

网络监督细粒度图像识别

一、赛题背景

细粒度图像识别是计算机视觉和模式识别领域的重要分支，其目标是对物体的子类别进行精确识别。例如，在犬类分类任务中，区分“哈士奇”和“阿拉斯加”这两类犬种。由于子类别间往往类间差异较小，而类内差异较大，使得细粒度识别相比传统图像分类更具挑战性。近年来，深度学习驱动的机器学习算法在多个领域取得了显著进展。这些成功广泛依赖于大规模高质量的数据标注。这种依赖性在细粒度图像识别任务中尤为明显，因为该任务通常需要专业领域知识进行精确标注。然而，人工标注大量高质量标签不仅困难而且成本高昂，限制了细粒度识别模型在实际应用中的可扩展性。因此，使用免费的网络数据训练细粒度识别模型吸引了越来越多研究者的关注。然而，网络数据通常存在标签噪声、数据偏差、长尾分布等问题，拥有强大拟合能力的深度网络利用此类数据进行训练时往往出现显著过拟合问题，无法在测试数据上获得良好的泛化表现。

本赛题希望通过网络监督细粒度识别模型的开发，探索在噪声数据环境下提升模型稳健性的方法，并利用网络数据训练高效的细粒度识别模型，以减少对大规模人工标注数据的依赖，同时提升算法的实用性和可扩展性。

二、赛题应用场景

网络监督细粒度图像识别技术作为计算机视觉领域的前沿方向，凭借其强大的数据驱动能力和对细节特征的精准识别，在学术界和产业界均展现出巨大的发展潜力。该技术通过有效利用网络爬取的低质量数据进行鲁棒训练，不仅显著降低了对高成本人工标注数据的依赖，还大幅提升了模型的泛化能力和鲁棒性，为实际应用场景中的技术落地提供了可靠保障。

网络监督细粒度图像识别技术不仅在学术研究中具有重要的价值，也在实际应用中展现出广泛的潜力。从物种监测到智能安防，从医学影像诊断到自动驾驶，这项技术为各个领域带来了显著的效率提升和成本节约。参赛者通过解决本赛题，将能够深入了解网络监督细粒度识别技术的挑战与应用，推动这一技术在现实世界中的发展，并为实际问题提供创新的解决方案。

三、赛题任务

本赛题聚焦于网络监督的细粒度图像识别，参赛者需利用赛题方提供的细粒度网络图像训练数据集，训练高效的细粒度识别模型。赛事方提供的训练集中的数据均来源于网络爬取，包含标签噪声、数据分布不均衡、类别间差异小等现实挑战。参赛者需充分利用这两个数据集，通过优化数据处理、模型架构、训练策略等手段，提高模

型在细粒度分类任务中的表现。由于网络数据存在标签噪声，如何优化数据质量，是本赛题的重要考察点。同时，模型需具备优秀的泛化性，能够在不同的数据集上均保持良好的识别性能。

四、数据集及数据说明

（一）数据来源

数据来源于南京理工大学公开数据集 WebFG-496 和 WebiNat-5089，是基于网络图像的细粒度识别数据集，类别覆盖飞机、汽车、鸟类以及各类自然物种。训练数据均爬取自网络，标签未经人工核查校验，因此具有较强的现实代表性。

（二）数据情况概要

细粒度网络数据集 WebFG-496：包含了飞机、汽车、鸟这 3 个大类下的 496 个子类，其中包括 200 种鸟类（Web-Bird）、100 种飞机（Web-Aircraft）和 196 种汽车模型（Web-Car），共 5 万多张训练图像。

细粒度网络数据集 WebiNat-5089：包含自然物种中的 5089 个子类别，共 110 多万张网络训练图像，是有史以来最大的网络监督细粒度数据集，其类别涵盖多种元类别，如植物、昆虫纲、鸟类、爬行动物、真菌、原生动物、软体动物、动物等。

在实际赛题中，赛事方对 WebFG-496 和 WebiNat-5089 中的类别进行了采样，分别随机选取其中 400 类和 5000 类的数据，进行数据脱敏后作为最终的比赛数据 WebFG-400 以及 WebiNat-5000。为了进行评估，赛事方将使用均匀分布的人类精确标记的数据集作为测试集进行精度比较。最终，WebFG-400 含有 43087 个训练数据及 13882 个测试数据，WebiNat-5000 含有 580865 个训练数据及 100000 个测试数据。训练数据将按照类别文件夹的形式给出（即类别 0001 内的数据都在 0001 文件夹内）。示例数据可从 <https://pan.baidu.com/s/1I1oXGinmXlqM78-514bS7g?pwd=dih6> 获取，正式数据将在报名后开放下载。

