

ConoHN: 层级知识增强的牙形石细粒度图像分类框架

夏雨萌, 樊媛媛, 梁庆中, 李新川, 姚宏

(中国地质大学(武汉)计算机学院, 湖北 武汉 430078)

摘要: 牙形石的鉴定工作是古生物领域的一项重要任务, 而现有牙形石图像分类研究较为稀缺, 且现有研究存在特征学习不足和类间多区域判别性差异难以挖掘的问题。为了解决上述问题, 提出了ConoHN——一个引入层级知识的层级多标签图像分类框架。它主要将寻找各子类别判别性区域的困难任务解耦为分别找到子类所有父层级和当前层级判别性区域的两类简单子任务。它利用注意力机制, 将骨干网络提取的特征离散为各层级的特征, 同时擦除高层级语义信息, 以强迫模型关注其他部分的低层级判别性区域。实验结果显示, 该模型在多层级图像分类领域显著优于对比方法, 验证了模型有效性。

关键词: 牙形石; 化石图像分类; 层级多标签图像分类; 细粒度图像分类

DOI: 10.11907/rjdk.251256

中图分类号: TP391.41

文献标识码: A

文章编号: 1672-7800(2026)007-0140-12

扫描二维码阅读全文:



ConoHN: A Hierarchical Knowledge-Enhanced Framework for Fine-Grained Conodont Image Classification

XIA Yumeng, FAN Yuanyuan, LIANG Qingzhong, LI Xinchuan, YAO Hong

(School of Computer Science, China University of Geosciences(Wuhan), Wuhan 430078, China)

Abstract: Conodont identification is a critical task in paleontology, yet current image classification methods remain limited by insufficient feature learning and difficulties in capturing discriminative multi-region variations across species. To address these challenges, we propose ConoHN, a hierarchical multi-label classification framework that incorporates taxonomic knowledge. Our key innovation decomposes the complex task of locating sub-category discriminative regions into simpler hierarchical subtasks through attention-based feature discretization across taxonomic levels (family-genus-species), and high-level feature erasure to force learning of low-level discriminative patterns. Experiments demonstrate that ConoHN significantly outperforms state-of-the-art methods in multi-level classification accuracy, while providing interpretable region-focused attention maps.

Key Words: conodont; fossil image classification; hierarchical multi-label image classification; fine-grained image classification

0 引言

牙形石(Conodont)是一类已灭绝的水生脊椎动物, 广泛分布于寒武纪至三叠纪末期的海洋中, 历经前三次大灭绝事件(包括二叠纪末大灭绝), 最终于三叠纪末灭绝事件中消亡。凭借其演化迅速、分布广泛、易于获取等特性, 牙

形石成为生物地层划分和古环境重建的关键指标^[1]。其形态特征不仅能指示地质年代、划分和对比地层, 还可通过元素地球化学分析(如牙形石颜色蚀变指数CAI)推断岩石有机变质程度和古海水温度^[2]。

牙形石分类准确性直接决定其科学应用价值, 其分类的核心挑战之一在于识别形态高度相似物种间的细微差异。例如, *C.changxingensis* 由 *C.yini* 进化而来, 其差异仅在

收稿日期: 2025-05-07

基金项目: 国家自然科学基金项目(42071382); 湖北省自然科学基金项目(2023AFA006)

作者简介: 夏雨萌(1999-), 女, 中国地质大学(武汉)计算机学院硕士研究生, 研究方向为计算机视觉; 樊媛媛(1979-), 女, 博士, 中国地质大学(武汉)计算机学院副教授, 研究方向为计算机视觉; 姚宏(1976-), 男, 博士, 中国地质大学(武汉)计算机学院教授、博士生导师, 研究方向为知识图谱。本文通讯作者: 樊媛媛。

于下部齿形, *C. changxingensis* 小齿则更为尖锐(见图1), 这对模型的细粒度区分能力提出了极高要求。尽管早在20世纪80年代就已引入计算机辅助化石识别技术, 但由于浅层机器学习算法(如神经网络、支持向量机)的局限性, 其泛化能力较差。早期方法多结合手动特征提取与浅层学习算法, 需要人工干预^[3-6]。随着卷积神经网络的发展, 这类方法逐渐被自动识别的方法所替代。现有化石分类方法在部分任务中取得了一定进展, 然而其仅针对鉴定可分度较高的样本, 在处理形态高度相似的牙形石分类任务时仍面临显著局限性, 阻碍了计算机辅助化石识别技术在古生物学中的实际应用。

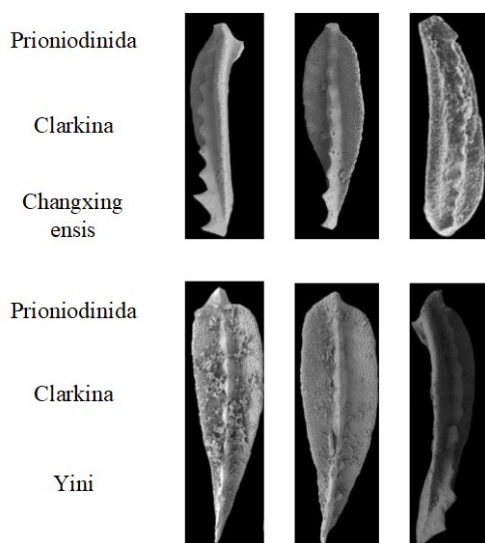


Fig. 1 Three-view displays of two conodont species from the genus

Clarkina

图1 Clarkina属下两种牙形石的3种视图展示

细粒度图像分类方法(Fine-Grained Image Classification, FGVC)的核心在于识别特征高度相似的子类对象, 能够捕捉图像之间的细微差异, 并准确定位不同子类的强区分区域。由此, 细粒度图像分类方法为解决此类问题提供了新的思路。在当前FGVC研究中, 特征编码方法和定位判别性区域方法是两大主流方向。特征编码方法的研究重点在于通过高阶特征表达增强图像的分类能力, 大多针对双线性池化(HBP)方法进行改进, 以获得图像的高阶编码特征。在牙形石自动鉴定领域, Duan等^[7]尝试利用Bilinear-ResNet18模型识别牙形石 *Hindeodus* 属下的9个子类, 取得了67.6%的分类准确率。然而, 这类方法的分类精度较低, 且数据集种类较为单一, 类别数量和信息丰富度远低于本文研究的更复杂数据集(包含3科、8属、30种), 难以直观反映不同种类间的分类依据。与特征编码方法不同, 定位判别性区域的方法通过捕获区域, 能够更直观地揭示子类之间的细微差异, 可解释性往往更强。近年来, 随着注意力机制的引入, 这类方法在学习图像特征和定位关键区域方面展现出显著优势。例如, RA-CNN通过多次对图像中的关键区域进行聚焦, 依次提取细微区域

特征, 从而实现了更精准的分类^[8]。然而, 这类方法通常孤立地检测对象各部分, 忽略了不同区域间的语义联系。这种局限性导致模型注意力往往集中在单一图像区域, 而缺乏捕获具有区分能力的多个关键区域的能力。同时, 牙形石判别性区域并非孤立存在, 而是嵌入于科—属—种的多级分类体系中, 属级分类依赖整体轮廓(如对称性、主体膨大程度)的全局表征, 种级分类则需协同局部细节(如齿片间距、齿脊曲率)与区域结构(如齿片复合体形态)的层级关联。现有方法通常假设物种标签独立互斥, 忽视了牙形石分类固有的层级先验知识, 导致判别性区域间语义约束缺失, 本质上是割裂了形态特征间的生物学依赖关系, 削弱了模型对判别性区域关联性的建模能力。

针对以上问题, 本文提出了一种融合领域知识的层级多标签细粒度牙形石图像分类框架 ConoHN。首先, 结合牙形石天然的类间层级关系, 构建了科、属、种三级标签体系, 并设计了层级特征提取与交互机制, 通过深入建模层级特征间的语义联系, 解决了类间多区域判别性差异难以挖掘的问题, 为后续研究提供了数据和方法基础。实验结果表明, 本文方法在属种级分类任务中均取得了显著性能提升, 为牙形石智能鉴定提供了高效、可靠的解决方案, 同时为其他古生物分类任务提供了可推广的技术范式。

1 相关工作

1.1 化石图像分类

化石图像分类作为智能技术与地质学研究的结合, 为地质年代划分、古生物群落研究等提供了科学支持, 对相关学科的研究具有重要意义。近年来, 化石图像分类研究随着智能技术的发展而取得了显著进展。通过计算机视觉和深度学习技术, 对化石图像进行自动分类已成为地质学和古生物学研究的重要工具。

