面向噪声标签数据的细粒度图像识别鲁棒微调

一、赛题背景

近年来，以 CLIP 为代表的视觉基础模型在开放域图像理解任务中展现出了优异的迁移能力与泛化能力，成为视觉智能领域的重要技术基础。然而，在细粒度图像识别等需要较强专业知识支撑的下游任务中，基础模型的零样本能力往往仍然不足，难以直接满足实际应用对准确性和鲁棒性的要求，因此需要通过任务适配与模型微调进一步释放其潜力。

传统微调方法通常依赖大规模、精确标注的数据，但在自然动植物等细粒度场景中，高质量人工标注往往需要较强领域知识支撑，标注成本高、获取周期长，严重制约了基础模型在专业场景中的落地应用。与之相比，来源广泛、规模庞大的网络图像数据为模型微调提供了新的可能，但网络数据不可避免地包含大量标签噪声，并可能伴随类别分布不均衡、样本难度差异显著等问题。若直接利用此类数据对基础模型进行微调，容易导致模型对噪声样本过拟合，甚至破坏基础模型原有的预训练知识，引发“灾难性遗忘”。

因此，如何在标签含噪条件下对视觉基础模型进行安全、有效、鲁棒的微调，降低对精确人工标签的依赖，并兼顾精度、泛化性与稳定性，已成为当前视觉智能与机器学习交叉领域的重要研究课题。本赛题旨在引导参赛者围绕这一前沿问题开展探索，推动鲁棒微调技术在真实复杂场景中的发展与应用。

二、赛题应用场景

面向标签含噪数据的视觉基础模型鲁棒微调技术，在多个真实场景中具有重要应用价值。例如，在生态监测与生物多样性保护中，需要对大量自然动植物图像进行精细分类，但人工标注成本极高；在农业生产中，需要基于含噪的田间图像识别作物品种、病虫害类型；在智慧林业、自然资源调查、科普教育等领域，也普遍面临“数据海量但标签不准”的问题。

通过研究鲁棒微调算法，可以充分利用低成本网络数据，将视觉基础模型快速适配到专业细粒度识别任务中，在降低数据采集与标注成本的同时，提升模型在真实场景中的可用性、稳定性与推广能力。

三、赛题任务

本赛题聚焦于标签含噪条件下的视觉基础模型鲁棒微调。参赛者需基于赛事方提供的细粒度图像训练数据，围绕自然动植物类别识别任务，设计并训练高效、稳健的细粒度图像识别模型。

为保证评测聚焦于“微调算法”本身的创新性与公平性，参赛者必须严格使用 CLIP 的 ViT-B/32 作为骨干网络，在此基础上设计适用于含噪标签场景的鲁棒微调方

法。参赛者可围绕参数高效微调、噪声标签学习、鲁棒优化、样本筛选、正则化约束、灾难性遗忘抑制等方向开展研究。

最终，赛事方将使用类别分布均衡、经人工精确标注的测试集对模型性能进行评估，重点考察参赛方法在含噪训练数据条件下的识别精度、泛化能力与鲁棒性。

四、数据集及数据说明

（一）数据来源

赛题数据集由赛事方构建，面向自然动植物细粒度图像识别任务，训练数据主要来源于互联网公开图像资源，经整理后形成含噪标签训练集。测试集由赛事方组织人工精确标注，用于统一评测。