WebFG-400 和 WebiNat-5000 的数据将被划分为初赛数据（A 榜）和复赛数据（B 榜）进行分阶段比赛评估。

五、算法设计要求

鼓励参赛者提出创新性的深度学习算法或者改进现有算法，提升模型在使用网络数据进行训练时的准确性和鲁棒性。算法应具备良好的可扩展性，能够在处理不同规模的数据时保持性能稳定。

竞赛过程中若使用预训练模型，只能采用在 ImageNet-1k 上预训练的模型，不能使用额外数据。为鼓励算法创新，本次赛题不允许使用多模型集成（model ensemble），最终提交的结果应仅包含一个模型。复赛结束后，赛事方将对各参赛队伍提交的代码进行复现，若使用赛事方提供的数据集无法复现性能，将取消成绩。

六、性能指标要求

本竞赛性能得分以图像分类识别任务中普遍使用的测试准确率（Accuracy）作为评价指标。本次竞赛对模型推理时间及模型大小不做限制。测试准确率定义如下：

$\text{Accuracy} = \text{测试集中所有预测正确的样本数量} / \text{测试集中的所有样本的总数量}。$

七、功能要求

参赛者需要设计创新性方法或改进现有算法，利用赛事方提供的网络数据训练集训练细粒度识别网络模型，最终在赛事方提供的测试数据集（经过人工精确标注）上表现出良好的准确性和鲁棒性。参赛者设计的方法需要在赛事方提供的不同数据集上有良好的泛化性和通用性。

八、开发环境

参赛者需要使用 Python 语言进行开发，可以使用开源的算法框架，如 Pytorch 等。

九、成绩评价

本赛题的赛事分为初赛、复赛和决赛三个阶段。

初赛测试集由机器自动评分（Accuracy），但不计入决赛总分，仅用于参赛者前期算法验证与调试。算法性能得分取为算法在 WebFG-400（A 榜子集）和 WebiNat-5000（A 榜子集）这两个数据集子集上的性能得分均值。

复赛测试集由机器自动评分（Accuracy），结果计入决赛总分。算法性能得分取为算法在 WebFG-400（B 榜子集）和 WebiNat-5000（B 榜子集）这两个数据集子集上的性能得分均值。

线下决赛最终成绩的计算方式为：70%客观评分+30%主观评分。这里客观评分指的是通过标准化处理后的复赛机器评测得分，而主观评分则是答辩分数经过标准化处理后的得分。答辩分数考量内容包括参赛者的答辩表现以及参赛费提供的技术方案及代码文档等。标准化处理的方式是将最高分设定为 100 分，其余各分数按照比例进行调整（国二、国三成绩不涉及 30%主观评分部分）。

十、参考资源

Mengmeng Sheng, Zeren Sun, Gensheng Pei, Tao Chen, Haonan Luo, Yazhou Yao, “Enhancing Robustness in Learning with Noisy Labels: An Asymmetric Co-Training Approach”, ACM International Conference on Multimedia (ACM MM), 2024.

Mengmeng Sheng, Zeren Sun, Tao Chen, Shuchao Pang, Yucheng Wang, Yazhou Yao, “Foster Adaptivity and Balance in Learning with Noisy Labels”, European Conference on Computer Vision (ECCV), 2024.

Zeren Sun, Yazhou Yao, Xiu-Shen Wei, Yongshun Zhang, Fumin Shen, Jianxin Wu, Jian Zhang, and Heng Tao Shen, “Webly Supervised Fine-Grained Recognition: Benchmark

Datasets and An Approach”, IEEE International Conference on Computer Vision (ICCV), 2021.

Chuanyi Zhang, Yazhou Yao, Xing Xu, Jie Shao, Jingkuan Song, Zechao Li, Zhenmin Tang, “Extracting Useful Knowledge from Noisy Web Images via Data Purification for Fine-Grained Recognition”, ACM International Conference on Multimedia (ACM MM), 2021.

Zeren Sun, Xian-sheng Hua, Yazhou Yao, Xiu-shen Wei, Guosheng Hu, Jian Zhang, “CRSSC: Salvage Reusable Samples from Noisy Data for Robust Learning”, ACM International Conference on Multimedia (ACM MM), 2020.