根据已有研究, 化石的智能识别主要依靠计算机视觉领域的机器学习模型和深度学习模型。Hsiang等^[4]使用3个CNN训练了27 737张现代浮游有孔虫图像, 性能最好的分类器在测试集共6 903张图像上达到了87.4%的准确率。Mitra等^[9]对古海洋学家广泛研究的6种现存浮游有孔虫进行了自动鉴定, 研究表明, 其自动识别准确率略高于专家识别, 远高于新手识别, 且由于先验知识的局限性, 人工识别的召回率远低于自动识别。Keçeli等^[10]使用预训练的VGG16网络模型进行微调, 在化石放射性数据集上实现了90.22%的准确率。Liu等^[11]使用从18加深到152层的5种ResNet网络模型, 并采用迁移学习对8个奥陶纪锥齿象的数据集进行训练, 共包含1 761个图像数据。安玉钊等^[12]对4种中生代鸵鸟化石进行了分层智能识别, 先对化石进行初始分类的目标检测, 然后在此基础上采用卷积神经网络(Convolutional Neural Network, CNN)和支持向量机(Support Vector Machine, SVM)进行更详细的分类,

最终识别准确率为 95%。以上化石分类方法均使用简单的骨干网络,且数据集每个类具有较为充足的样本量。

上述研究覆盖了从古植物、无脊椎动物到脊椎动物等多种化石类型,其中牙形石因其特殊的形态和地层分布,成为化石图像分类的重要研究方向。牙形石分类不仅需要应对化石分类的共性问题,如样本变异性 and 数据稀缺性问题,还需要处理其独特的层级关系和细粒度分类需求。因此,对牙形石图像分类的研究现状进行探讨,能够进一步推动化石分类技术发展,并为更广泛的古生物学研究提供借鉴。当前,牙形石图像研究大多属于人工鉴定。Yuan 等^[13]首次提出基于单个形态特征识别牙形石物种的困难性,对中二叠纪 *Neogondolella* 属下的样本进行了研究,因为该属的物种显示出相当大的形态变异性。并且,经过对整个样本种群的分析,包括幼年到老年阶段的标本,研究人员认识到样本节段性元件形状存在明显的种内变化。也有少量研究使用神经网络对牙形石进行自动鉴定。Duan 等^[7]尝试利用 Bilinear-ResNet18 识别牙形石 *Hindeodus* 属下的 9 个子类,达到了 67.6% 的准确率。分类精度较低,且数据集的样本种类单一,类别数量和信息丰富度远少于本文的数据集(3 科、8 属、30 种)。同时,研究人员建议所有用于 CNN 模型训练的化石样本都属于同一属,因而多属牙形石分类研究仍处于相当匮乏的状态。

1.2 细粒度图像分类

现有细粒度识别方法主要基于数量较为充足的样本进行研究,通常被划分为如下两广泛认可的范式:①基于定位判别性区域;②基于端到端特征编码。基于定位判别性区域方法的核心思想在于捕获细粒度类别之间共享的语义部分(如头部和躯干),并从中挖掘能够区分类别的细微差异^[14]。这类方法模型通常由定位子网络和分类子网络组成^[14-16]。定位子网络的主要功能是定位细粒度对象的判别性区域(例如头部、躯干等),提取这些部件的局部特征向量;分类子网络则通过结合这些局部特征与对象级全局图像特征,实现最终分类。Zhang 等^[17]采用空间金字塔策略,从对象提议中生成部件提议。随后,通过聚类方法生成了部件提议的原型聚类,并进一步筛选出有用的聚类以获取具有判别力的部件级特征。在弱监督场景下,基于共同分割的方法也被广泛使用。一种常见的方法是在无监督条件下利用共同分割生成对象掩码,随后通过启发式策略,如部件约束,以定位细粒度部件^[18]。

上述方法通过减少对高质量人工注释(如边界框或关键部件标注)的依赖,大幅降低了数据标注成本,同时在实际应用中表现出较强的灵活性。然而,上述方法通常关注独立区域的区分能力,而忽略了多个部件特征之间可能存在的内在语义联系。这种局限性使得模型在处理复杂的细粒度分类任务时,无法充分挖掘部件间的相关信息和协同作用。针对这一问题,有研究开始尝试联合学习部件级特征之间的相互依赖关系,以提升特征表达的区分能力。

例如,Wang 等^[19]利用交叉图传播方法在判别区域定位阶段通过交叉加权的方式聚合其他区域的信息,从而增强区域之间的相关性。同时,在分类阶段通过迭代增强有用特征并抑制无用特征,以探索判别性块特征向量之间的内在语义相关性。

尽管上述方法已多个领域取得良好效果,但针对牙形石分类任务仍然存在一些特定挑战。牙形石的形态特征通常表现为复杂的细粒度差异,例如不同种类在齿形结构、形态比例及细微纹理上存在着极难察觉的变化,这对现有方法的局部特征建模能力提出了更高要求。对于牙形石分类,某些重要区域(如对称的主齿和轻微外膨的体部)可能共同构成某种类别的重要判别依据,但这种区域之间层级语义上的相互依赖关系未被充分挖掘,导致模型无法捕捉到复杂的部件协同特征。

1.3 层级多标签图像分类

在通用图像分类任务中,各类别之间通常是两两互斥,不存在包含关系。然而,层级多标签分类任务则较为特殊,类别之间具有明确的层次结构关系,通常可以用树形结构或有向无环图(DAG)加以表示。在这种任务中,单个样本的标签会同时包括层次结构中的父类和子类,形成所谓的层级多标签分类(HMC)问题^[20]。

HMC 方法最初应用于细粒度文本分类任务^[21],但在诸如 CUB200-2011^[22]、iNat2017^[23]等自然生物数据集中,其广泛存在分层的类别层级关系。对于非生物数据集,例如 Stanford-Car^[24]和 Aircraft^[25],其类别也存在类似的层级结构。因此,对于有着明确类间层级关系的数据集,可以将它们的细粒度识别任务建模为层级多标签图像分类任务^[26-28]。HMC 的常用方法主要分为两类:局部方法和全局方法^[29]。

全局方法通常直接利用层级结构信息加以建模,仅由 1 个分类模块构成。例如:利用层次结构,构造递归正则化损失项以对分类参数进行约束。Deng 等^[30]将任意两个标签之间的语义连接形式化为有向无环图(DAG),其构建了一个修改后的连接树算法,该算法在连接树上的消息传递期间包含多个循环,以计算定义在 DAG 上的概率分类损失。C-HMCNN 修改了二元交叉熵损失函数,加入父子约束,以确保层级结构的一致性,即当样本被预测为某子类时,模型自动保证其属于父类^[31]。全局方法的优点在于建模简单,但由于无法充分利用层级信息中的细粒度特征,模型容易欠拟合,从而表现出较低的分类性能。

局部方法则通过将层级关系编码到网络架构或损失函数中以帮助模型进行预测。通过学习不同层类的类别特征,并聚合不同层的特征以得到最终预测结果。这类方法通常由多个针对不同层级的分类模块组成,通过学习各层级特征并聚合不同层级特征的方式,得到最终分类结果。自顶向下的层次结构方法将在每个非叶子节点层级上训练一个局部分类器。部分研究尝试通过优化不同层

级的特征表征以提升分类性能。例如,Chang等^[32]利用分类头离散粗略级特征与细粒度特征,并允许细粒度特征参与粗略级标签预测。Chen等^[33]提出层级残差网络,编码了类标签结构的父子关系以传递层级知识。ENN结合证据神经网络依据迪利克雷分布建模父子层级之间的概率关系,显著提升了分类准确性^[34]。还有方法则重点利用高层级的分类结果指导低层级的特征学习。例如,HSE逐层预测得分向量,并利用上一层的得分向量作为软目标引导更细粒度的特征表示^[35]。基于局部的方法能够利用更细粒度的层级信息,但往往需要构建多个分类模块,同时也容易受到误差传播的影响。为减少以上误差,Wang等^[36]提出一种有效的一致性感知特征学习方法,通过对标签层次结构进行前向推导和后向归纳,将预测一致性约束编码为弱监督机制,同时提高预测一致性和分类精度。

