**Project Two: chrysanthemum Classification**

1. **数据集**

数据集总共包含5种菊花，类别分别亳（bo）、滁（chu）、贡（gong）、杭（hang）、槐（huai），以文件夹命名划分为训练集（training，每类100张）、验证集（validation，每类40张）和测试集（testing，30张）。在训练集和验证集中，图片命名格式为“class\_id.png”，如“bo\_161.png”；在测试集中，图片命名格式为“id.png”，如“1.png”。示例如下：

* 训练集和验证集示例：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| bo\_1.png | chu\_101.png | gong\_201.png | hang\_301.png | huai\_401.png |

* 测试集示例：



2.png

1. **任务**

对给定的数据集，结合所学知识，采用合适的图像处理技术，包括预处理、特征提取，结合分类器实现菊花分类。要求模型在训练集上训练后，在验证集中验证模型性能，使用最终的模型预测出测试集中每张图片的类别。

可调用现有的工具包，如scikit-learn、openCV等，但是不允许使用深度学习模型。

1. **提交内容**

提交内容包含了实验报告和测试集的预测结果。实验报告中要求体现模型在训练集和验证集上的准确率（如，95.01%）。测试集的预测结果通过csv格式提交，命名为“result.csv”，每行为“id.class”，如“1.bo”。将结果存入csv的代码可参考附录。

**报告要求**：

* 英文撰写，无字数限制。
* 需要体现分析过程，阐述具体的实现步骤，以合理的方式展示、分析实验结果。

**提交要求：**

* 可自由分组，每组不超过4人，鼓励单独完成。
* 包含代码（含注释）、报告、以及在测试集上的预测结果“result.csv”。
* 报告以及文件夹的命名方式为：“学号+姓名”，提交zip、rar格式压缩包。
* Deadline: 8 May, 2021

**评分标准：**

* 主要参考预测准确率、算法流程、报告内容及代码。

**写入csv参考代码附录（python）：**

**import** csv  
  
  
**def** write\_csv():  
 path = **"result.csv"  
 with** open(path, **'w'**, newline=**''**, encoding=**'utf8'**) **as** f:  
 csv\_write = csv.writer(f)  
  
 id = [**'1'**, **'2'**]  
 *# your results* pred = [**'\*\*\*'**, **'\*\*\*'**]  
  
 **for** pair **in** zip(id, pred):  
 csv\_write.writerow(pair)  
  
write\_csv()