十一、提交要求

（一）初赛提交内容及要求

参赛方需要在赛事方提供的初赛测试集上进行预测，并将预测结果保存为一个 CSV 格式的文件，在该 CSV 文件中，每行包含两个元素，第一个元素是图片文件名，第二个元素是类别名（四位数字，不满四位的前面补 0）。输出格式参考如下：

```
xxxxxxxxxxxx.jpg, 0000
xxxxxxxxxxxxxy.jpg, 1111
xxxxxxxxxxxxxz.jpg, 0812
.....
```

赛事方依据该结果文件自动对参赛者的模型性能进行性能指标分数计算。

注意事项：

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。

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，否则可能会出现匹配错误的情况。

类别名需要保证四位数字，不满四位的前面补 0，否则可能会出现匹配错误的情况。

（二）复赛提交内容及要求

参赛选手需在官方提供的复赛测试集上进行预测，输出格式和初赛相同。

（三）决赛提交内容及要求

可复现的 docker 文件，包括但不限于：

1. 完整的训练代码和验证代码，包括复现脚本
2. 代码、环境、操作说明文档

技术方案文件（PDF）：要求包含算法设计动机，方法说明，伪代码等内容辅助证明所设计的方法的创新性和有效性；

答辩 PPT。

（具体内容参考《算法赛题作品提交要求》进行组织和提交。）

十二、其他说明

（一）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。

（二）知识产权

1. 竞赛仅允许使用比赛提供的数据集。为了维护竞赛公平性，严禁使用其他公开或私有的数据集。复赛结束后，赛事方会对选手提交的代码进行训练复现，若使用赛事方提供的数据集无法复现结果，将取消竞赛成绩。

2. 参赛者提交的作品必须为原创。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。

3. 参赛团队提交算法及可执行模型的知识产权归双方所有，竞赛视频、图像数据由南京理工大学所有。经协商，参赛队伍的优胜算法可以用于与本次比赛经费相关的相关研究中。

4. 各参赛队需要承诺本队提交的结果可复现，组织方承诺履行保密义务，并不用于除本比赛外的任何其他用途。

5. 参赛队伍应保证所提供的方案、算法属于自有知识产权。组织方对参赛队伍因使用本队提供/完成的算法和结果而产生的任何实际侵权或者被任何第三方指控侵权概不负责。一旦上述情况和事件发生参赛队伍必须承担一切相关法律责任和经济赔偿责任并保护组织方免于承担该等责任。

十三、联系方式

赛项交流 QQ 群：731220935

邮 箱：zerens@njust.edu.cn

报名官网：www.aicomp.cn

附录、比赛流程及奖项设置

（一）报名阶段

参赛者在比赛官方网站上完成报名注册，提交个人或团队信息，获取初赛数据下载链接。

（二）初赛阶段

参赛者利用赛事方提供的训练数据集进行算法模型设计，利用提供的初赛测试集进行相应方法的验证与调试。初赛阶段参赛者每天提交结果的次数不限，但是初赛排行榜每隔 1 小时刷新一次。

（三）复赛（省赛）阶段

初赛结束后进入复赛阶段，开放复赛数据下载链接。仅有初赛阶段提交有效结果的参赛团队可以进入复赛。复赛期间，参赛者利用赛事方提供的复赛阶段数据进行算法模型调试，提交对复赛测试数据的推理结果。复赛阶段持续 3 天，每个参赛队伍每天仅能提交 2 次。复赛排行榜每隔 1 小时刷新一次。

（四）复赛（省赛）成绩公布

在比赛官方网站上公布复赛成绩。以进入复赛参赛团队数量作为计奖基数，按照不超过大赛省赛设奖比例，评选出复赛一、二、三等奖（颁发省赛获奖证书）。评选复赛奖过程中，参赛者提交的算法性能低于赛事方提供的基线参考分数的判定为无效成绩，不予授奖。复赛一、二等奖晋级参加国赛总决赛。

（五）决赛（国赛）阶段

1. 决赛线上评选。晋级决赛的参赛团队，依据复赛排行榜结果，以进入决赛参赛团队数量作为计奖基数，按照不超过大赛国赛设奖比例，评选出国赛一等奖候选名单及国赛二、三等奖获奖名单（颁发国赛二、三等奖证书）。

2. 决赛作品提交。国赛一等奖候选团队在规定时间内提交技术文档、算法代码和模型文件、演示视频、补充材料等。提交截止后，不再接受任何形式的修改和补充。

3. 决赛审核阶段。由专业评审团队对国赛一等奖候选参赛团队的参赛作品进行结果复现与审核。评审过程中如有疑问，可要求参赛者进行解释说明。

4. 决赛线下答辩。国赛一等奖候选团队在规定时间内提交完善后的技术文档、算法代码和模型文件、演示视频、补充材料，参加国赛线下总决赛复核答辩，最终依据算法性能得分和线下答辩得分确定国赛一等奖获奖名单及其排名（未参加线下复核答辩视同放弃奖项）。国赛一等奖颁发荣誉证书。