

OCR_for_HHU 程序 1.0 使用说明

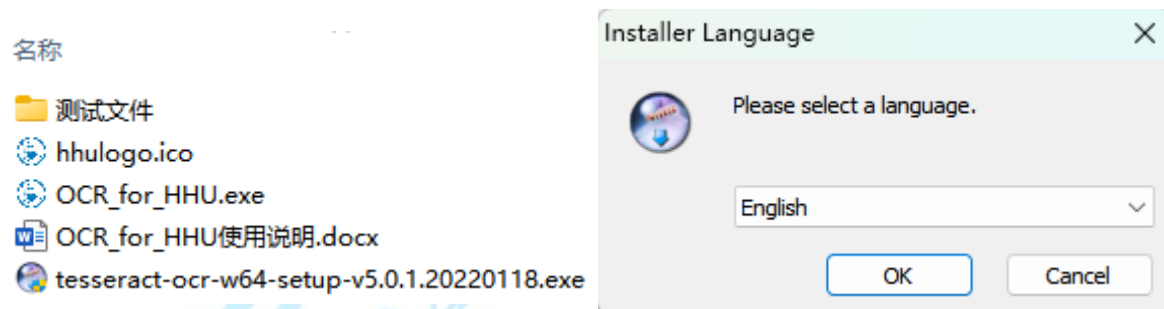
目录

一、 安装 Tesseract	1
二、 配置环境变量	4
三、 使用 OCR_for_HHU	8

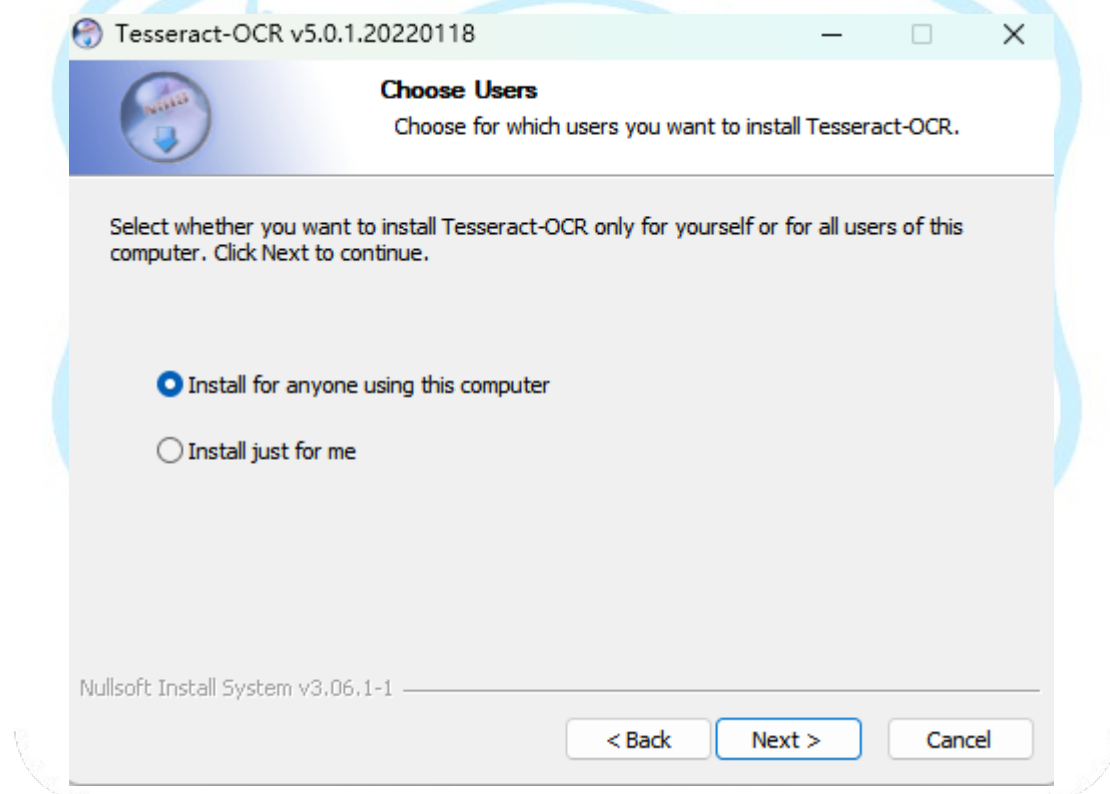
2022 年 4 月 9 日

一、 安装 Tesseract

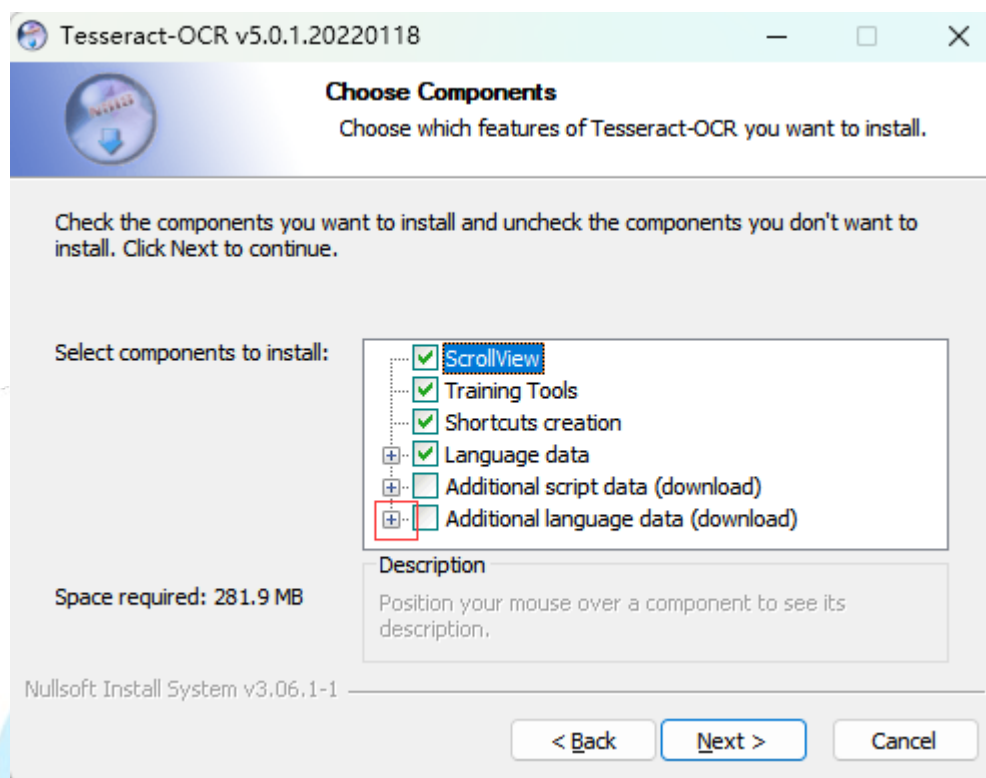
1. 系统要求： Windows10 及以上， 64 位操作系统。
2. 运行 tesseract-ocr-w64-setup 来安装 tesseract。



