**题目**：基于 ResNet50 的水果分类

**背景**：使用基于卷积的深度神经网络 ResNet50 对 30 种水果进行分类

**任务**

> 1. 划分训练集和验证集
> 2. 按照 MMPreTrain CustomDataset 格式组织训练集和验证集
> 3. 使用 MMPreTrain 算法库，编写配置文件，正确加载预训练模型
> 4. 在水果数据集上进行微调训练
> 5. 使用 MMPreTrain 的 ImageClassificationInferencer 接口，对网络水果图像，或自己拍摄的水果图像，使用训练好的模型进行分类
>
> 需提交的验证集评估指标（不能低于 60%）

- ResNet-50

```
Epoch(val) [5][34/34]    accuracy/top1: 82.4675  data_time: 0.0038  time: 0.0880
```

**作业数据集下载：**
链接：https://pan.baidu.com/s/1YgoU1M_v7ridtXB9xxbA1Q
提取码：52m9

**课程中猫狗数据集下载地址：**
https://download.openmmlab.com/mmclassification/dataset/cats_dogs_dataset.tar

## 对数据集进行划分

```python
import os

# 获取数据集文件夹路径
CustomDatasetPath = r'D:\workspace\datasets\fruit30_train'
# 获取数据集文件夹下的所有文件
CustomDatasetFile = os.listdir(CustomDatasetPath)
# 如果文件夹中不存在train、val、test文件夹，则创建
dataset_type = ['train', 'val', 'test']
for type in dataset_type:
    if type not in CustomDatasetFile:
        os.mkdir(os.path.join(CustomDatasetPath, type))
    else:
        # 清空文件夹
        os.removedirs(os.path.join(CustomDatasetPath, type))

# 遍历所有文件
for fruit_name in CustomDatasetFile:
```

```
    for type in dataset_type:
        os.mkdir(os.path.join(CustomDatasetPath, type, fruit_name))
    # 水果文件夹路径
    fruit_path = os.path.join(CustomDatasetPath, fruit_name)
    # 获取水果文件夹下的所有文件
    fruit_file = os.listdir(fruit_path)
    # 将水果文件夹下的所有文件分为训练集、验证集、测试集
    train_file = fruit_file[:int(len(fruit_file)*0.8)]
    val_file = fruit_file[int(len(fruit_file)*0.8):int(len(fruit_file)*0.9)]
    test_file = fruit_file[int(len(fruit_file)*0.9):]
    # 将训练集、验证集、测试集分别放入对应文件夹
    for file in train_file:
        os.rename(os.path.join(fruit_path, file), os.path.join(CustomDatasetPath, 'train',
fruit_name, file))
    for file in val_file:
        os.rename(os.path.join(fruit_path, file), os.path.join(CustomDatasetPath, 'val',
fruit_name, file))
    for file in test_file:
        os.rename(os.path.join(fruit_path, file), os.path.join(CustomDatasetPath, 'test',
fruit_name, file))
    # 删除空文件夹
    os.removedirs(fruit_path)
```

## 配置文件

见MMPretrain下**resnet50_8xb32_in1k_fruit_classify.py**

## 验证集评估指标

```
06/06 22:26:30 - mmengine - INFO - Epoch(val) [100][55/55]    accuracy/top1: 67.8899
accuracy/top5: 92.8899  data_time: 0.0008  time: 0.0159
```