由于牙形石种类繁多且形态特征具有天然的层次关系,层级信息在牙形石分类任务中具有天然的适用性。牙形石在演化过程中,其主要特征往往在粗粒度类别(如科或属)上具有一致性,而在细粒度类别(如种)上则表现出显著差异。这种层级结构与HMC方法的建模方式高度契合,使其成为牙形石分类的潜在解决方案。然而,以上HMC方法在建模层级信息时,通常将不同区域的特征作为独立表征进行处理,忽略了区域之间可能存在的语义联系。这种方法往往仅关注父类与子类之间的从属关系,而缺乏对父类与子类层级特征在某一类中高度共现的建模。例如,在牙形石分类中,对称的主齿和外膨的体部可能共同形成对某个类别的重要判别依据,但现有方法通常难以捕捉它们之间的联系,这限制了模型对复杂形态特征的理解能力。因此,如何在层级分类方法中引入区域间的语义联系建模,将是提升牙形石分类性能的关键研究方向。

2 数据集构建

由于缺乏牙形石相关公开数据集,本文通过爬虫等技术收集了大量牙形石学术论文,以期获取研究所用数据集。同时,为了改善牙形石类间差异小、数据稀疏带来的影响,提出了一种层级多标签结合细粒度特征的牙形石分类方法。该方法利用层级标签信息对样本进行多粒度组织与建模,将标注到不同粒度的样本,通过区分不同层级特征,并优化细粒度特征提取与交互方式,显著提升了牙形石分类准确性和鲁棒性。

牙形石是古生物学研究中的重要化石类型,其形态多样性为分类学、进化学和地层学研究提供了关键依据。然而,牙形石研究面临样本获取和分类效率的双重挑战。传统的牙形石研究主要依赖于显微镜下的人工观察和分类,这种方式不仅耗时费力,而且主观性强,难以在大规模数据上实现高效、可靠的分析。同时,现有的公开牙形石数据库普遍存在样本规模有限、标注信息不足的问题,无法

满足深度学习模型对大规模高质量数据的需求。

近年来,随着学术论文的数字化发展,越来越多的牙形石化石图像及其相关信息发表在学术文献中。这些文献不仅包含大量高质量的牙形石图像,还附带了对应的图注和上下文描述,为构建牙形石数据集提供了重要资源支持。基于此,本文充分利用这些现有资源,设计了一个高效的数据集构建流程,以服务于牙形石自动分类任务。具体而言,从文献资源出发,依托自动化技术和少量人工干预,完成了从数据收集到目标检测模型训练的全流程。由于研究人员通常研究地层的特定部分,收集到的化石样本不可避免地集中在几个属中或代表属中的多个物种,因而本文数据集尽可能完整地包含特定地层属于该属的所有物种以使得模型有用性得以最大化。

本文将详细介绍数据集构建的3个步骤:①论文图像图注数据获取;②标注与目标检测模型训练;③化石图像与图注语义对齐。

2.1 论文图像图注数据获取

牙形石图像数据的来源主要是公开发表的文献(共225篇)。作为学术产出的重要指标,会议论文、期刊等文献不仅仅阐述了作者本人的观点,还包含了大量作者收集归纳得到的数据信息和图形内容。因此,可将文献中的大量图片作为数据来源,解决数据渠道问题。同时,还可以将文献中对应的图注信息作为专家指导,为图像打上正确标签。

为了从大量学术论文中系统性地提取牙形石相关数据,本文开发了一套论文爬虫工具。该工具能够批量下载学术论文,并对文献内容进行解析,自动提取牙形石化石图像及其相关图注信息,初步构建了一个包含牙形石图像的基础数据集。这一步不仅解决了数据来源问题,还为后续牙形石图像目标检测训练打下了基础。

2.2 数据标注与目标检测模型训练

尽管从论文中提取的牙形石图像质量较高,但这些图像中通常包含多个化石图像,如图1所示。因此,本文结合古生物学领域的专业知识,对图像中的牙形石目标进行了人工标注。标注过程中,为图像中的牙形石目标添加边框,并附上相应的文本描述,以帮助模型理解目标的特征。需要指出的是,与传统图像分类任务不同,本文目标并不是对每一种牙形石进行精细区分,而是首先区分图像中的前景(牙形石化石)和背景。

基于构建的标注数据集,本文进一步训练了目标检测模型,以实现牙形石目标的自动化识别。具体而言,通过采用轻量化的U-Net网络模型,在像素级别上对牙形石图像的特征进行提取和分类。U-Net模型能够通过卷积操作学习牙形石前景和背景的区别特征,其轻量化设计既满足了高效性要求,又确保了识别精度。此外,在训练过程中,通过数据增强技术对标注数据进行扩展,进一步提高了模型鲁棒性和泛化能力。

2.3 化石图像与图注语义对齐

在构建训练数据集的过程中,为了实现图像中的牙形石标本与对应图注文本的准确匹配,必须对化石图像与图注信息进行语义对齐。这一步对于确保每个牙形石标本的图像区域能够正确关联其属种信息至关重要。

2.3.1 基于规则和遗传算法的对齐方法

由于单个化石图像通常包含多个牙形石标本,而图注文本中往往同时描述了这些标本的属种信息,因而需要一种有效的方法将图像中的标本与图注中的描述逐一对应起来。为此,本文采用了基于规则和遗传算法的对齐方法。

首先,基于规则的方法主要利用图注文本中标本编号与图像中标本的物理位置关系。学术论文中的图注通常按照一定的顺序列出标本编号和对应属种信息,例如“Fig.1(A)Species X;(B)Species Y;...”根据这些编号,可以初步建立图注描述与图像中各标本之间的对应关系。通过解析图注文本,提取出每个标本编号和对应属种信息,然后在图像中根据标本排列顺序、位置等先验知识,初步匹配图像中的标本与图注文本。

然而,仅依靠规则匹配可能无法完全准确地对齐,特别是在标本排列复杂或图注描述不明确的情况下。为了进一步提高对齐准确性,本文引入了遗传算法。遗传算法是一种基于自然选择和遗传机制的优化算法,适用于搜索全局最优解。在本次对齐任务中,遗传算法用于优化标本匹配的组方案。

具体而言,本文将图像中的标本区域和和图片中的序号(对应图注中的描述)视为两组待匹配的元素。首先,生成初始的匹配组合作为种群中的个体,然后定义适应度函数以评估每个个体的匹配质量。适应度函数可以考虑标本的形状特征、大小、位置等视觉信息与图注描述的关联程度。通过选择、交叉和变异等遗传操作,不断迭代优化种群,最终找到使适应度函数最大化的匹配方案。

这种基于规则和遗传算法相结合的对齐方法,既利用了图注文本提供的明确编号和顺序信息,又通过遗传算法的全局搜索能力,克服了仅靠规则匹配可能出现的误差。经过多次迭代优化,最终实现了化石图像与图注文本的精确对齐,为后续图像分类训练数据集提供了高质量的标注。

2.3.2 评价指标

为定量验证规则与遗传算法相结合的化石图像与图注文本对齐方法的有效性,本文实验选取机器学习分类任务中的精确度(Precision)、召回率(Recall),以及目标检测领域的平均交并比(mIoU)作为核心评价指标。TP表示实验预测结果为古生物化石目标,实际结果也为古生物化石目标;FP表示实验预测结果为背景,实际结果却为古生物化石目标;FN表示实验预测结果为古生物化石目标,实际结果却为背景;TN表示实验预测结果为背

景,实际结果也为背景。

其中,准确率(Precision)表示检测到的古生物化石目标正确的数量(将背景当作一个目标)占有所有检测出的古生物化石目标数量的比例。召回率(recall)表示检测到的古生物化石目标正确数量占有所有标注的检测到的古生物化石目标数量的比例。而mIoU则表示检测到的前景(古生物化石)、背景两种目标框与实际正确目标框的相交部分占实际正确目标框的比例。