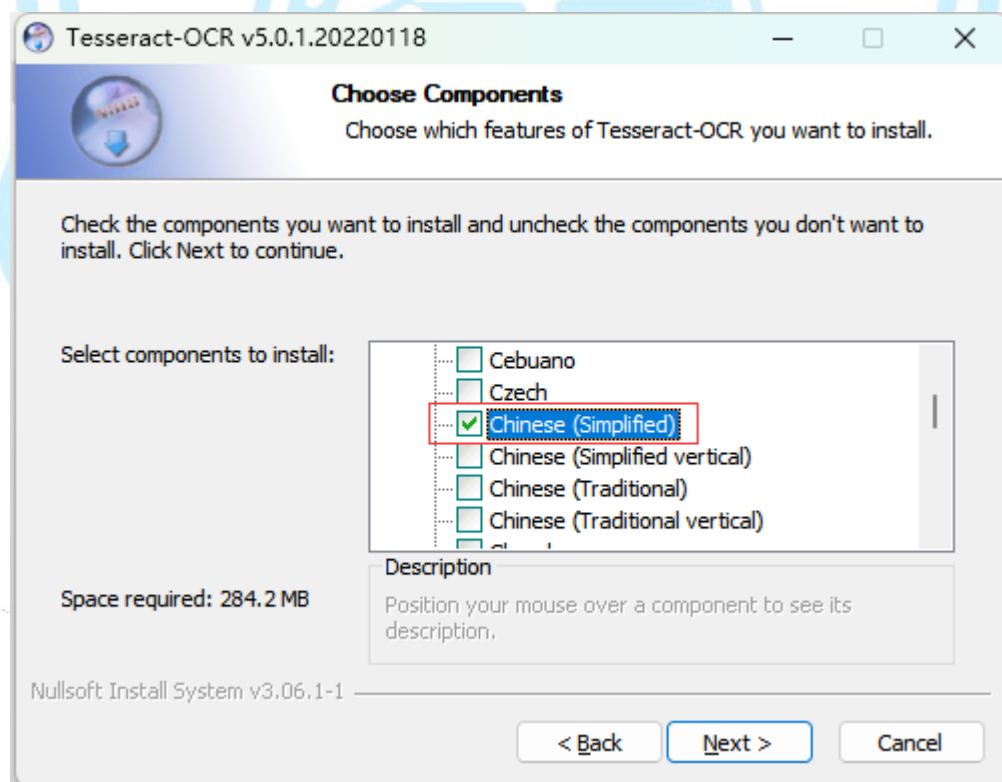
默认选择 English



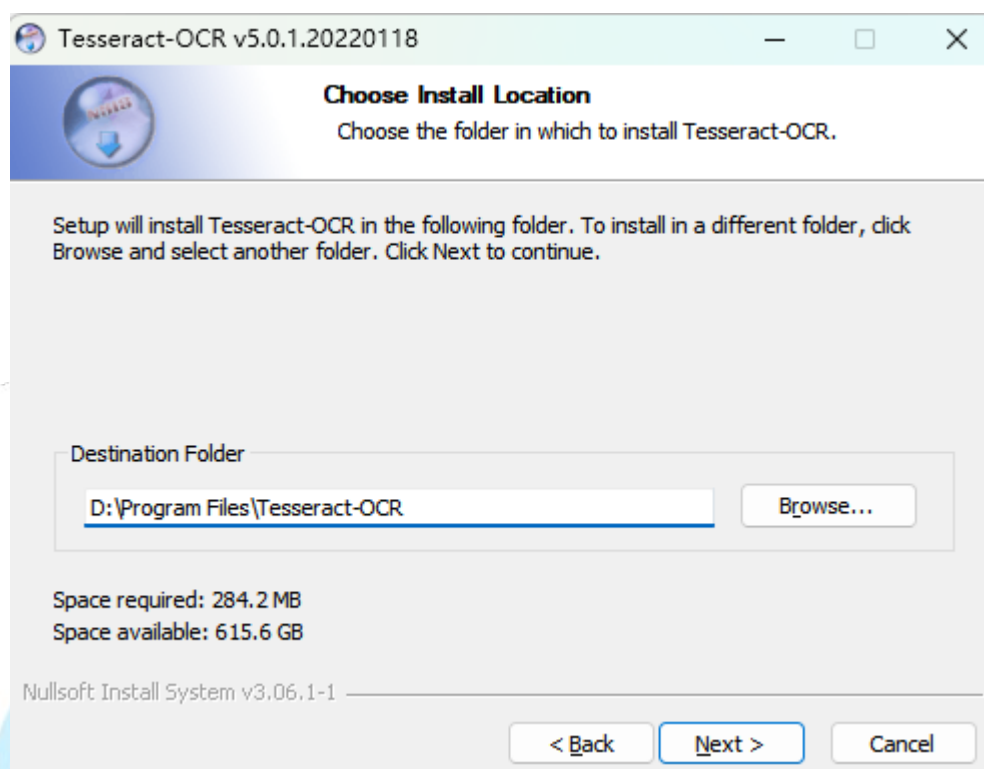
选择第一项： Install for