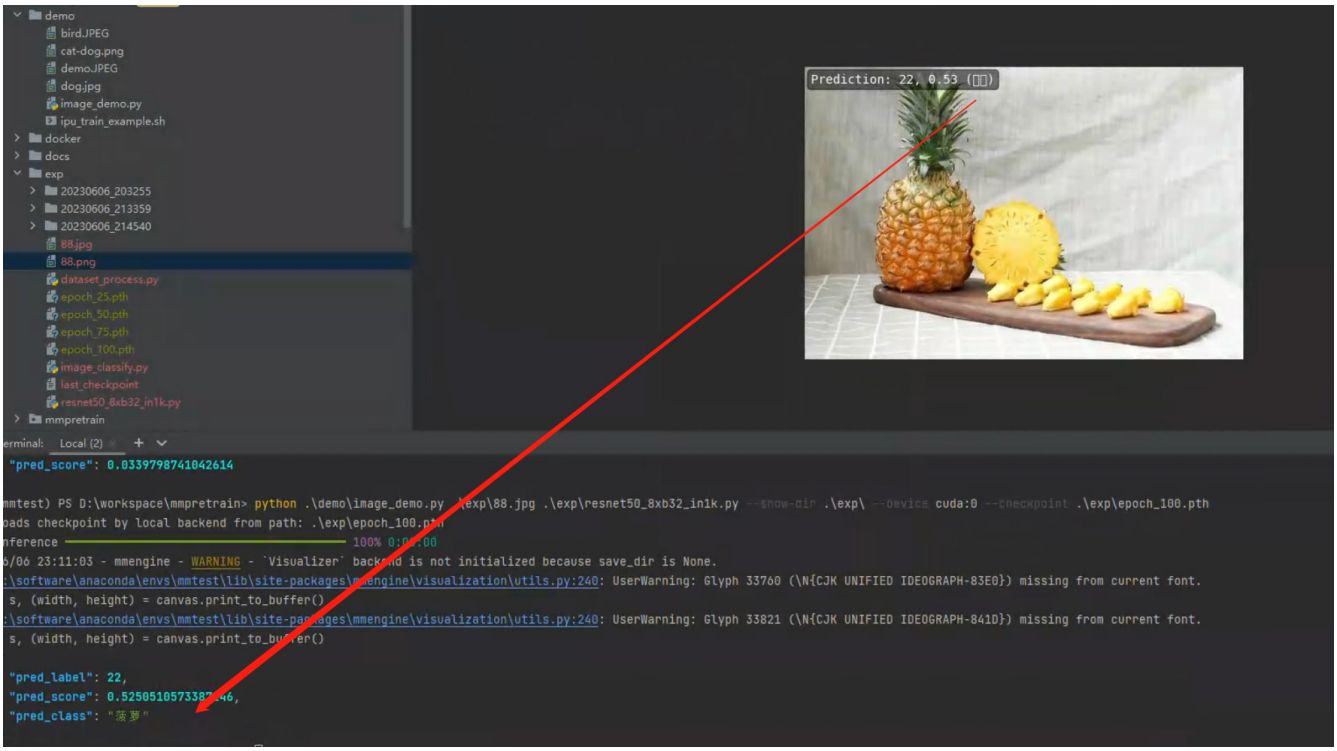
可以看出验证集的评估指标较低，原因，初始学习率设置过大，收敛困难，将学习率调小后验证集评估指标如下

```
2023/06/08 22:15:35 - mmengine - INFO - Epoch(val) [50][28/28]    accuracy/top1: 94.4954
accuracy/top5: 99.0826  data_time: 0.0253  time: 0.0380
```

达到了较好的94.4954

使用ImageClassificationInferencer 接口进行分类


Prediction: 22, 0.53 (□□)


demo
bird.JPEG
cat-dog.png
demo.JPEG
dog.jpg
image_demo.py
ipu_train_example.sh
docker
docs
exp
20230606_203255
20230606_213359
20230606_214540
88.jpg
88.png
dataset_process.py
epoch_25.pth
epoch_50.pth
epoch_75.pth
epoch_100.pth
image_classify.py
last_checkpoint
resnet50_8xb32_in1k.py
mmpretrain
Prediction: 22, 0.53 (□□)
"pred_score": 0.0339798741042614
mmtest) PS D:\workspace\mmpretrain> python .\demo\image_demo.py .\exp\88.jpg .\exp\resnet50_8xb32_in1k.py --show-dir .\exp\ --device cuda:0 --checkpoint .\exp\epoch_100.pth
oads checkpoint by local backend from path: .\exp\epoch_100.pth
nference 100% 0:00:00
6/06 23:11:03 - mmengine - WARNING - `Visualizer` backend is not initialized because save_dir is None.
:\software\anaconda\envs\mmtest\lib\site-packages\mmengine\visualization\utils.py:240: UserWarning: Glyph 33760 (\N{CJK UNIFIED IDEOGRAPH-83E0}) missing from current font.
s, (width, height) = canvas.print_to_buffer()
:\software\anaconda\envs\mmtest\lib\site-packages\mmengine\visualization\utils.py:240: UserWarning: Glyph 33821 (\N{CJK UNIFIED IDEOGRAPH-841D}) missing from current font.
s, (width, height) = canvas.print_to_buffer()
"pred_label": 22,
"pred_score": 0.5250510573387146,
"pred_class": "菠萝"