

第十届“泰迪杯”数据挖掘挑战赛——

A 题 农田害虫图像识别

一、问题背景

病虫害一直是农业生产中无法避免的问题，每年都会由此造成巨大的经济损失。为了对农田病虫害进行有效的预防和控制，需要收集有害生物信息，在此基础上进行虫情分析。由于农田害虫的多样性和信息类型的复杂性，通过人工观察统计的传统害虫监测方式已经难以满足现代大规模农业生产对虫害预防工作的需要。近年来出现的虫情测报灯是虫情信息采集的智能设备，可以在无人监管的情况下，实现自动诱集、杀虫、虫体分散、拍照等作业，并实时地将虫情信息上传至云平台。虫情测报灯的投入使用可帮助植保人员高效地进行虫情分析，提高测报工作效率和准确率，避免农药的滥用和误用，减少农产品的农药残留，改善农田生态环境。

本赛题提供虫情测报灯采集到的图像文件，用于识别对某作物有危害的 28 种害虫，具体害虫种类见表 1。

表 1 害虫种类名称及其编号

编号	0	6	7	8	9	10
名称	无（不含虫子）	大螟	二化螟	稻纵卷叶螟	白背飞虱	褐飞虱属
编号	25	41	105	110	115	148
名称	地老虎	蝼蛄	粘虫	草地螟	甜菜夜蛾	黄足猎蝽
编号	156	222	228	235	256	280
名称	八点灰灯蛾	棉铃虫	二点委夜蛾	甘蓝夜蛾	蟋蟀	黄毒蛾
编号	310	387	392	394	398	401
名称	稻螟蛉	紫条尺蛾	水螟蛾	线委夜蛾	甜菜白带野螟	歧角螟
编号	402	430	480	485	673	
名称	瓜绢野螟	豆野螟	石蛾	大黑鳃金龟	干纹冬夜蛾	

二、解决问题

1. 利用附件 1 和附件 2 的信息，建立确定害虫位置及种类的模型和算法。
2. 应用问题 1 所建立的模型和算法对附件 3 中提及的图像进行识别，并确定害虫的位置，将结果存放在“result2.csv”文件中（模板文件见附件 4 中的 result2.csv）。
3. 根据问题 2 得到的结果，对附件 3 中提及的图像文件中不同种类的害虫数量进行统计，将统计结果存放在“result3.csv”文件中（模板文件见附件 4 中的 result3.csv）。

三、附件说明

附件 2-4 是 csv 文件，采用 ANSI 编码。

附件 1：虫情测报灯采集到的图像文件。

附件 2：附件 1 中部分图像的识别信息，其中左上角坐标和右下角坐标表示识别出的害虫包围框的两个角点（见图 1 中绿色矩形的蓝色顶点和橙色顶点），中心点是指矩形包围框的中心点（见图 1 中绿色矩形的红色中心点）。

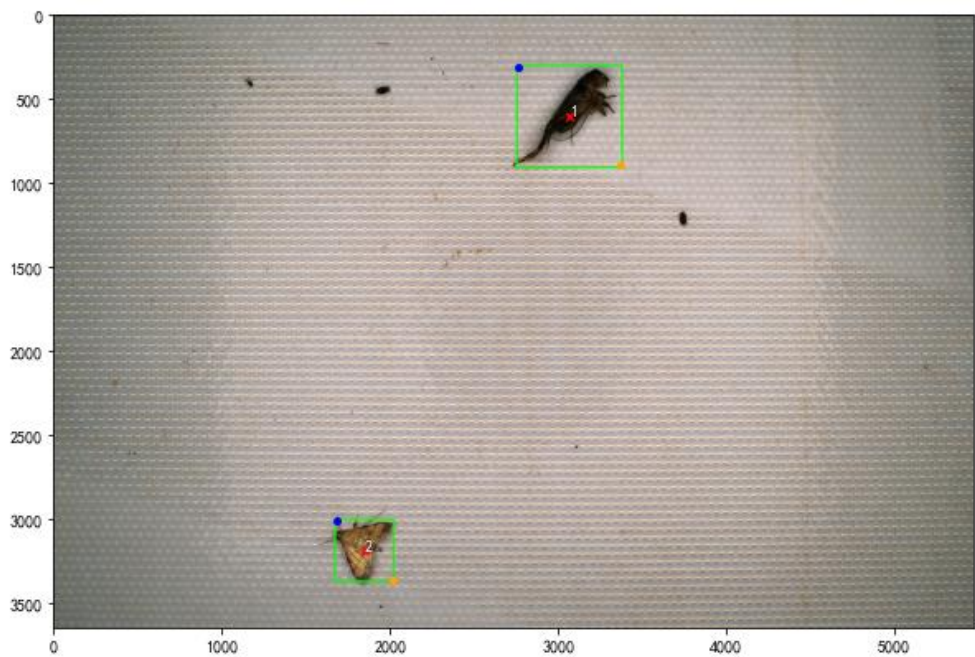


图 1 害虫坐标点标记图

附件 3：附件 1 中未识别的图像文件名。  
 附件 4：问题 2 和问题 3 结果文件的模板文件，具体字段名和样例见表 2 和表 3。

表 2 问题 2 的结果样例

序号	文件名	虫子编号	中心点 x 坐标	中心点 y 坐标	左上角 x 坐标	左上角 y 坐标	右下角 x 坐标	右下角 y 坐标
1	04001.jpg	256	2209	591	2036	397	2382	786
2	04001.jpg	256	2413	2710	2170	2344	2656	3077
3	04001.jpg	6	4180	2046	4033	1806	4328	2286
4	04001.jpg	156	2976	600	2732	392	3221	809
5	04002.jpg	0						
6	04003.jpg	156	323	1621	85	1921	1921	561
7	04004.jpg	156	1501	183	1330	4	1673	362
...	...	...	...	...	...	...	...	...

表 3 问题 3 的结果样例

序号	文件名	虫子编号	数量
1	04001.jpg	256	2
2	04001.jpg	6	1
3	04001.jpg	156	1
4	04002.jpg	0	0
5	04003.jpg	156	1
6	04004.jpg	156	1
...	...	...	...