anyone using this computer



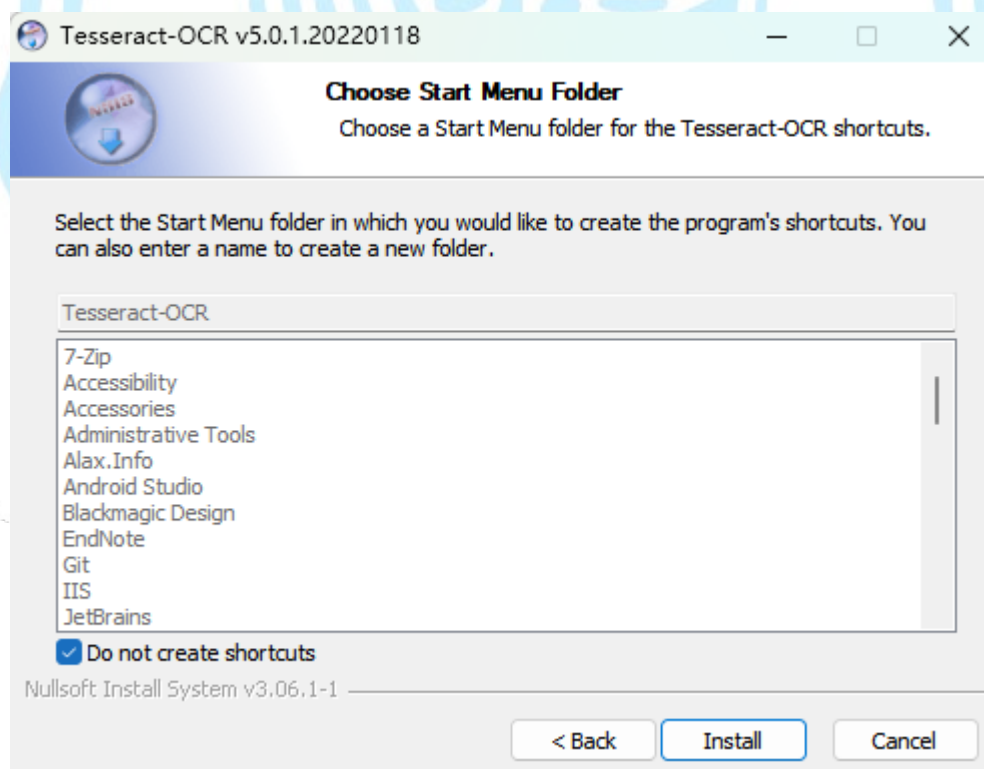
点击 Additional language data(download)左边的加号



