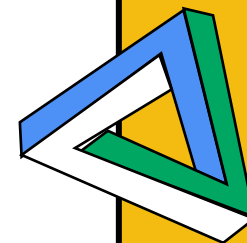


2026/03/08

实验一：做一名“数据厨师”—— 揭秘AI模型背后的秘密

周宇文



CONTENTS

目录

01 | 实验背景

02 | 实验平台

03 | 实验任务

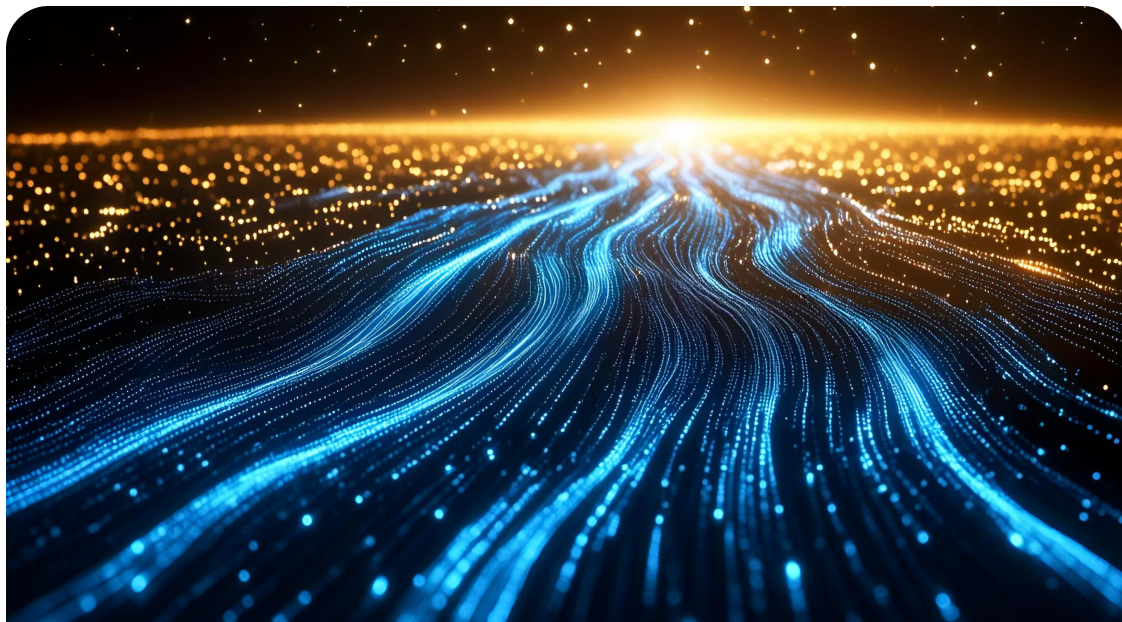
04 | 实验步骤



01

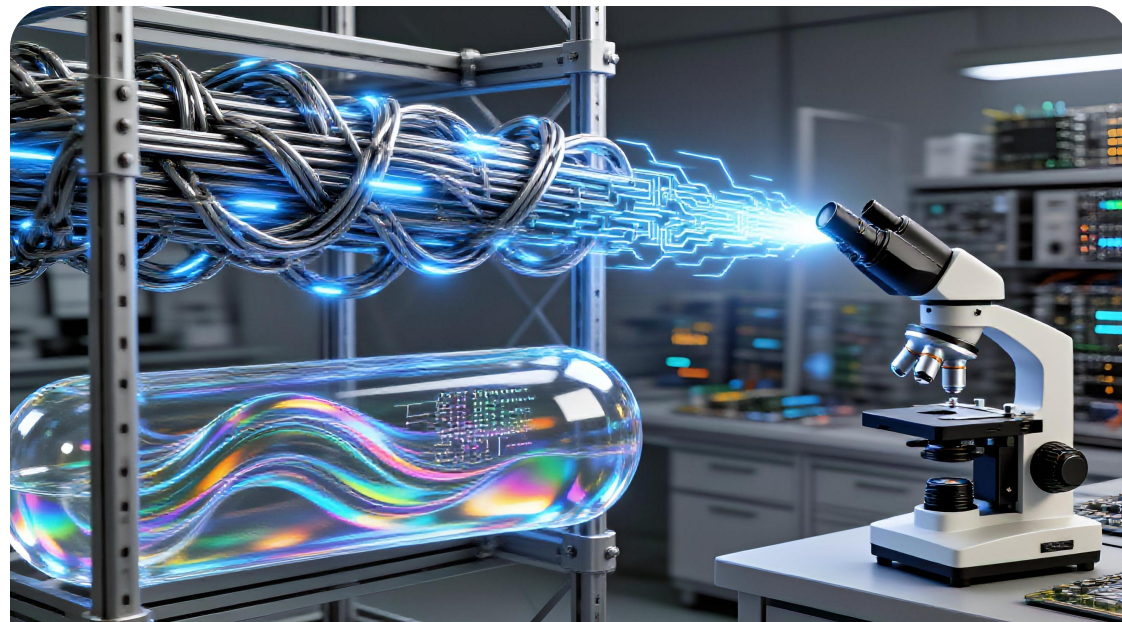
【实验背景】

数据决定AI模型质量



AI三要素

数据如食材，算法似菜谱，算力像灶台，共烹AI佳肴。



实验设计

利用EasyDL，亲自动手，组间攻防，探索数据对模型影响。



02

【实验平台】

百度智能云 EasyDL图像分类

EasyDL零门槛AI开发平台

[立即使用](#)[使用文档](#)

EasyDL入选Forrester领导者象限

零算法基础定制
高精度AI模型

- 【了解】EasyDL简介 
- 【上手】新手教程 > NEW
- 【实践】场景范例 > NEW
- 【购买】平台定价 >

AI场景范例实战

零基础AI建模，30分钟快速落地应用

[立即前往](#)

百度智能云 EasyDL (图像分类)



03

【实验任务】

AI模型识别生活物品挑战赛

训练一个能识别生活中常见物品的AI模型，通过分组对抗测试，体验数据量与数据质量对模型决定性的影响





04

实验步骤

第一阶段：厨房初探——平台与环境准备

千帆大模型平台已正式面向企业和个人开放

平台，为零基础基础的开发者提供定制高精度AI模型的服务，包括数据处理、模型训练、服务管理、模型部署功能模块。
适用场景：适合有多个目标主体，或要识别目标位置及数量的场景。常见应用场景：视频监控、工业检测、零件计数



数据处理

数据上传、标注、训练到回流
[在线标注](#)



2 模型训练

提供丰富的训练方式，零代码轻松获得高精度模型
[模型训练](#)

提供详细的模

登录与创建

登录百度智能云，进入“EasyDL控制台”，点击【图像分类】->【创建模型】，模型名称为学号_物品识别。

确定识别对象

