
上海理工大学 具身智能赛事规则

初心高于胜负 成长胜于输赢

上海理工大学机器人创新社

2026 年 1 月

一、比赛背景介绍

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、比赛应用场景

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、比赛任务

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、数据集说明

赛题采用公开数据集 **ImageNet-1k**, 这是 **ImageNet** 最常用的子集。该数据集跨越 1000 个对象类, 包含 1,281,167 个训练图像、50,000 个验证图像和 100,000 个测试图像。数据集可从 <https://opendatalab.com/OpenDataLab/ImageNet-1K> 获取。

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。

竞赛过程中不可以使用预训练模型, 只能在 **ImageNet-1k** 上训练模型, 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。比赛结束后, 赛事方将对各参赛队伍提交的代码进行复现, 若赛事方在 **ImageNet-1k** 上无法复现性能, 将取消成绩。

算法参考方案:

ResNet-RS - <https://arxiv.org/abs/2103.07579>

ConvNeXt-V2 - <http://arxiv.org/abs/2301.00808>

GhostNet-V2 - <https://arxiv.org/abs/2211.12905>

EfficientViT (MIT) - <https://arxiv.org/abs/2205.14756>

MobileViT-V2 - <https://arxiv.org/abs/2206.02680>

StarNet - <https://arxiv.org/abs/2403.19967>

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:

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

七、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。

八、成绩评价

比赛成绩的计算方式为: 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 **Accuracy** 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者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 100 分, 其余各分数按照比例进行调整。

九、提交要求

(一) 预测结果

参赛方需要在赛事方提供的测试集上进行预测,并将预测结果保存为一个 CSV 格式的文件并自行计算 Accuracy。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:

xxxxxxxxxxxxx.jpg, 0000

xxxxxxxxxxxxxy.jpg, 1111

xxxxxxxxxxxxxz.jpg, 0812

赛事方将检查依据该 CSV 文件计算 Accuracy 是否正确。

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

(二) 其它文档

可复现的 docker 文件,包含但不限于:

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(具体参考见附录);

答辩 PPT。

十、命题组与裁判组

命题组:罗云、赵胤君、吴晨亮、李荣炜

裁判组:罗云、赵胤君、吴晨亮、李荣炜

十一、参赛人员要求

参赛范围:上海理工大学全日制本科生

参赛人数:每支参赛队伍限 2 至 5 人,每人限加入 1 支队伍 指导老师:每队至多 2 指导老师(可以选择无指导老师)

附录：

技术方案大纲

一、算法概述

简要阐述算法作品目标、算法主要特点、创新性、应用成效和应用价值等。语言简洁明了,重点突出。概述字数控制在 100-300 字。

二、实现方案

详细说明解题思路、数据处理、算法设计与开发、模型训练与优化、测试与验证等内容。

三、算法创新点

说明算法的主要特点和创新性,如采用的独特模型架构、创新的数据处理方法、算法优化等。

四、问题与思考

开发过程中遇到的困难、问题以及对应解决方案,进一步优化思考等内容。

五、过程进度

使用表格形式说明从开始到结束的各个阶段,包括数据准备阶段、算法设计与开发阶段、模型训练与优化阶段、测试与验证阶段、文档撰写与提交阶段等,明确每个阶段的开始时间、结束时间以及主要完成的内容。

六、团队分工

列出团队成员的姓名、角色(如算法工程师、数据分析师、文档撰写员等)以及具体负责的工作内容,如算法设计、模型训练、数据处理、文档编写等。

七、解题参考

解题过程中用到的解题资源、参考资料等。