下拉勾选 Chinese(Simplified)



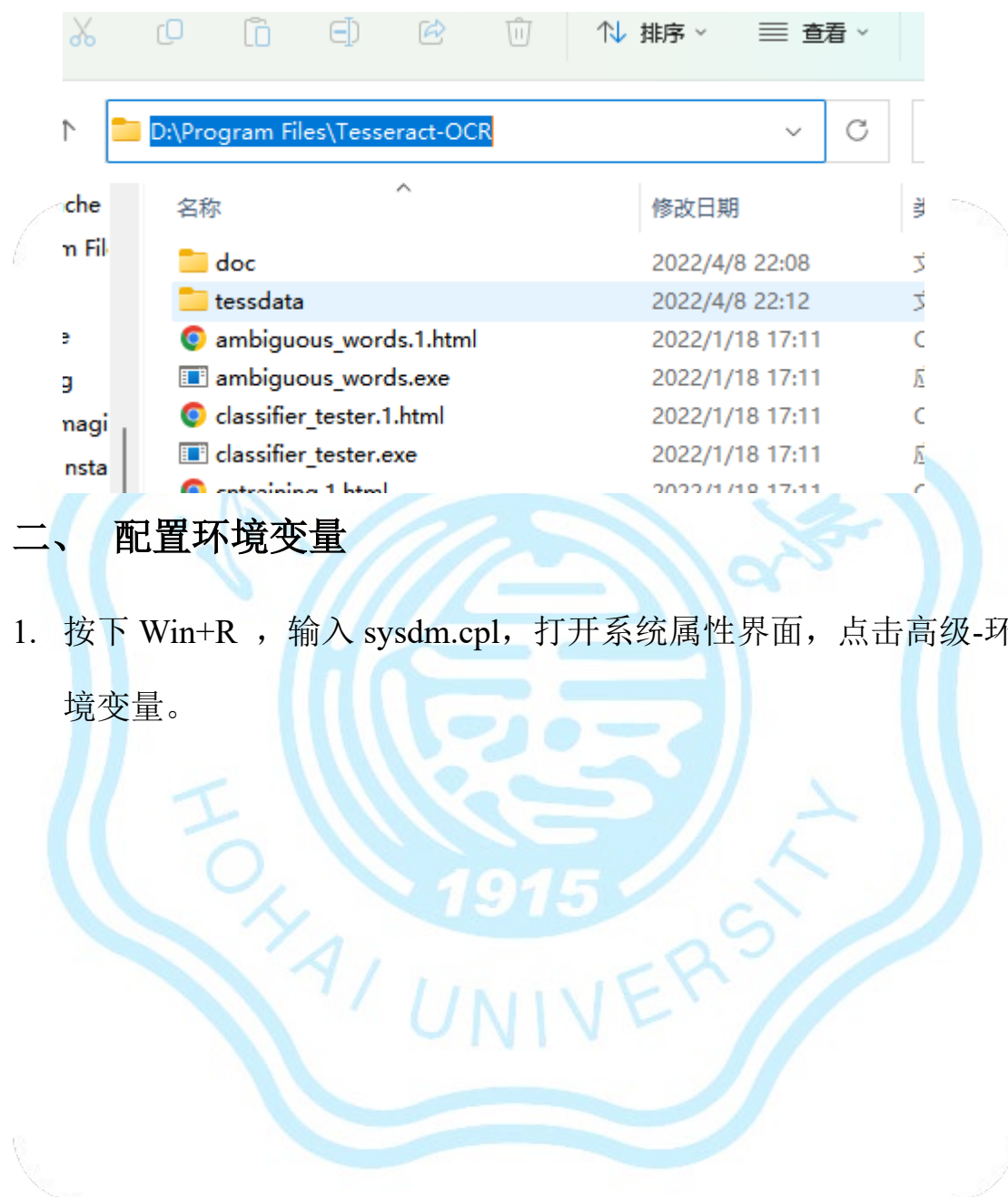
选择安装位置,这里可以直接把 C 盘改为 D 盘,要记住这个路径,下面需要用到。



