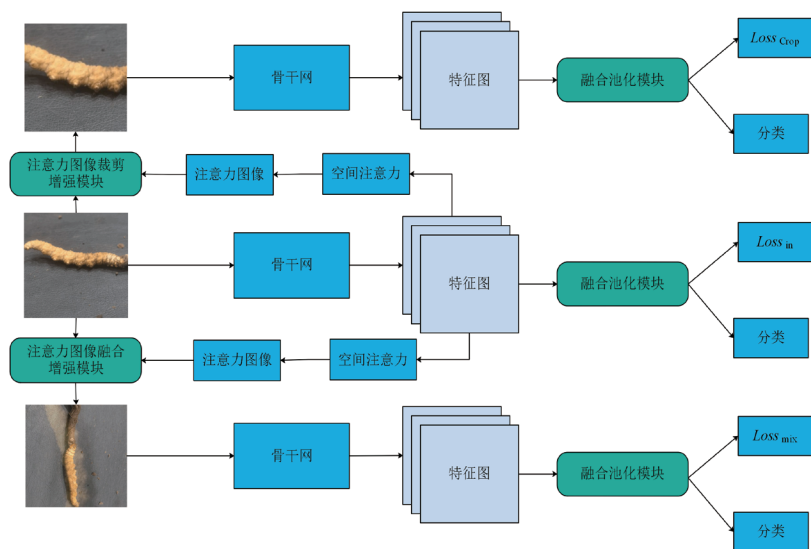


图 2 细粒度视觉分类网络

在原始输入分支中,首先将冬虫夏草图像 $X_{in} \in \mathbf{R}^{C \times H \times W}$,其中 C 、 H 、 W 分别表示图像的通道数、高度和宽度。经过骨干网 (backbone) 提取特征后,得到冬虫夏草的特征图,接下来对其特征图进行空间注意力^[11]操作,其具体操作流程如图 3 所示。

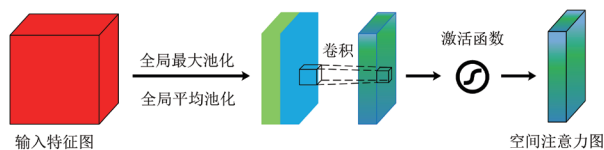


图 3 空间注意力机制

首先在通道维度上进行全局平均池化和全局最大池化,因此得到两个特征图 $F \in \mathbf{R}^{1 \times H \times W}$, 拼接后得到池化特征图 $F^p \in \mathbf{R}^{2 \times H \times W}$; 其次使用 1×1 卷积核提取特征图后经过激活函数进行归一化操作得到空间注意力图 $A \in \mathbf{R}^{1 \times H \times W}$; 再次将注意力图分别输入到注意力图像裁剪增强模块和注意力图像融合增强模块中,这样就可以使网络聚焦贡献度大的空间区域特征,从而分别得到和原始输入图像 X_{in} 尺寸一致的 X_{mix} 和 X_{crop} ; 然后,本文将 X_{mix} 和 X_{crop} 输入到不同通道的注意力裁剪分支和注意力融合分支的骨干网中,得到特征图 F_{mix} 和 F_{crop} ; 最后,分别将 F_{in} 、 F_{mix} 和 F_{crop} 特征图输入到特征融合模块中,并相加 3 个分支的误差 $Loss_{in}$ 、 $Loss_{mix}$ 和 $Loss_{crop}$ 。由于 3 个通道的权重互相共享,在训练过程中能不断进行迭代更新,从而得到最优的网络。

3 实验结果分析

3.1 实验环境

本文实验选择百度 AI Studio 平台,采用 Paddle2.2.2 框架, GPU 为 Tesla V100, CPU 为 4 Cores, Video Mem 为 32 GB,

RAM 为 32 GB, Disk 为 100 GB。本文选用 ImageNet^[12] 上预训练的 ResNet50^[13] 作为骨干网。其中,模块的参数 AFM、FPM、ACM 中 Q 、 k 和 θ 分别设置为 64、8 和 0.5。在冬虫夏草细粒度视觉分类训练任务中,本文采用随机梯度下降优化器 SGD 进行优化,其中,动量为 0.9,权重衰减为 $5e-4$, epochs 为 120。并且在 12 个 epochs 内每个 GPU 的批量大小为 12。学习率为 0.01,在 120 个 epochs 内采用 Cosine Annealing Warm Restarts 对学习率进行调整。

3.2 数据采集

本文使用的是自建的冬虫夏草图像细粒度品质数据集 CSCS-3,在单一的背景中,首先采用微单拍摄优、良和差 3 种品质的冬虫夏草图像共计 4500 张,其次按照随机裁剪和旋转的增强策略把数据扩充到 9000 张,然后将图像尺寸归一化为 224×224 ,最后将图像尺寸按照 8:2 的比例划分为训练集和测试集。