2.3.3 实验结果

本文选择U-Net模型进行训练,化石识别准确率达到90%以上,其结果如图2所示。

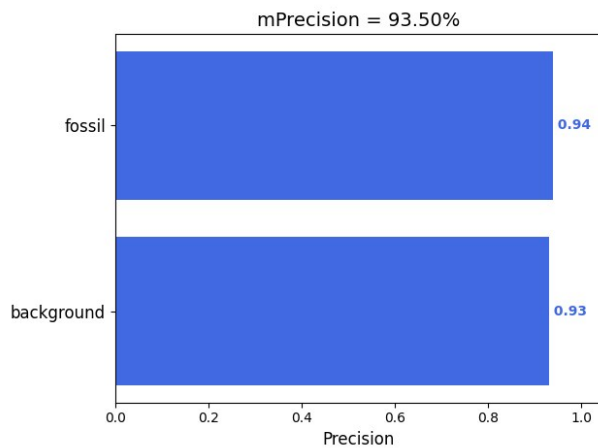


Fig. 2 mPrecision metric

图2 mPrecision 指标

2.3.4 所构建数据集结果

通过以上方法,最终收集到11 948张与牙形石相关的图像数据,如表1所示。这些图像涵盖了丰富的分类层级,包括3科、8属和30种,为研究提供了广泛的数据支持。然而,为保证数据质量与分类结果可靠性,在数据预处理中剔除了图像质量较差和样本量极少的类别。经过筛选后,最终保留下6 437个粗粒度标注样本和3 885个细粒度标注样本,形成了一个结构清晰且覆盖范围广泛的数据集。数据集各层级不同种类详细信息如表1所示,其给出了收集到的从粗到细粒度层级的各类样本数量,共3科7属30种。在“Fam_N”和“Gen_N”列中,括号内的数据表示在相应粒度下标注的图像数量。

3 实验方法

层级多标签图像分类(Hierarchical Multi-Label Image Classification, HMIC)是一种重要的图像分类任务,旨在利用类别之间的层级关系以提升分类性能。这种任务通常针对具有层级结构的类别数据展开,如生物分类中的界、纲、目、科、属、种关系,或工业产品分类中的类别、子类别、具体型号等。在细粒度分类任务中,类别之间不仅存在细微差异,还往往呈现出层级结构的特点,这使得层级信息

Table 1 Conodont species numerical distribution statistics in the dataset									
表 1 数据集中牙形石各种类统计情况									
Order	Ord_N	Family	Fam_N	Specie	Spe_N	Train	VAL/Test		
Gondolellidae	6 434(3 004)	Clarkina	2 800(1 799)	changxingensis	428	300	64		
				yini	208	146	31		
				subcarinata	194	136	29		
				longicuspidata	141	99	21		
				meishanensis	208	146	31		
				orientalis	214	150	32		
				postbitteri	148	104	22		
				liangshanensis	113	79	17		
				wangi	145	102	22		
				quadrata	180	126	27		
		Epigondolella	847(415)	rigoï	104	73	16		
				triangularis	131	92	20		
				dieneri	336	235	50		
		Neospathodus	1 318(557)	pakistanensis	96	67	14		
				waageni	125	88	19		
		Quadralella	826(233)	polygnathiformis	166	116	25		
				praecommunisti	67	47	10		
				changxingensis	110	77	17		
		Anchignathodontidae	3 065(1 957)	Hindeodus	1 916(1 461)	eurypyge	106	74	16
						latidentatus	93	65	14
parvus	621					435	93		
praeparvus	332			232	50				
typicalis	199			139	30				
isarcica	90			63	14				
Isarcicella	549(496)			lobata	90	63	14		
				staeschei	188	132	28		
				turgida	128	90	19		
Novispathodontidae	823(590)			Novispathodus	705(590)	abruptus	160	112	24
		pingdingshanensis	216			151	32		
		waageni	214			150	32		
Sum	10 322	7	8 961	30	5 551	3 885	833		

在分类任务中的重要性日益凸显。

相比于传统的平面分类任务,层级多标签图像分类的一个显著特点是:类别之间的父子关系和层级结构能够提供更多先验知识,有助于在标签稀疏或类别相似的情况下提升模型区分能力。然而,现有的 HMIC 方法通常忽略了层级间的关系,这种方法无法充分挖掘层级特征间的潜在关联,尤其是在类别层级较深的情况下,往往难以取得较好的分类效果。

在层级结构中,类别的父子关系可以用树状结构或有向无环图(DAG)表示。以前的细粒度分类方法往往仅关注于种级分类,忽略了粗粒度标注的潜在价值。这种方法对数据稀疏性较强的牙形石数据集表现不佳,尤其是样本量不足的种级类别,无法通过有限的学习数据到充分的特征表示。而粗粒度标注(如属级或科级标注)能够提供更高层次的语义信息,可帮助弥补种级数据不足的问题。类别的层级关系不仅可以帮助模型在父类中准确定位子类,还能够通过高层次信息(如科级特征)辅助低层次类别(如种级特征)的分类。然而,现有方法缺乏利用粗粒度标注的先验信息以辅助细粒度分类,导致模型无法充分利用粗粒度标注以增强种级分类性能。而且,从古生物论文中抽

取到的数据集存在大量标注为粗粒度的样本。

针对上述问题和难点,本文提出了一种结合层级关系的多标签细粒度图像分类模型(Hierarchical Multi-Label Fine-Grained Image Classification, HML-FGIC),该模型通过充分利用粗粒度标注样本的全局特征信息,指导细粒度类别的分类学习,以解决以下两个核心挑战:①如何有效利用大规模粗粒度样本?②如何缓解类间差异细微的问题?对于挑战一,粗粒度样本在类别层级上具有覆盖广泛的特点,包含丰富的全局语义信息。然而,现有的细粒度分类模型难以直接整合粗粒度标注样本,主要是因为类别层级信息的不对称性会对模型训练造成干扰。例如,粗粒度样本中的属级或科级类别标注无法直接映射到种级类别分类任务中。这种类别层级间的语义脱节导致粗粒度样本在细粒度分类任务中利用率不足。因此,本文模型通过层级标签的分层编码机制对粗粒度样本进行语义建模,并提出一种层级损失函数,分别计算不同层级的分类损失,利用损失函数约束模型优化结果。对于挑战二,细粒度分类的关键在于捕捉类别间的细微差异,而粗粒度特征主要提供类别的全局信息,二者在本质上存在表征目标的差异性。本文模型设计了一种特征擦除模块(Semantic

Erasure Module, FEM), 能够让粗粒度特征更好地支持细粒度分类。通过利用粗粒度样本生成的类激活图(Class Activation Map, CAM), 动态筛选并抑制高响应区域。具体而言, SEM 模块会通过粗粒度样本的特征分布生成激活图, 并对高响应区域进行特征擦除, 从而减少模型对粗粒度显著区域的过度依赖, 同时引导模型关注细粒度类别中更具判别力的区域特征。随后, 这些细粒度特征将通过一组权重共享机制与粗粒度特征进行交互, 增强模型对关键细节的捕捉能力。

本文模型在设计过程中, 不仅充分考虑了粗粒度标注样本的语义价值, 还通过特征交互机制和一致性约束, 建立了粗粒度特征对细粒度分类的有效支持关系。通过验证, 该模型在处理牙形石种级分类任务中表现出显著的性能提升。鉴于此, 本文提出了 ConoHN 框架, 对牙形石图像进行分类。

3.1 问题定义与数据建模

在单标签分类任务中, 每张图像 x 通常被标记为 l_i , 其中 $i \in \{1 \dots k\}$, 且标签 y 是不可分的, 各标签之间互斥。然而, 在多标签分类任务中, 每个图像的标注不仅仅局限于单一标签, 而是通过层次化的结构丰富了标签的描述信息。具体而言, 每张图像 x 被标记为 $L = (l_1, l_2, \dots, l_h)$, 其中 h 表示数据集的层级数量, l_m 属于第 m 层级的标签集合 C_m , $C_m = \{c_1 \dots c_{n_m}\}$, n_m 为第 m 层的类别数量。

以本文数据集为例, 其层级数为 $h = 3$, 且 $n_1, n_2, n_3 = 3, 8, 30$, 分别对应 3 科、8 属和 30 种。基于数据集中牙形石

的种属关系, 可以构建出如图 3 所示的层级化标签结构。需要注意的是, 构建的所有类别结构空间实际上是一个森林结构, 其中包含不止一棵树。尽管这些树的根节点共享同一个虚拟的公共祖先, 但在实际分类任务中, 这个公共祖先不会直接出现在标签中。

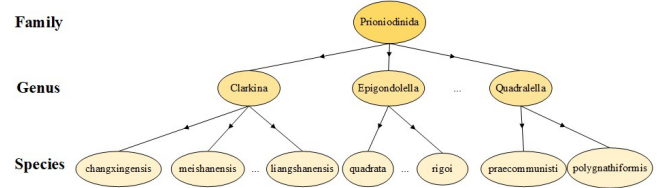


Fig. 3 Hierarchical label tree structure visualization

图3 分层标签树结构展示

在此基础上, 为了更好地利用数据集中多层次标签的层次信息, 本文设计了一种结合层级信息的多标签分类模型。该模型不仅能够充分利用粗粒度标签的全局语义信息, 还能够通过层级化的标签结构有效指导细粒度特征的学习, 从而提升分类性能。