勾选 Do not create shortcuts

点击 Install 即可完成安装。

完成后去复制一下安装文件夹的路径。

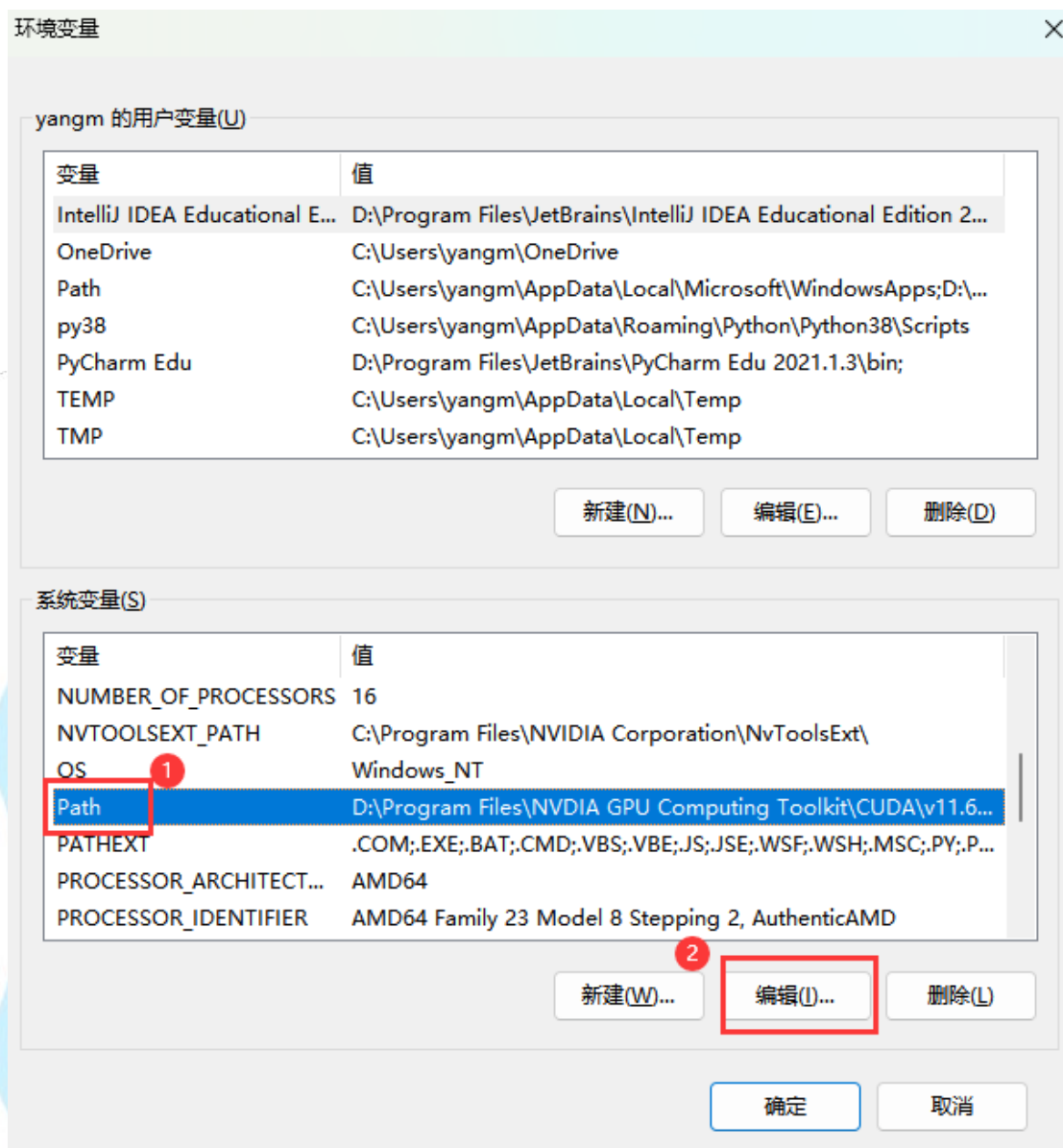