（二）数据情况概要

本赛题分为初赛、复赛、半决赛和总决赛四个阶段，其中初赛、复赛、半决赛为线上赛，总决赛为线下答辩与复核评审。

1. 初赛数据集：包含 500 个细粒度类别，103218 个训练样本以及 24967 个测试样本，训练数据包含标签噪声。

2. 复赛数据集：包含 1500 个细粒度类别，297282 个训练样本以及 74896 个测试样本，训练数据包含标签噪声，并存在长尾分布。

3. 半决赛数据集：包含 1000 个细粒度类别，180274 个训练样本以及 49857 个测试样本，训练数据包含标签噪声，并存在长尾分布。

三个阶段的数据类别均以自然动植物为主，任务形式均为单标签细粒度图像分类。训练数据按类别文件夹组织，测试集提供待预测图像列表，标签不公开。

（三）数据特点

1. 标签含噪：训练集标签来源于网络数据，存在错误标注、弱相关标注及难样本干扰。

2. 细粒度性强：部分类别间视觉差异较小，对模型表征能力和判别能力要求高。

3. 长尾分布：复赛与半决赛阶段存在明显类别不均衡问题，进一步考验模型对少样本类别的适应能力。

4. 规模较大：赛题训练数据总体规模较大，对算法设计、训练效率和工程实现能力提出较高要求。

（部分数据可能由于图片文件部分截断问题导致在文件系统中出现无法正常显示的情况，但是所有数据经试验确认，均可被 python 的 pillow 库进行正常读取。）

五、算法设计要求

1. 参赛者必须严格使用 CLIP 的 ViT-B/32 作为骨干网络，不得替换为其他视觉基础模型或更大规模模型。

2. 在 CLIP ViT-B/32 基础上开展鲁棒微调算法设计，包括但不限于 Prompt

Tuning、Adapter、LoRA 等参数高效微调方法，以及鲁棒损失函数、噪声样本过滤、伪标签优化、表征约束等技术。

3. 鼓励参赛者重点关注以下方向：在含噪标签条件下抑制错误监督带来的负面影响；最大程度保留和利用 CLIP 预训练先验知识；提升模型在不同阶段数据集上的泛化能力与稳定性。

4. 为保证竞赛公平性，最终提交方案不得采用多模型集成。最终评测结果应来自单一模型或单一推理流程。

5. 提交到总决赛的代码须可复现。若赛事方无法基于提供的代码、环境和赛事数据复现主要实验结果，将取消相应成绩。

6. 为确保模型训练的完整可复现性，整个训练过程（包括噪声筛选等环节）必须能够通过最终提交的代码直接复现。因此，人工数据清洗等手段仅可作为算法开发过程中的辅助参考，不应成为最终模型训练方案的必要前置环节。如需引入数据清洗环节，建议采用自动清洗或筛选策略，以实现对噪声样本的自动甄别。

六、性能指标要求

本赛题以图像分类准确率（Accuracy）作为线上阶段的主要评价指标：

Accuracy = 测试集中预测正确的样本数量 / 测试集样本总数量。

线上各阶段将分别基于对应测试集计算准确率，并用于排行榜展示与晋级评定。为避免过度依赖单一阶段结果，赛事方将综合复赛与半决赛成绩计算线上综合分数，重点衡量算法在更复杂含噪场景下的稳定表现。

注：上述指标专指 Top-1 Accuracy，每张测试图像均只有一个真值标签。具体提交文件格式要求说明详见“十三、提交要求”。

七、功能要求

参赛者提交的解决方案应至少具备以下功能：

1. 能够在赛事方提供的含噪训练集上完成模型训练或微调。

2. 能够基于官方测试集生成规范的类别预测结果。

3. 具备一定的跨阶段适应能力，在不同类别规模、不同噪声水平和长尾分布条件下保持较好的性能表现。

八、开发环境

（一）软件环境

参赛者需使用 Python 语言进行开发，可使用 PyTorch 等主流深度学习框架。允许使用公开可获取的 CLIP ViT-B/32 实现与相关训练工具，但需保证最终方案可复现。

（二）硬件环境

建议使用具备 GPU 加速能力的计算设备开展训练与推理。由于赛题数据规模较大，推荐使用具有较大显存的 GPU 环境完成模型训练、参数调优与结果验证。

九、成绩评价

1. 初赛主要用于参赛者熟悉任务、验证方案与完成晋级，不计入最终线上综合成绩。

2. 复赛与半决赛成绩共同构成线上综合成绩，其中复赛线上成绩占 40%，半决赛线上成绩占 60%。

3. 总决赛延续“客观评测 + 线下答辩”相结合的评价方式。线上综合成绩作为客观评分的重要依据，线下答辩重点考察方法创新性、技术完整性、复现质量与表达展示效果，最终成绩以组委会公布的决赛细则为准。

4. 若参赛者提交结果低于赛事方公布的基线成绩，赛事方有权将其认定为无效成绩。

十、解题思路

（一）考核知识点

1. 数据挖掘与处理能力：从海量网络图像中识别有效样本、分析噪声特征，并开展数据清洗、样本重加权或难例挖掘。

2. 基础模型微调能力：重点考察参赛者在标签含噪条件下对 CLIP 大模型进行鲁棒优化的能力，包括参数高效微调、特征约束、鲁棒损失设计及灾难性遗忘抑制等。

3. 创新思维：鼓励参赛者突破传统全参数监督微调范式，提出新颖的面向含噪数据的视觉基础模型微调算法。

4. 实践与工程能力：考察参赛者在大规模细粒度识别任务中的训练组织、算力分配、参数调优与实验复现能力。

（二）思路引导

参赛者可从“数据”和“模型”两个层面联合优化。一方面，可针对噪声标签、类别不均衡和难样本问题，设计数据过滤、样本选择、置信度建模、课程学习或重标注机制；另一方面，可围绕 CLIP ViT-B/32 设计更稳健的微调策略，例如仅更新少量参数、增加正则化约束、约束表征漂移，或结合教师模型、自训练等方法增强鲁棒性。