3.2 模型架构

ConoHN 模型由骨干网络、层级特征提取模块、语义擦除模块、1 个特征聚合和分类模块组成, 如图 4 所示。主干网络用于从输入图像中提取特征, 同时引入了特征金字塔网络 (Feature Pyramid Network, FPN) 以融合不同尺度的特征^[37]。层级注意力特征提取模块包含对不同层级特征提取结构、不同层级共享相同结构、特征聚合模块聚合不同层级的特征, 通过以上步骤得到的特征通过分类模块实现最终预测。

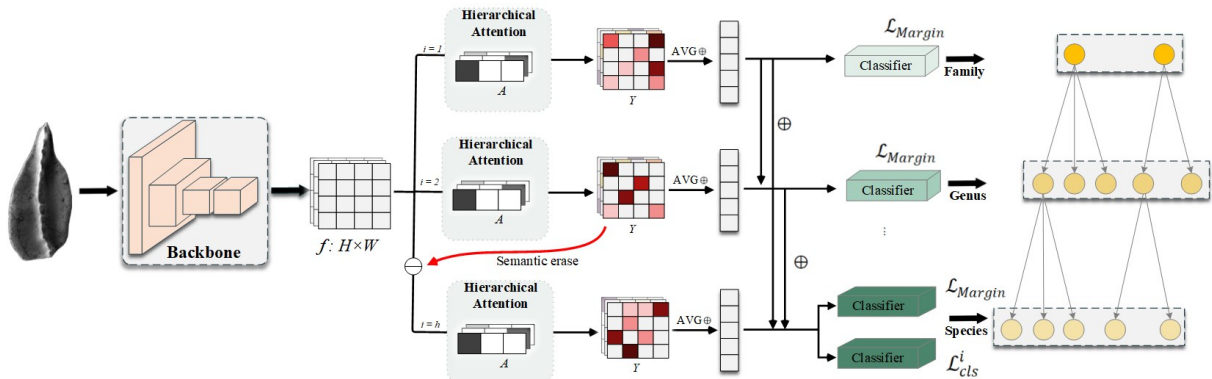


Fig. 4 ConoHN model architecture

图4 ConoHN 模型架构

该模型由主干网络 (ResNet-50)、分层特征交互模块和 4 个并行输出通道组成。前 3 个通道结点通过 Sigmoid 激活计算边缘概率损失 L_{margin} , 最后一个输出通道和最后一个层级的输出分类节点数量相同。网络结构和牙形石标签的 3 个层次级别相对应。层级注意力模块处理来自主干网络的每个分层级别特征图, 然后将这些分层特征通过残差连接交互, 即将父级别的特征作为跳层连接添加到子级别的特征中。

3.3 层级注意力特征提取模块

根据上述层级关系, 模型利用注意力机制提取每个层级的特征。该模块引入注意力机制, 在通道维度增加注意力机制, 通过自动学习的方式, 获取到特征图的每个通道的重要程度。之后, 利用该重要程度为每个特征赋予一个权重值, 从而让神经网络重点关注对当前任务有用的特征图通道, 并抑制对当前任务用处不大的特征通道。

经过骨干网络前向传播, 获得了一个特征图 $f \in \mathbb{R}^{C \times H \times W}$, 其中 C 表示通道数, H 和 W 分别表示特征图的

高度和宽度。通过全局平均池化(Global Average Pooling, GAP),将每个通道的二维特征 $H \times W$ 压缩为 1,将特征图从 $f \in \mathbb{R}^{C \times H \times W}$ 转换为 $f' \in \mathbb{R}^{C \times 1 \times 1}$,其公式如式(1)所示。

$$f' = \text{GAP}(f) = \text{AvgPool2d}(f) \quad (1)$$

通过两个全连接层(FC)构建通道间的相关性,使用 Sigmoid 函数给特征通道生成一组权重值 $A = [a_1, \dots, a_i, a_c]$,如式(2)所示,输出的权重值数目与输入特征图的通道数相同。

$$A = \sigma(\text{FC1}(\text{FC2}(f'))) \quad (2)$$

其中, $A \in \mathbb{R}^{C \times 1 \times 1}$ 。将前面得到的归一化权重 A 加权到 f 的 C 个通道特征上,得到加权后的特征图 Y ,其中 $Y \in \mathbb{R}^{C \times H \times W}$,如式(3)所示。

$$Y = A \cdot f \quad (3)$$

通过两个 FC 层和 RELU 激活,得到层级特征 Y' ,如式(4)所示。

$$Y' = \text{RELU}(\text{FC4}(\text{FC3}(\text{Avg}(y_i)))) \quad (4)$$

Y'_i 是 i^{th} 层级生成的特征图,每个层级都将共享该分支结构。

根据上述层级关系,模型利用注意力机制提取每个层级的特征。在通道维度增加注意力机制,通过自动学习的方式,获取到特征图每个通道的重要程度。之后,利用该重要程度为每个特征赋予一个权重值,从而让神经网络重点关注对当前任务有用的特征图通道,并抑制对当前任务用处不大的特征通道。

3.4 语义擦除模块

经过上述模块,模型提取到不同层级的特征,每个层级的特征都输入到对应层级的分类器中,以用于分配到不同层级的类标签。让模型分别提取属的特征和区分属内子类的特征要比直接提取区分所有类的特征更容易且具有可解释性。由于在属以及更高层级上的分类任务相较于种分类而言属于粗粒度分类,是一个较为简单的任务。换言之,模型可以很容易地学习到代表属的足够具有辨识度的表征。已经证明的是,有意擦除图像的高响应区域可以强制神经网络专注于其余语义部分。擦除部分相对应高层级特征区域以对抗学习,有助于找到难以区分的低层级显著性区域。该模块采用渐进式擦除策略,通过如下机制确保底层级特征的完整性:首先,擦除操作仅作用于高层级(属级)特征响应区域,使用属级别的语义特征 Y_2 对最底层的种层级进行对抗擦除,强迫网络发掘更具有辨识度的种层级特征;其次,引入特征正交性约束,确保不同层级的特征表示位于相互正交的子空间;最后,采用残差连接保留原始特征,使模型能够回退到未擦除的特征表示,以确保底层级特征不受影响。

经过层级特征提取器,得到属于不同层级语义的特征图 Y 和一组类加权向量集合 A 。具体而言,将分类器的输入特征在属层级分类器产生的可区分区域指导下进行部

分擦除,通过混合原始特征和从相邻层次结构中提取的正交分量以增强区域特征表示。通过以上过程,细粒度级别可通过在前向传播过程中找到与粗粒度级别相比的差异以增强观察区域的可辨性。

Y_i 是 i^{th} 层级加权后的特征图输入,将其由线性投影得到的 Sigmoid 输出 $S_i \in \mathbb{R}^n$,如式(5)所示。其中, $i \in \{1 \dots h\}$, n_i 是标签对应层级下类的数量。对于计算的每个 Sigmoid 节点对应于不同层次结构中的类标签。

$$S_i = \begin{cases} \sigma(Y_i), i = 1 \\ \sigma(Y_{i-1} \oplus Y_i), i = 2 \\ \sigma(Y_1 \oplus W_{i-1} \cdot f \oplus Y_i), i = 3 \end{cases} \quad (5)$$

