删除多余文件

导入工具包

In [2]:
```python
import os
import cv2
from tqdm import tqdm
# import imageio
```

准备样例数据集

In [2]:
```python
# 下载测试数据集压缩包
!wget https://zihao-openmmlab.obs.cn-east-3.myhuaweicloud.com/20220716-mmclassificat
```

删除gif格式的图像文件

In [3]:
```python
dataset_path = r'C:\Users\baisichang\img_classification\img_cla\1_building_dataset\d
```

In [7]:
```python
# for fruit in tqdm(os.listdir(dataset_path)):
#     for file in os.listdir(os.path.join(dataset_path, fruit)):
#         file_path = os.path.join(dataset_path, fruit, file)

#         img = cv2.imread(file_path)
#         print(img)
#         # "cv2.imread可以获取jpg,png这类正常图片,但gif就不行
#         if img is None:
#             print(file_path, '读取错误，删除')
#             os.remove(file_path)
```

In [13]:
```python
# windows
from PIL import Image

for fruit in tqdm(os.listdir(dataset_path)):
    for file in os.listdir(os.path.join(dataset_path, fruit)):
        file_path = os.path.join(dataset_path, fruit, file)
        # print(file_path)
        # img = Image.open(file_path)
        # print(img)
        # print(img.format)
        if img.format == 'GIF':
            print(file_path, '读取错误，删除')
            img.close()
            os.remove(file_path)
```

```
100%|████████████████████████████████████████
███████████████████████████████████████████
██| 3/3 [00:00<00:00, 423.32it/s]
```

删除非三通道的图像

In [16]:
```python
import numpy as np
from PIL import Image
```

```python
for fruit in tqdm(os.listdir(dataset_path)):
    for file in os.listdir(os.path.join(dataset_path, fruit)):
        file_path = os.path.join(dataset_path, fruit, file)
        img = np.array(Image.open(file_path))
        # print(img.shape)
        # print(img.shape)
        try:
            channel = img.shape[2]
            if channel != 3:
                print(file_path, '非三通道，删除')
                os.remove(file_path)
        except:
            print(file_path, '非三通道，删除')
            os.remove(file_path)
```

```
100%|████████████████████████████████████████████████████████
████████████████████████████████████████████████████████████
████| 3/3 [00:03<00:00,  1.00s/it]
```

In [ ]: