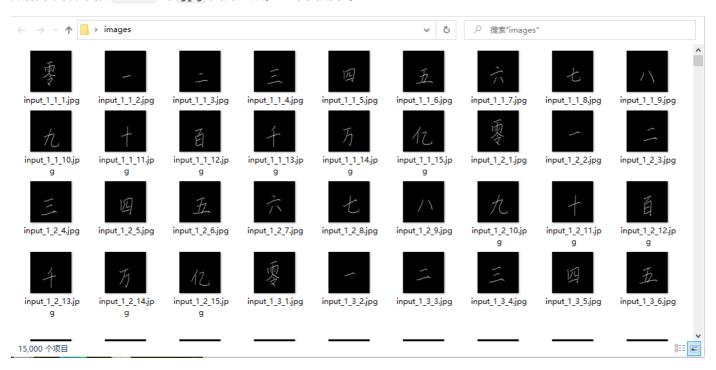
手写汉字数字识别

任务描述

本任务是一个图像分类任务,数据集包含了若干张手写的汉字数字图片,共包含 15 种中文汉字字符。

数据集说明

数据集由若干张 64*64 的 jpg 图片组成,如下图所示。



数据集已进行随机切分,划分为 train 、 dev 和 test , 其中 test 不公布, 目录结构如下:

```
1 data
2 ├── dev # 验证集,包含 3000 张图片
3 ├── dev_dict.csv # 验证集标签
4 ├── train # 测试集,包含 9000 张图片
5 └── train_dict.csv # 训练集标签
```

train dict.csv 文件包含了 train 对应图片的标签, 部分数据如下:

```
1
   suite_id,sample_id,code,label,character
2
   60,9,8,7,七
3
   9,7,9,8,八
4 8,4,2,1,-
5 47,9,14,13,万
   21,9,2,1,-
6
7
   24,4,1,0,零
   10,2,8,7,七
8
9
   84,9,4,3,<u>=</u>
10
```

每张图片的命名规则为: input_{suite_id}_{sample_id}_{code}.jpg 。

以 train_dict.csv 中第一条数据 60,9,8,7,七 为例,其表示 train 文件夹中文件名为 input_60_9_8.jpg 的图片的在数据集中的分类标签为 7,图片中实际的汉字数字为 七;

dev_dict.csv 文件结构与上图相同,包含了 dev 对应图片的标签。

硬件需求

1 内存: 最低4G 2 CPU: 最低2核