其中,“ $\oplus$ ”表示 add 操作。

最后融合得到的总特征为 T ,在最低层级经过 Softmax 激活后送入分类器计算二值交叉熵损失 $\mathcal{L}_{CE}$ 。

$$T = Y_1 \oplus \dots \oplus Y_{h-1} \oplus Y_h \quad (6)$$

3.5 模型优化

计算联合概率以传递层级知识信息。给定一个图像 x ,其标记为 $L = (l_{j_1}, l_{j_2}, \dots, l_{j_h})$,采用式(7)计算 x 在第 m 层属于 l_{j_m} 的概率。

$$P(L[m] = l_{j_m} | x) = \prod_{m=1}^{n_m} e^{\phi(x, l_{j_m})} \quad (7)$$

其中, $\phi(x, l_{j_m})$ 是 x 在标签 l_{j_m} 下对应的经过 RELU 激活后的 Sigmoid 结点输出。计算边界概率 $P_m(\cdot)$ 公式如式(8)所示。

$$P_m(x \in l|x) = \frac{1}{N(x)} \sum_{m=1} P(L[m] = l_{j_m} | x) \quad (8)$$

其中, $N(x)$ 用于对 $P_m(\cdot)$ 归一化,公式如式(9)所示。

$$N(x) = \sum_{m=1} \sum_{i=1} P(L[m] = l_{j_m} | x) \quad (9)$$

对每一个样本计算边缘概率分类损失 $\mathcal{L}_{\text{margin}}(x)$,如式(10)所示。

$$\mathcal{L}_{\text{margin}}(x) = -\ln(P_m(x \in l|x)) \quad (10)$$

只有在最后一层得到的融合特征经过 Softmax 激活后得到的概率分布向量将会用于计算交叉熵损失 $\mathcal{L}_{CE}$ 。对于 t 个样本,最终损失结果定义如式(11)所示。

$$L = \mathcal{L}_{CE} + \frac{1}{t} \sum_{i=1}^t \mathcal{L}_{\text{margin}}(x_i) \quad (11)$$

4 实验与结果分析

4.1 实验设置

4.1.1 参数设置

模型的输入图像大小为 448×448,优化器使用随机梯度下降优化器 SGD,批量大小设置为 8,模型训练 50 个周期。在数据增强中采用了随机水平翻转和随机裁剪,但在测试阶段仅使用中心裁剪。在训练过程中,骨干网络的学习

习率设置为 0.001, HASM 和 SEM 模块的学习率初始化为 0.01。训练完成大约需要 5 h。

4.1.2 实验环境

在 Intel (R) Core (TM) CPU i7-10700 和 GPU GeForce RTX 3090 (24 GB 内存) 的服务器上对 AE-GERL 与其他基准模型进行了评估。实验环境基于 Ubuntu 18.04 和 CUDA 11.1。所有方法都在 Python 3.11 和 Pytorch 1.7.1 中实现。

4.2 评估指标

本文使用混淆矩阵 (Confusion Matrix) 和总体准确率 (Overall Accuracy, OA) 以评估模型性能。

4.2.1 总体准确率

准确率 (Accuracy) 定义的是模型整体预测结果与标签的一致性。本文将总体准确率 (OA) 作为主要评估指标, 用于衡量所提模型在分类任务中的性能。

总体准确率 (OA) 的正式定义为: 正确预测的实例数量 (包括真正类和真负类) 与总实例数量的比值。准确率可通过式 (12) 进行计算。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

其中, TP (True Positive) 表示标记为正类且被预测为正类的样本数量, FP (False Positive) 表示标记为负类但被预测为正类的样本数量, TN (True Negative) 表示标记为正类但被预测为负类的样本数量, FN (False Negative) 表示标记为负类且被预测为负类的样本数量。

4.2.2 混淆矩阵

除总体准确率 (OA) 外, 本文还采用了混淆矩阵 (Confusion Matrix) 作为关键的评估指标, 以更细致地分析模型在分类任务中的性能。混淆矩阵是一种用于描述分类模型性能的表格工具。它能够直观地展示模型预测结果, 包括真正类 (True Positives, TP)、真负类 (True Negatives, TN)、假正类 (False Positives, FP) 和假负类 (False Negatives, FN)。

4.3 结果分析

本文对比实验的基线模型如下:

(1) HSE^[35]。该模型将更高级别的预测分数向量作为先验知识以学习更细粒度的特征表示, 通过使用更高级别的预测分数向量作为相应子类别的软目标以规范标签预测。

(2) GLM-Loss^[38]。该模型使用 GLM 损失可以使给定的 DCNN 模型显式地探索细粒度图像类的层次结构和相似性规律。GLM 损失不仅可以减少学习特征类间相似性和类内方差, 还可以使同一粗略类别的子类在特征空间中更相似。

(3) MGE-CNN^[39]。该模型依次学习 3 个不共享参数的网络作为 3 个特定粒度的专家, 并使用 KL 散度最大化不同专家的决策区别。

(4) API-Net^[40]。该模型引入了一个专注的成对交互网络, 以捕获成对图像之间的对比线索, 并通过成对交互

以区分它们。

(5) PIM^[41]。该模型衡量每个像素的分类能力并以此为依据从骨干网络每层输出中选取特征点, 并通过图神经网络聚合每层的特征点对图像进行分类。

(6) TransFG^[42]。该模型通过自注意机制捕捉最具判别性的区域, 并引入对比损失以增强模型区分能力。

为了全面评估本文方法的性能, 将所提方法与当前最先进方法 (SOTA), 包括 HMC 方法和 FGVC 方法进行比较。此外, 本文还评估了使用不同主流骨干网络的分类性能, 将网络 ResNet-50^[33]、Swin-T^[34] 和 ViT^[35] 作为骨干网络, 并使用在数据集 ImageNet22k 上预训练的权重对其进行初始化。实验结果如表 2 所示。

Table 2 Comparison results of ConoHN against baseline models

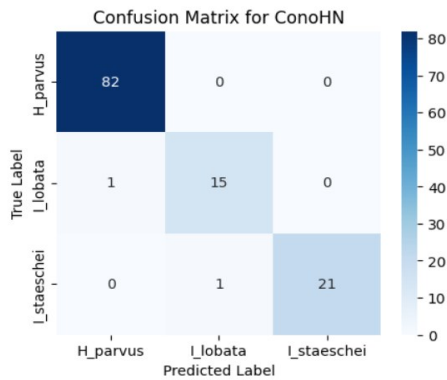
表 2 ConoHN 与基线模型对比结果

	Method	Backbone	Family	Genus	Specie ACC/
			ACC/%	ACC/%	OA /%
HMC	GLM-Loss	VGG	97.29	91.13	70.44
	HSE	ResNet-50	97.96	91.04	69.04
	HRN	ResNet-50	98.80	90.72	71.10
FGVC	MGE-CNN	ResNet-101	-	-	62.30
	API-Net	DenseNet-161	-	-	69.72
	PIM	Swin-t	-	-	67.60
ConoHN (Ours)	PMG	ResNet-101	-	-	52.50
	TransFG	TransFG	-	-	71.39
		ResNet-50	98.82(98.91)	93.76(94.17)	74.13(74.77)
		Swin-t	98.78	93.20	73.35
		Vit-B	98.20	92.10	72.76

在与 HMC 方法的比较中, 本文计算了不同层级的分类总体准确率。可以观察到, 所提出的方法在属和种层级上取得了更优结果。此外, 当骨干网络设置为 ResNet-50 时, 本文方法展示了更佳性能。具体分析如下: MGE-CNN^[39] 使用 3 个独立的骨干网络在每个层级提取特征; HSE^[32] 利用高级类别预测结果以引导低级分类; HRN 使用层级概率损失在不同层级之间转移知识。这些方法的 OA 分别为 70.44%、69.04% 和 73.10%。相比之下, 本文方法在所有层级的分类结果中均超越了这些方法, 其可能的原因在于本文方法充分利用了不同层级特征之间的关系。通过擦除容易识别的高级判别区域, 迫使模型关注其他区域, 从而在低层级发现判别区域, 最终实现了更好的性能。

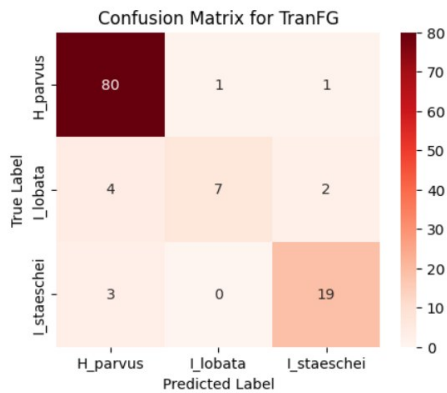
将本文方法与 FGVC 方法进行比较时, 仅在最细粒度层次上比较了总体准确率, 可以观察到 FGVC 方法中的主要误分类案例, 其主要原因在于缺乏层级知识。具体而言, 对于引入对比学习的方法, 如 TransFG^[42] 和 API-Net^[30], 它们通过对比学习以扩展细粒度类别的边界。在同一空间中, 当在相同层次观察时, 属于同一类别的样本应彼此靠近, 而来自不同类别的样本应相互远离。然而, 负样本选择的不同方式可能导致特征空间中表示分布的差异。根据 FGVC 方法 TransFG^[42] 的分类结果绘制了混淆矩阵, 如图 5 所示。本文 ConoHN 模型在所有物种上表现出强劲的性能, 且误分类更小, 突显了其在物种区分方面