（三）注意事项

1. 不宜直接采用无约束的全参数微调，以免标签噪声导致预训练知识退化。

2. 需特别关注复赛与半决赛中的长尾分布问题，避免模型过度偏向头部类别。

3. 应充分考虑不同阶段数据集之间的差异，提升方法的跨阶段泛化能力。

十一、 赛题约束条件

（一）算法约束

1. 必须使用 CLIP ViT-B/32 作为骨干网络。

2. 不得使用其他视觉基础模型、商业闭源模型 API 或在线大模型推理接口替代核心识别流程。

3. 不得使用多模型集成、模型融合或多个独立模型投票等方式提升成绩。。

（二）数据使用约束

1. 参赛者每个赛事阶段仅可使用赛事官方提供的当前阶段的数据集进行训练、验证和测试。需注意：测试数据仅可用于测试，不得以自监督、无监督等方式参与到模型训练过程中。

2. 不得在训练过程中引入任何形式的额外数据集（包括公开的、私有的、或人工补充标注的等）。这里的限制包括：复赛期间不得使用初赛的数据，半决赛期间不得使用初赛和复赛的数据。

3. 可以使用 CLIP ViT-B/32 的预训练模型权重，但仅限 OpenAI 官方公开的预训练权重（如 openai/clip 库或 HuggingFace Transformers 库中的对应版本）。

4. 赛事数据仅限用于本次比赛，严禁泄露、传播或用于与赛事无关的用途。

十二、参考资源

1. Alec Radford, Jong Wook Kim, Chris Hallacy, et al. Learning Transferable Visual Models From Natural Language Supervision. ICML, 2021.

2. Kaiyang Zhou, Jingkang Yang, Chen Change Loy, Ziwei Liu. Learning to Prompt for Vision-Language Models. IJCV, 2022.

3. Edward J. Hu, Yelong Shen, Phillip Wallis, et al. LoRA: Low-Rank Adaptation of Large Language Models. ICLR, 2022.

4.Yuncheng Guo and Xiaodong Gu. JoAPR: Cleaning the lens of prompt learning for vision-language models. CVPR. 2024.

5. Xueyi Zhang, Peiyin Zhu, Yuan Liao, et al. TrustCLIP: Learning from Noisy Labels via Semantic Label Verification and Trust-aligned Gradient Projection. ACM MM. 2025.

6. 近年来关于学习含噪标签、鲁棒优化、样本筛选、表征一致性约束及灾难性遗忘抑制等方向的代表性论文。

十三、提交要求

（一）初赛提交要求

参赛者需在官方提供的初赛测试集上进行预测，并提交 CSV 格式结果文件。每行包含两个字段：图片文件名、类别编号。类别编号统一使用四位数字表示，不足四位前补 0。输出格式示例如下：

xxxxxxxxxxxx.jpg, 0001

xxxxxxxxxxxy.jpg, 0123

xxxxxxxxxxxz.jpg, 0456

注意事项：

（1）请确保提交的 CSV 文件符合上述格式规范，否则可能导致提交结果无效。

（2）图片文件名需与测试集中的文件名保持完全一致，包括大小写和扩展名。

（3）生成的预测结果文件应保证命名为“pred_results.csv”，随后压缩为一个 zip 文件进行提交。

（二）复赛提交要求

复赛阶段提交格式与初赛一致。参赛者需在规定时间内提交复赛测试集预测结果，作为复赛线上评分依据。

（三）半决赛提交要求

半决赛阶段提交格式与前两阶段一致。半决赛成绩将作为线上综合成绩的重要组成部分。提交作品包含：

1. 可复现的代码与运行说明文档；
2. 完整的训练、验证、推理脚本及环境配置说明；
3. 可执行模型文件或容器化复现环境；
4. 技术方案文档（PDF），内容包括算法动机、方法设计、关键模块、实验分析等；

（三）总决赛提交要求

提交内容及具体要求以组委会后续正式通知为准。

十六、其他说明

公平性：严禁任何形式的作弊行为，包括但不限于数据泄露、模型预训练数据与测试数据重叠、抄袭他人代码等。一经发现，立即取消参赛资格，并追究相关责任。

知识产权：参赛者提交的作品必须为原创，未在其他比赛中获奖或公开发表。比赛主办方有权对参赛作品进行展示、宣传等相关活动，但知识产权仍归参赛者所有。

十七、联系方式

赛题交流 QQ 群：1090224462

邮　箱：zerens@njust.edu.cn

报名官网：www.aicomp.cn