3.3 对比实验

本文选择 ResNet50^[13] 作为骨干网在 RA-CNN 和 WS-DAN 细粒度视觉分类中的弱监督网络进行对比实验,在 CBU-200-2011 和 CSCS-3 细粒度视觉分类数据集集中进行实验,各网络分类的准确率如表 1 所示。

表 1 不同网络在公开数据集和自建数据集上的分类准确率

网络	骨干网络	CUB-200-2011	CSCS-3
RA-CNN	ResNet50	86.1%	90.7%
WS-DAN		88.5%	93.5%
本文网络		89.7%	96.8%

通过表 1 可以看出,本文提出的网络在公开数据集 CUB-200-2011 和自建数据集 CSCS-3 均取得了最优的性能,在冬虫夏草图像细粒度品质数据集的准确率为 96.8%。

3.4 消融实验

为了验证本文提出网络的有效性,本文在骨干网 ResNet-50 的基础上,分别加入 FFM、ACM 和 AFM,所得的实验结果如表 2 所示。

表 2 各网络模块在 CUB-200-2011 数据集上的消融实验结果

FFM	ACM	AFM	CUB-200-2011
—	—	—	84.8%
✓	—	—	87.2%
✓	✓	—	88.3%
✓	✓	✓	89.7%

通过表 2 可以看出,ResNet-50 结合 FFM 在 CUB-200-2011 数据集上准确率上升 2.4%,然后再结合 ACM 后,准确率上升 1.1%,最后结合 AFM,准确率上升 1.4%。综上所述,本文发现 3 个模块对细粒度分类任务均有提升,其中 FFM 对模型性能提升的贡献度最大。

4 结语

本文所提出的细粒度视觉分类网络结合了特征融合模块和基于注意力机制引导的图像增强模块。特征融合模块能够关注不同尺度物体的判别性特征。此外,为了实现冬虫夏草图像细粒度品质数据的精准增强,本文特别设计了基于注意力机制引导的图像裁剪增强模块和图像融合增强模块。通过这两个数据增强模块的联合协同作用,有效提升了冬虫夏草图像细粒度的识别率。在未来的研究中,可以继续深入探索和优化注意力图像的生成方式,从而进一步提升细粒度网络的分类性能。

参考文献:

[1] ZHANG N,DONAHUE J,GIRSHICK R, et al.Part-based R-CNNs for Fine-grained Category Detection[M].Berlin:Springer International Publishing,2014:834-849.

[2] HUANG S,XU Z,TAO D , et al. Part-stacked cnn for fine-grained visual categorization[J].Computer Vision and Pattern Recognition, 2015,12:27-30.

[3] ZHANG H,XU T,ELHOSEINY M,et al. SPDA-CNN: unify-

ing semantic part detection and abstraction for fine-grained recognition[EB/OL].(2016-11-12)[2023-12-06].<https://ieeexplore.ieee.org/document/7780498>.

[4] FU J, ZHENG H,MEI T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition[EB/OL].(2017-11-09)[2023-12-12].<https://ieeexplore.ieee.org/document/8099959>.

[5] HU T ,QI H . See better before looking closer: weakly supervised data augmentation network for fine-grained visual classification[EB/OL].(2019-01-26)[2023-11-08].<https://arxiv.org/abs/1901.09891>.

[6] ZHANG H,CISSE M,DAUPHIN Y N,et al. Mixup: beyond empirical risk minimization[EB/OL].(2017-10-25)[2023-11-22].<https://arxiv.org/abs/1710.09412>.

[7] YUN S,HAN D,OH S H, et al. CutMix: regularization strategy to train strong classifiers with localizable features[EB/OL]. (2019-05-13)[2023-12-06].<https://arxiv.org/abs/1905.04899>.

[8] GOODFELLOWIAN I J, ABADIE J P,MIRZA M,et al. Generative adversarial networks[EB/OL].(2014-06-10)[2023-11-29].<https://arxiv.org/abs/1406.2661>.

[9] MEHDI M,OSINDERO S. Conditional generative adversarial nets[EB/OL].(2014-11-06)[2023-12-01].<https://arxiv.org/abs/1411.1784>.

[10] 肖斌,郭经伟,张兴鹏,等. 基于融合池化和注意力增强的细粒度视觉分类网络 [J]. 模式识别与人工智能, 2023, 36(7): 661- 670.

[11] WOO S,PART J,LEE J Y,et al.Cbam: Convolutional block attention module[EB/OL].(2018-07-17)[2023-11-28]. <https://arxiv.org/abs/1807.06521>.

[12] DENG JIA ,DONG W, SOCHER R , et al. Imagenet: A large scalehierarchical image database[EB/OL].(2009-08-18)[2023-12-06]. <https://ieeexplore.ieee.org/document/5206848>.

[13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway:IEEE,2016: 770-778.

【作者简介】

祁晋东（1997—），男，山西永济人，硕士研究生，研究方向：计算机视觉和模型轻量化等。

周艳霞（1992—），通信作者，女，四川南充人，硕士研究生，实验师，研究方向：图像识别和科技史等。

（收稿日期：2024-01-04）