二、配置环境变量

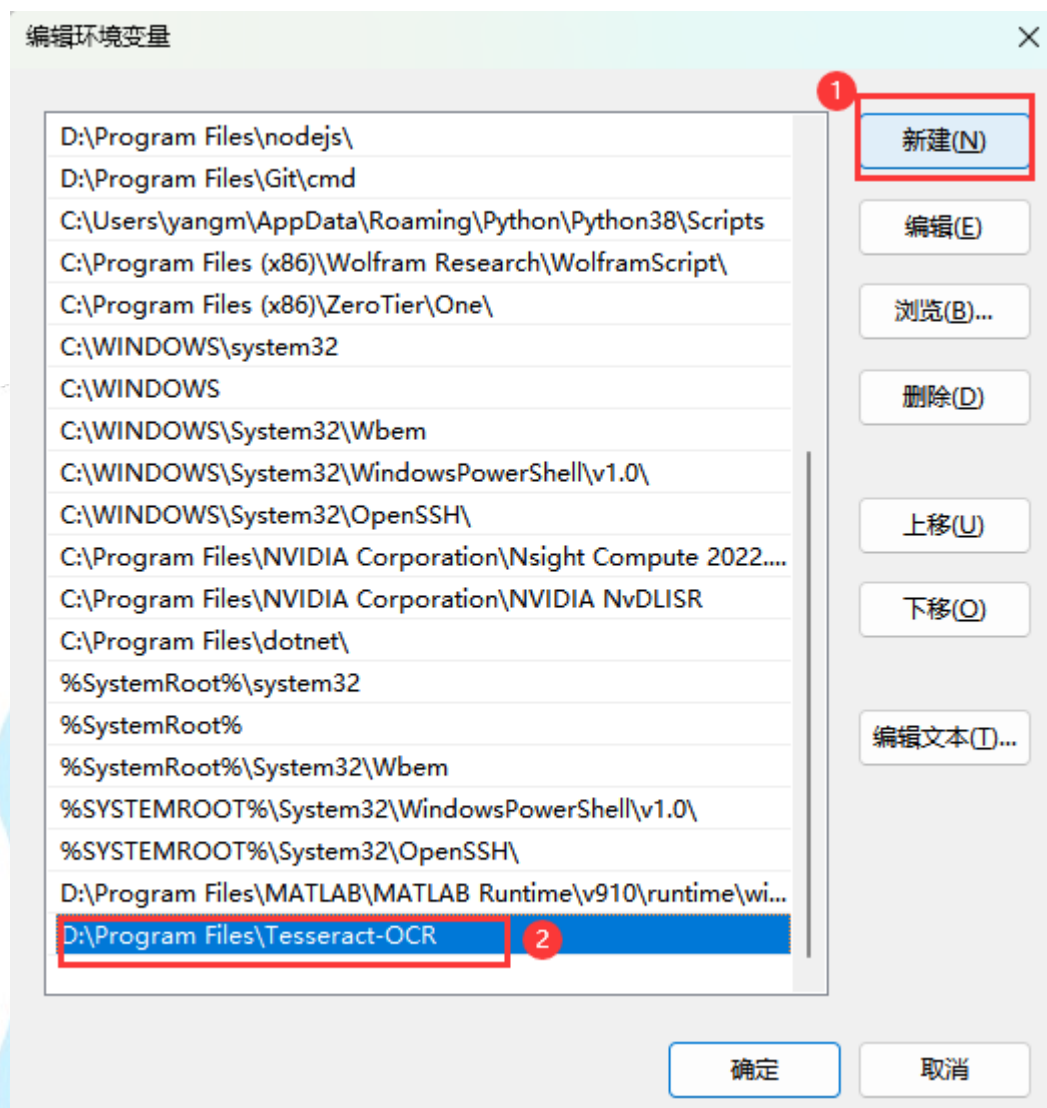
1. 按下 Win+R ， 输入 `sysdm.cpl`，打开系统属性界面，点击高级-环境变量。