的有效性。与TransFG模型相比,ConoHN模型在不同属下物种的误分类大幅减少,尤其是在*H.parvus*和*I.staeschel*这两个物种上。一个合理的解释是,这些对比学习方法仅在物种分类层次选择负样本,导致其表示没有按上述方式分布,从而导致分类性能较差。此外,本文方法在准确性方面显著超越了PMG方法。PMG方法旨在解决细粒度视觉信息在渐进训练学习中的集中问题,通过在不同层次的粒度中引入拼图生成器。然而,实验结果(模型准确率52.50%)表明,上述方法框架在牙形石图像分类中并未取得有益成果。



(a) ConoHN confusion matrices

(a) ConoHN 混淆矩阵



(b) TranFG confusion matrices

(b) TranFG 混淆矩阵

Fig. 5 Performance comparison of misclassified cases

图5 误分类案例性能比较

4.4 消融实验

本文开展消融实验,以探究所提出关键模块HAM和SEM对于模型效果的影响。其中,HASM代表层级注意力特征提取模块,SEM代表语义擦除模块。

如表3所示,对于本文数据集,仅使用骨干网络(Backbone Only)方法,表示仅使用ResNet-50作为分类网络,仅使用CE Loss作为损失函数优化网络,达到了69.26%准确率。在骨干网络之后添加层级注意力特征提取模块HAM,同时利用最大化层级边缘概率以优化分类过程,损失函数计算了Margin Loss。未加入粗粒度标签数据的时

候精度比Resnet-50增加了3.34%,证明了层级知识的引入对于提高分类精度的巨大贡献。在HASM之后添加语义擦除模块SEM,进一步帮助模型在最细粒度上区分同属子类,最终分类精度为74.13%。本文使用梯度回流生成了属和种层级的类激活图(CAM),如图6所示。其中,左1是随机选取的牙形石样本原图,左2是该样本属层级的注意力图可视化结果,左3和左4分别展示了不使用和使用特征擦除模块的可视化对比结果。可以看出,加入SEM后模型提取到的种特征更为精细,定位判别性区域更为准确。上述消融实验和可视化结果都验证了所提出模块的有效性。

Table 3 The results of ConoHN ablation experiment

表3 ConoHN 消融实验结果

Component	w/o coarse data/%	with coarse data/%
Backbone only	67.26	—
+HASM	73.10(+3.34)	73.68(+0.58)
+SEM	74.13(+1.03)	74.77(+0.57)

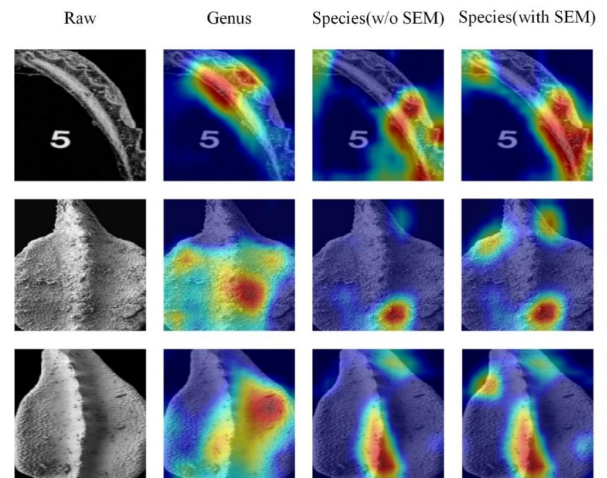


Fig. 6 Class activation map at the genus and species levels

图6 属和种层级的类激活图

4.5 样例研究

牙形石分类的关键是找到细微的鉴别特征,而ConoHN框架解决了以往基于注意力的细粒度图像分类方法存在层级特征信息利用不足等问题,总体取得了较好效果。本文统计了所提方法在不同种上的分类效果,如表4所示,并对部分种案例的典型错误进行了分析与总结。

*C.changxingensis*被错误分类为*C.yini*,同时*C.yini*也被错误分类为*C.changxingensis*。从形态特征来看,两者都是Clarkina属,在形态上有一定的相似度,其中*C.changxingensis*的典型特征是在齿台末端有一个明显且稍微倾斜的主齿。*C.changxingensis*由*C.yini*进化而来,前者特征相对明显和特化。

无法利用物种特有的属性对信息进行描述。例如,*H.praeparvus*被错误分类为*H.parvus*,两者都为Hindeodus属,形态具有一定相似性,同时*H.praeparvus*是*H.parvus*的先驱分子,*H.parvus*的后部末端形态继承于*H.praeparvus*。两

Table 4 Dataset classification results achieved by ConoHN**表 4 ConoHN 在数据集上的分类结果**

Family	Genus	Number/ OA/%	Specie	Number	OA/%
Gondolellidae	Clarkina	1 799/ 99.6	changxingensis	428	76.9
			yini	208	51.6
			subcarinata	194	66.6
			longicuspadata	141	90.9
			meishanensis	208	71.4
			orientalis	214	64.6
			postbitteri	148	95.5
			liangshanensis	113	64.7
	Epigondolella	415/98.4	wangi	145	81.8
			quadrata	180	64.2
			rigoi	104	68.7
			triangularis	131	65.5
	Neospathodus	557/93.9	dieneri	336	88.0
			pakistanensis	96	64.0
			waageni	125	78.9
	Quadralella	233/99	polygnathiformis	166	97.8
			praecommunisti	67	88.0
			changxingensis	110	88.2
			eurypyge	106	75.0
Anchignathodontidae	Hindeodus	1 461/ 91.3	latidentatus	93	57.1
			parvus	621	86.0
			praeparvus	332	50.0
	Isarcicella	496/81.8	typicalis	199	65.5
			isarcica	90	64.2
			lobata	90	68.0
			staeschei	188	78.6
			turgida	128	75.0
Novispathodontidae	Novispathodus	590/87.2	pingdingshanensis	160	79.1
			abruptus	216	81.8
			waageni	214	68.9
Overall	—	5 551/ 93.7	—	5551	74.1

者的形态差别表现在: *H.praeparvus* 的分子整体尺寸、细齿紧密程度大于 *H.parvus*, 主齿、细齿之间的大小差异小于 *H.parvus*。其中, 最明显的特征分子大小在图像上并不能表现出来。还有一些因素也可能会影响模型最终学习效果, 例如牙形石本身不完整, 导致某些重要识别特征丢失, 或者模型侧重学习了无关紧要的特征。此外, 不同视角下的图像也可能造成重要判别信息丢失, 牙形石图像主要分口视、反口视和侧视, 通常口视或者斜口视信息含量多, 当识别到侧视或者反口视时可能出现错误决策。

5 结语

本文提出了一种结合层级关系的多标签细粒度图像分类框架 (ConoHN), 通过充分利用粗粒度标注样本和细粒度样本之间的层级关系以优化细粒度分类任务。实验结果表明, ConoHN 能够通过粗粒度特征对细粒度类别进行有效指导, 尤其是在数据稀疏和类间差异较小的场景下, 显著提升了分类准确性和鲁棒性。该方法设计了层级

损失函数和特征擦除模块, 不仅有效利用了粗粒度样本中的全局语义信息, 还通过特征回流策略加强了细粒度特征的判别力。此外, 模型通过层级一致性约束确保了不同层级分类结果的逻辑一致性, 进一步提升了整体分类性能。需要指出的是, 当输入图像存在重要特征丢失或视角变化时, 本文模型性能仍会受到一定影响。

针对这些局限性, 未来工作可从如下两个方面加以改进: ①设计自适应特征补全机制, 通过生成对抗网络重构缺失的判别性特征; ②引入多视角特征融合模块, 聚合不同成像角度下的关键特征表示。这些改进将进一步提升模型在复杂场景下的适用性。本文研究为细粒度图像分类领域提供了新的思路, 层级关系引导的特征学习和多粒度语义融合机制的核心方法可扩展应用于其他具有层次化结构的分类任务。通过结合粗粒度和细粒度特征, ConoHN 为拥有复杂层级结构的数据分类提供了方法学参考, 也为其他细粒度图像分类任务提供了有益的实践探索。

参考文献:

- [1] TROTTER J A, WILLIAMS I S, BARNES C R, et al. Did cooling oceans trigger Ordovician biodiversification? evidence from conodont thermometry [J]. *Science*, 2008, 321: 550–554.
- [2] SUN Y D, WIGNALL P B, JOACHIMSKI M M, et al. Climate warming, euxinia and carbon isotope perturbations during the Carnian (Triassic) crisis in South China [J]. *Earth and Planetary Science Letters*, 2016, 444: 88–100.
- [3] KONG S, PUNYASENA S, FOWLKES C. Spatially aware dictionary learning and coding for fossil pollen identification [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016: 1305–1314.
- [4] HSIANG A Y, BROMBACHER A, RILLO M C, et al. Endless forams: > 34, 000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks [J]. *Paleoceanography and Paleoclimatology*, 2019, 34(7): 1157–1177.
- [5] YAYAN K, YAYAN U. A new model on benthic foraminifer image classification and definitions based on conventional neural network (CNN) [J]. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, 2023, 31(1): 481–490.
- [6] ELSAYED O R, ELKOT Y G, ELREFAAI D A, et al. Automated identification and classification of teeth fossils [C]//2023 International Mobile, Intelligent, and Ubiquitous Computing Conference, 2023: 179–186.
- [7] DUAN X. Automatic identification of conodont species using fine-grained convolutional neural networks [J]. *Frontiers in Earth Science*, 2023, 10: 1046327.
- [8] FU J, ZHENG H, MEI T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition [C]//Hawaii: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [9] MITRA R, MARCHITTO T M, GE Q, et al. Foraminifera optical microscope images with labelled species and segmentation labels [DB/OL]. <https://doi.org/10.1594/PANGAEA.897873>.
- [10] KEÇELİ A S, KEÇELİ S U, KAYA A. Classification of radiolarian fossil images with deep learning methods [C]//Izmir: 2018 26th Signal Processing and Communications Applications Conference, 2018.
- [11] LIU T Z, CHEN X, LI X, et al. Conodont image recognition based on deep residual neural network transfer learning [J]. *Acta Palaeontologica Sinica*,

- 2020,59(4):512-523.
- 柳天滋,陈昕,李想,等. 基于深度残差神经网络迁移学习的牙形刺图像识别[J]. 古生物学报,2020,59(4):512-523.
- [12] AN Y C, CHEN Y, HUANG Y N, et al. Hierarchical identification of ostracod fossils based on deep learning[J]. Geological Review, 2022, 68(2):673-684.
- 安玉钊,陈雁,黄玉楠,等. 基于深度学习的介形类化石层次化识别[J]. 地质论评,2022,68(2):673-684.
- [13] YUAN D X, SHEN S Z, HENDERSON C M, et al. Revised conodont-based integrated high-resolution timescale for the Changhsingian stage and end—Permian extinction interval at the Meishan sections, South China[J]. Lithos, 2014, 204:220-245.
- [14] ZHANG Z L, LI Y J, LIU Z Y, et al. Applications of deep learning in fine-grained image recognition: a survey[J]. Journal of Beijing University of Technology, 2021, 47(8):379-416.
- 张志林,李玉鑑,刘兆英,等. 深度学习在细粒度图像识别中的应用综述[J]. 北京工业大学学报,2021,47(8):379-416.
- [15] DUAN L J, GUO Y N, QIAO Y H, et al. Video-based human action recognition method using discriminative region extraction[J]. Journal of Beijing University of Technology, 2017, 43(10):1480-1487.
- 段立娟,郭亚楠,乔元华,等. 基于判别性区域提取的视频人体动作识别方法[J]. 北京工业大学学报,2017,43(10):1480-1487.
- [16] ZOU Z A, HE R H. Fine-grained apparel image retrieval based on cost-sensitive method[J]. Software Guide, 2023, 22(10):185-190.
- 邹子安,何儒汉. 基于代价敏感的细粒度服装图片检索[J]. 软件导刊,2023,22(10):185-190.
- [17] ZHANG Y, WEI X S, WU J, et al. Weakly supervised fine-grained categorization with part-based image representation[J]. IEEE Transactions on Image Processing, 2016, 25(4):1713-1725.
- [18] HE X, PENG Y. Weakly supervised learning of part selection model with spatial constraints for fine-grained image classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022:123-130.
- [19] WANG Z, WANG S, LI H, et al. Graph-propagation based correlation learning for weakly supervised fine-grained image classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020:12289-12296.
- [20] SILLA C N, FREITAS A A. A survey of hierarchical classification across different application domains[J]. Data Mining and Knowledge Discovery, 2011, 22:31-72.
- [21] CERRI R, BARROS R C, DE CARVALHO A C P L F, et al. Reduction strategies for hierarchical multi-label classification in protein function prediction[J]. BMC Bioinformatics, 2016, 17:373.
- [22] WAH C, BRANSON S, WELINDER P, et al. The Caltech-UCSD birds-200-2011 dataset[J]. International Journal of Computer Vision, 2011, 91(2):153-198.
- [23] VAN HORN G, AODHA O, SONG Y, et al. The iNaturalist species classification and detection dataset[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:8769-8778.
- [24] KRAUSE J, DENG J, STARK M, et al. Collecting a large-scale dataset of fine-grained cars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(11):2392-2405.
- [25] MAJI S, RAHTU E, KANNALA J, et al. Fine-grained visual classification of aircraft[DB/OL]. <https://arxiv.org/abs/1306.5151>.
- [26] ZHOU B W, GAO J. Feature-reordered weighted deep forest[J]. Software Guide, 2021, 20(9):7-13.
- 周博文,皋军. 特征重排序的加权深度森林[J]. 软件导刊,2021,20(9):7-13.
- [27] SONG Y T, YU B G. Multi-label classification by fusing multi-granularity features and label semantic co-occurrence[J]. Science Technology and Engineering, 2023, 23(16):6959-6966.
- 宋宇婷,余本功. 融合多粒度特征和标签语义共现的多标签分类[J]. 科学技术与工程,2023,23(16):6959-6966.
- [28] CHEN J H, YU S S, LIANG J. Fine-grained image classification method based on convolutional semantic enhancement[J]. Software Guide, 2024, 23(3):142-149.
- 陈建华,余松森,梁军. 基于卷积语义增强的细粒度图像分类方法[J]. 软件导刊,2024,23(3):142-149.
- [29] JIAN Y S, ZHU D M, FU Z T, et al. Remote sensing image segmentation network with multi-level feature optimization fusion[J]. Laser & Optoelectronics Progress, 2023, 60(4):422-428.
- 管永胜,朱大明,付志涛,等. 多层次特征优化融合的遥感图像分割网络[J]. 激光与光电子学进展,2023,60(4):422-428.
- [30] DENG J, DING N, JIA Y, et al. Large-scale object classification using label relation graphs[C]//Proceedings of the European Conference on Computer Vision, 2014:48-64.
- [31] GIUNCHIGLIA E, LUKASIEWICZ T. Coherent hierarchical multi-label classification networks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020:9662-9673.
- [32] CHANG D, PANG K, ZHENG Y, et al. Your “flamingo” is my “bird”: fine-grained, or not[C]//Nashville: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [33] CHEN J, WANG P, LIU J, et al. Label relation graphs enhanced hierarchical residual network for hierarchical multi-granularity classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(5):1234-1245.
- [34] XU Z, YUE X, LYU Y, et al. Trusted fine-grained image classification through hierarchical evidence fusion[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2023:10657-10665.
- [35] CHEN T, LIN L, CHEN R, et al. Knowledge-embedded representation learning for fine-grained image recognition[C]//Stockholm: The Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018.
- [36] WANG R, ZOU C, ZHANG W, et al. Consistency-aware feature learning for hierarchical fine-grained visual classification[C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023:2326-2334.
- [37] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Hawaii: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [38] SHI W, GONG Y, TAO X, et al. Fine-grained image classification using modified DCNNs trained by cascaded softmax and generalized large-margin losses[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(3):683-694.
- [39] ZHANG L, HUANG S, LIU W, et al. Learning a mixture of granularity-specific experts for fine-grained categorization[C]//Seoul: 2019 IEEE/CVF International Conference on Computer Vision, 2019.
- [40] ZHUANG P, WANG Y, QIAO Y. Learning attentive pairwise interaction for fine-grained classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020:13130-13137.
- [41] CHOU P Y, LIN C H, KAO W C. A novel plug-in module for fine-grained visual classification[DB/OL]. <https://arxiv.org/abs/2202.03822>.
- [42] HE J, CHEN J N, LIU S, et al. TransFG: a transformer architecture for fine-grained recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022:852-860.

(责任编辑:孙娟)