

# 实验 1 去除无效文件

实验时长：4 小时

实验难度：一般

**实验摘要：**对于从网络下载的图片，经常会有图像无法打开，图像文件无效的情况。在做初步的清洗时，应该首先删除这类文件。

当前任务下，有通过网络爬取获得的一系列大熊猫的图片，放在 download 目录下。需要通过脚本进行清理，去除无法正常打开的文件，包括动图、下载不完全的图等。去除灰度图像，只保留彩色图。去除尺寸小于 200\*200 像素的图像。最后，把保留的图像全部转换成 jpg 格式，并按照顺序命名。

**实验建议：**了解 os 和 opencv 库的相关知识。

**实验目标：**能够通过相关 Python 库完成去除无效文件的任务。

## 1、去除无效文件

- 

### 1.1、遍历文件

- 

创建 python 文件，可以命名为 image\_preprocess.py(文件名自行设置)。首先，引入相关的 python 库，因为包括文件操作和图像操作，需要 os 和 opencv 库。

- 

```
import cv2
```

- 

```
import os
```

- 

- 

需要整理的图像文件都在目录 download\_data 下。首先定义要读取的文件路径，再通过 os.listdir 函数得到目录下所有文件的列表，查看文件的数量。

- 

```
ROOT_PATH = './download_data'
```

- 

```
files = os.listdir(ROOT_PATH)
```