2. 在下方系统变量里面找到 Path，再点编辑



3. 点新建，再粘贴刚刚复制的路径。

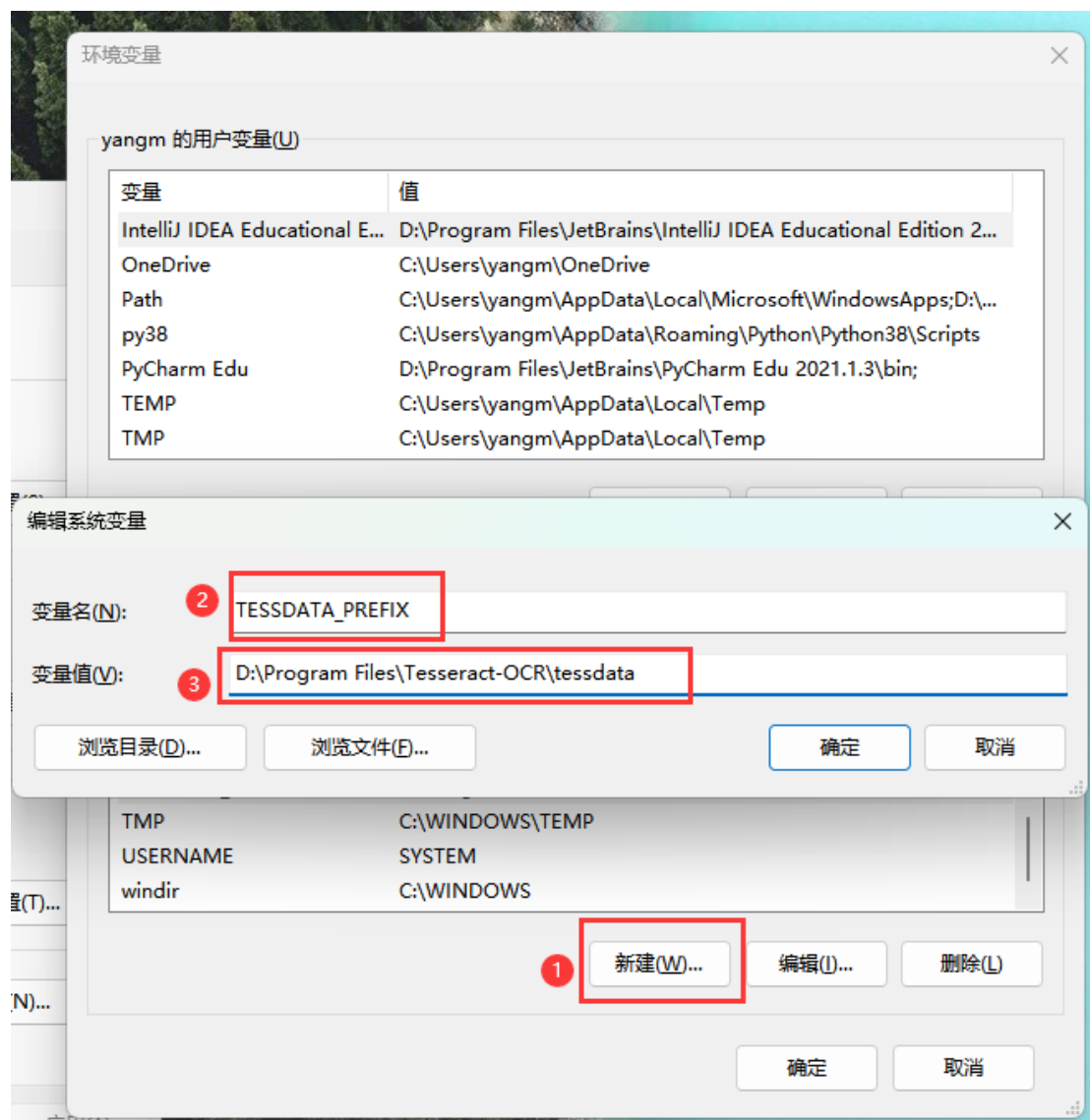


4. 然后点确定。

5. 在系统变量这里点新建，
在变量名中输入 TESSDATA_PREFIX

变量值输入 D:\Program Files\Tesseract-OCR\tessdata

这里的变量值就是安装路径的 tessdata 文件夹，如果你安装在其他地方，就更改一下。填完点击确定即可。



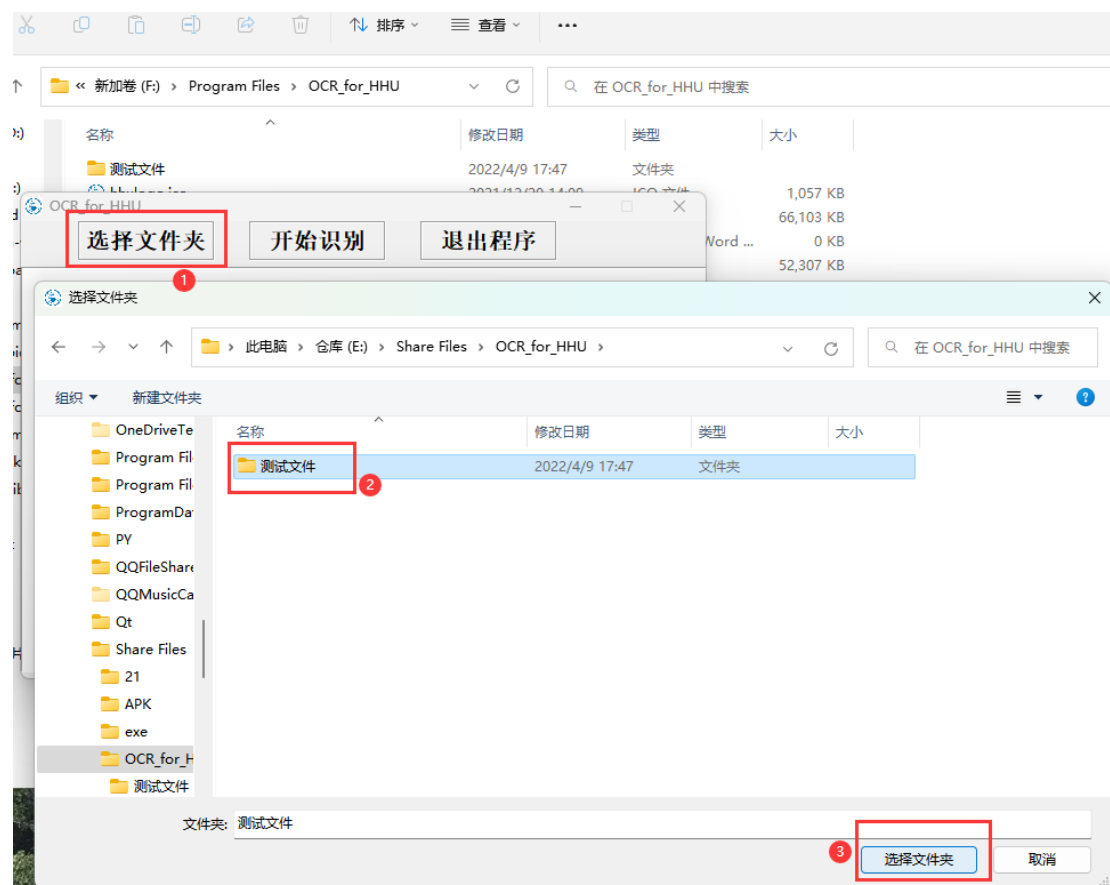
三、使用 OCR_for_HHU

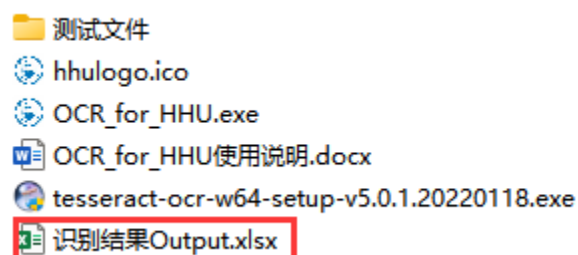
1. 软件界面非常简洁，首先选择需要识别的文件夹，确保文件夹内均为核酸检测截图，没有多余的文件和文件夹。截图必须是我的南京 APP 或者苏康码查询核酸检测结果的截图（如测试文件夹的图片），其他形式的图片可能会导致识别错误。
2. 选好文件夹再点击开始识别，根据电脑性能不同，识别速度会有差别，另外，图片分辨率的也会影响识别速度，分辨率高的

图片需要更长时间，一般情况是 1 秒识别一张图片。识别完成后结果会保存在“识别结果 Output.xlsx”文件夹内，因为每次识别都会对这个文件进行重写，所以识别完成后要保存结果的话，需要另存到其他位置。

3. 下面以测试文件夹进行演示







识别结果Output.xlsx - Excel

	A	B	C	D	E
1		0	1	2	3
2	0	文件名	姓名	检测日期	检测结果
3	1	2022-04-03-杨敏 - 副本 (2)	杨敏'	2022-04-03	阴性
