




树叶分类竞赛总结

 InClass Prediction Competition

Classify Leaves

Train models to predict the plant species

165 teams · 6 days ago

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Host](#)

Overview

Description

+ Add Page

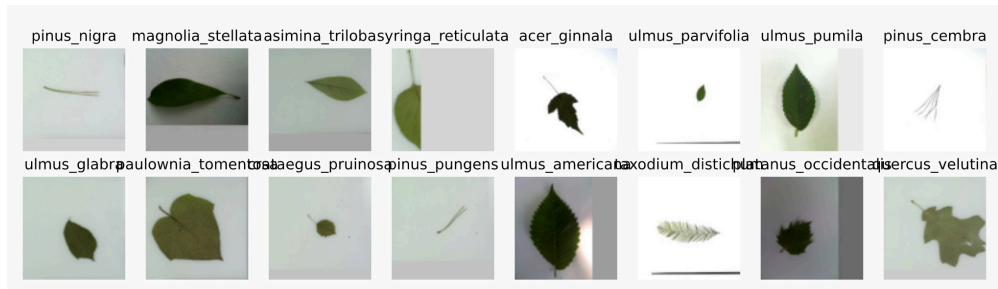
The task is predicting categories of leaf images. This dataset contains 8800 test images. Each category has at least 5 images in the public and private leaderboard.

The evaluation metric for this competition is [Classification Accuracy](#).

比赛结果



- 176类， 18353训练样本



- 165只队伍参加
 - 41只队伍精度 > 98% （非常好）
 - 83只队伍精度 > 95% （够用）



结果分析

- 16只队伍提供了代码
 - <https://www.kaggle.com/c/classify-leaves/code>
- 私榜排名1-5, 7-9, 11-14, 16-20将获得签名书
 - 请将你微信号邮件发给 mli@amazon.com
- 额外加上 Neko Kiku
 - 因为很多人用了你的代码:)



- 相比于课程介绍的代码，同学们主要做了下面这些加强
 - 数据增强，在测试时多次使用稍弱的增强然后取平均
 - 使用多个模型预测，最后结果加权平均
 - 有使用10种模型的，也有使用单一模型的
 - 训练算法和学习率
 - 清理数据



数据方面

- 有重复图片，可以手动去除
- 图片背景较多，而且树叶没有方向性，可以做更多增强
 - 随机旋转，更大的剪裁
- 跨图片增强：
 - Mixup：随机叠加两张图片
 - CutMix：随机组合来自不同图片的块



模型方面

- 模型多为ResNet变种
 - DenseNet, ResNeXt, ResNeSt, ...
 - EfficientNet
- 优化算法多为Adam或其变种
- 学习率一般是Cosine或者训练不动时往下调



- 15行代码，安装加训练耗时100分钟
 - <https://www.kaggle.com/zhrshold/autogluon-vision-0-96-with-15-lines>
- 精度96%
 - 可以通过定制化提升精度
 - 下一个版本将搜索更多的模型超参数
 - AG目前主要仍是关注工业界应用上，非比赛

总结



- 提升精度思路：根据数据挑选增强，使用新模型、新优化算法，多个模型融合，测试时使用增强
- 数据相对简单，排名有相对随机性
- 在工业界应用中：
 - 少使用模型融合和测试时增强，计算代价过高
 - 通常固定模型超参数，
而将精力主要花在提升数据质量