以小组为单位，从身边随手可得物品中确定3种识别对象，建议水杯/水瓶、教科书/笔记本、笔袋/文具盒，且邻组使用相同3种物品以便后续PK。



第二阶段：分组对抗——食材数量大比拼



分组约定

相邻实验小组结为“对抗组”，约定识别三类物品。左侧极简组每类拍5-10张，右侧严谨组每类拍30-50张（含不同角度、光照、背景）。



模型训练

按照各自的数据量配置，点击“训练模型”，等待训练完成（约需5-10分钟）。



测试集制作环节

制作“刁钻”测试集关键环节：拍摄5张极度刁钻照片作考题攻击对方，如特写局部、杂乱背景、逆光昏暗、遮挡物品等，保存待测试。



攻防PK

极简组与严谨组用同一刁钻测试集测试，记录模型准确率，思考数据量多更“抗揍”的原因及说明。



第三阶段：查漏补缺——食材质量大优化

难题攻克

将测试中识别错误的图片补充进训练数据集，如同考前看错题本，针对性补习弱项。

开启“数据增强”

在训练配置中开启“数据增强”功能（含自动旋转、裁剪、调节亮度），原理是将一份数据变为多份加工品，增加机器见识的可能性。

二次验证

重新训练模型，并再次让邻组进行测试。观察准确率是否提升？

第四阶段：实验思考 与法学场景迁移

请结合今天的实验体验与法学专业背景，完成以下思考

实验反思

在刚才的对抗中，你的模型在哪些“刁钻角度”下最容易识别错误？这体现了数据采集的什么原则？



第四阶段：实验报告与法学场景迁移：场景迁移：当AI遇见法律

AI图像分类在司法应用

01

识别法庭文件中关键信息，如案件编号、当事人姓名，提升文档处理效率，减轻人工负担。

数据采集方案

02

从公开法律数据库和文档中提取图像样本，包括判决书、起诉书等，确保数据多样性，涵盖不同法律领域。

场景一：知识产权保护（商标侵权判定）

03

训练AI模型自动识别商品图片中Logo是否为侵权高仿品，需收集正品Logo高清图、已知侵权山寨Logo图、不同角度和材质的Logo变形图。

场景二：卷宗与证据数字化管理

04

律师事务所希望AI自动区分合同原件、传票/法院文书、证据照片，需解决AI对手机拍摄的倾斜、模糊文件的识别问题。

场景三：法庭秩序维护

05

智慧法庭系统需识别庭审违规，面临手机数据多样性设计挑战，数据决定模型上限，需高质量无偏见数据训练。

THE END
谢谢