4	2	2022-04-03-杨敏 - 副本 (3)	杨敏'	2022-04-03	阴性
5	3	2022-04-03-杨敏 - 副本 (4)	杨敏'	2022-04-03	阴性
6	4	2022-04-03-杨敏 - 副本 (5)	杨敏'	2022-04-03	阴性
7	5	2022-04-03-杨敏 - 副本 (6)	杨敏'	2022-04-03	阴性
8	6	2022-04-03-杨敏 - 副本 (7)	杨敏'	2022-04-03	阴性
9	7	2022-04-03-杨敏 - 副本 (8)	杨敏'	2022-04-03	阴性
10	8	2022-04-03-杨敏 - 副本 (9)	杨敏'	2022-04-03	阴性
11	9	2022-04-03-杨敏 - 副本	杨敏'	2022-04-03	阴性
12	10	2022-04-03-杨敏	杨敏'	2022-04-03	阴性
13					
14					
15					
16					

第 0 列是读取的文件名，第 1-3 列是识别的结果。

做几点说明：

- 检测日期和检测结果的识别准确率很高，一般不会有问题，核查也是以这两个为主。
- 图片的分辨率过低可能无法正确识别到姓名，图片的字体如果是手写体的话也会无法正确识别到姓名。

- 对于繁体字或者笔画较为复杂的字也可能会识别错误，因为本程序为了加快识别速度，没有加入较多功能，只识别简体中文和数字。
- 下面是反面示例

姓名: [REDACTED]

采样点: 河海大学

采样时间: 2022-04-05 12:15

检测结果: 阴性

2	3
检测日期	检测结果
2022-04-05	阴性