摘要: 照片划分训练集和验证集,在 kaggle 云平台中进行图像识别模型训练。

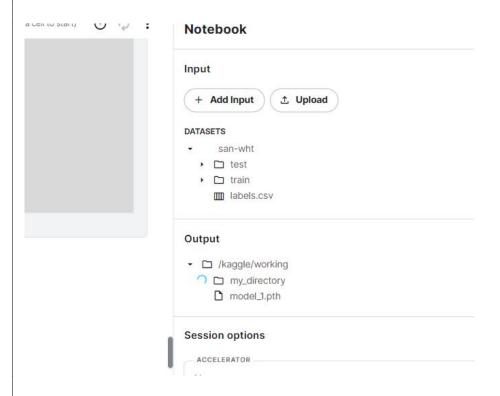
1. 实验准备

在 kaggle 中新建一个 notebook 后,将数据集上传到 kaggle 中,尝试使用 resnet34 模型去训练。

2. 实验步骤

2.1. 将照片划分为训练集和验证集

将数据集中的 train 划分训练集和验证集,至于 test 当作测试集



2.2. 编写 python 代码在 kaggle 云平台中进行图像识别模型训练

在 notebook 中编写代码进行训练:

('YNXSBN', 0.72041)

```
#从原始训练集中拆分验证集,然后将图像移动到按标签分组的子文件夹中
#三叶膏的图像识别
import collections
import math
import os
import shutil
import pandas as pd
import torch
import torchvision
from torch import nn
from d21 import torch as d21
data_dir = "/kaggle/input/san-wht/"
def read_csv_labels(fname):
     ""读取fname来给标签字典返回一个文件名"""
    with open(fname, 'r') as f:
       # 跳过文件头行(列名)
       lines = f.readlines()[1:]
    tokens = [1.rstrip().split(',') for 1 in lines]
    return dict(((name, label) for name, label in tokens))
labels = read_csv_labels(os.path.join(data_dir, 'labels.csv'))
print('# 训练样本 :', len(labels))
print('# 类别 :', len(set(labels.values())))
target_dir = '/kaggle/working/my_directory'
def copyfile(filename, target_dir):
    """将文件复制到目标目录
   os.makedirs(target_dir, exist_ok=True)
    shutil.copy(filename, target_dir)
def reorg_train_valid(data_dir, labels, valid_ratio):
     ""将验证集从原始的训练集中拆分出来"
    # 训练数据集中样本最少的类别中的样本数
   n = collections.Counter(labels.values()).most_common()[-1][1]
```

2.3. 比较不同分类方法的图像识别准确率

```
train loss 0.27, train acc 0.914, valid acc 0.965, valid loss 0.109
6551 examples/sec on [device(type='cuda', index=0), device(type='cuda', index=0)]
food: sared to //aggle/noring/osel_l.jpth
(gr 'coms', 1' 'coms', 2' 'cox', 3' 'casi', 4' 'aggo', 5' 'sox', 6' 'nsor', 7' 'nsosm', 8' 'ZZTZ', 9' 'ZZ
10 
Accuracy of GoSS: 985
Accuracy of MINISH: 985
Accuracy
```

3. 实验结果

然后发现似乎使用十分类,验证集的准确率反而更高,一个是96.5%; 一个是93.7%。总之模型的准

确率比较高。然后使用 pyqt 做了一个简单的识别。也能够识别。

