

OpenCV Image Processing Workshop

Pre-workshop materials

by: Lim Cher Liang

Disclaimer

This document contains installation guide for Python, VS code and Tesseract OCR on Windows. If you are using other operating system such as Linux or MacOS, please choose other installation version for your operating system. I have little experience in troubleshooting on other operating systems, but I will try my best to help.

If you encounter any problems, please try googling to solve it, the software used in this workshop is widely used, well documented, and discussed in a lot of forums. You can also reach out to me through Microsoft Teams chat or email me at efycl10@nottingham.edu.my

Please try to install the software before the workshop, we might not have enough time to troubleshoot during the workshop

Python Installation

<https://www.python.org/downloads/>

Download the latest version for Windows

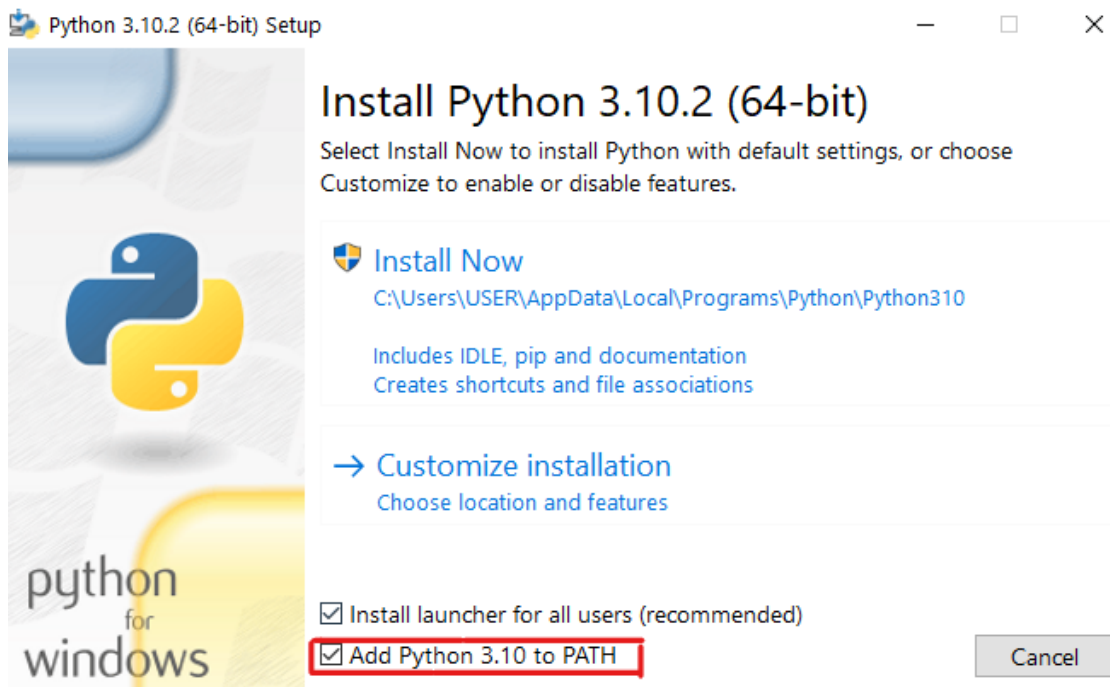
Download Python 3.10.2

Looking for Python with a different OS? Python for [Windows](#),
[Linux/UNIX](#), [macOS](#), [Other](#)

Want to help test development versions of Python? [Prereleases](#),
[Docker images](#)

Looking for Python 2.7? See below for specific releases

Any python version after 3.7 should work just fine.

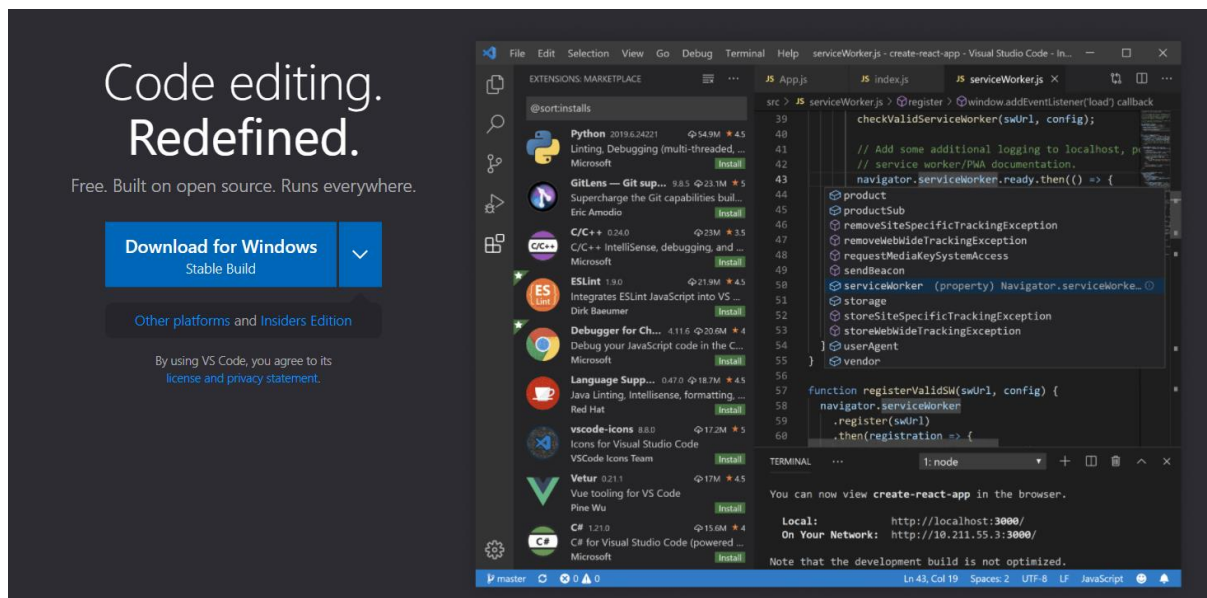


Remember to “check” the box for “Add Python 3.x to PATH”

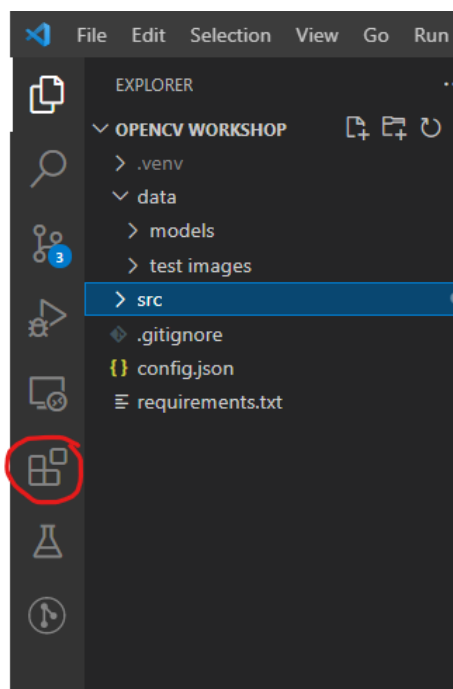
VS Code Installation

VS Code is the code editor of my choice, it very customisable with tons of features. It's widely used by programmers throughout the world

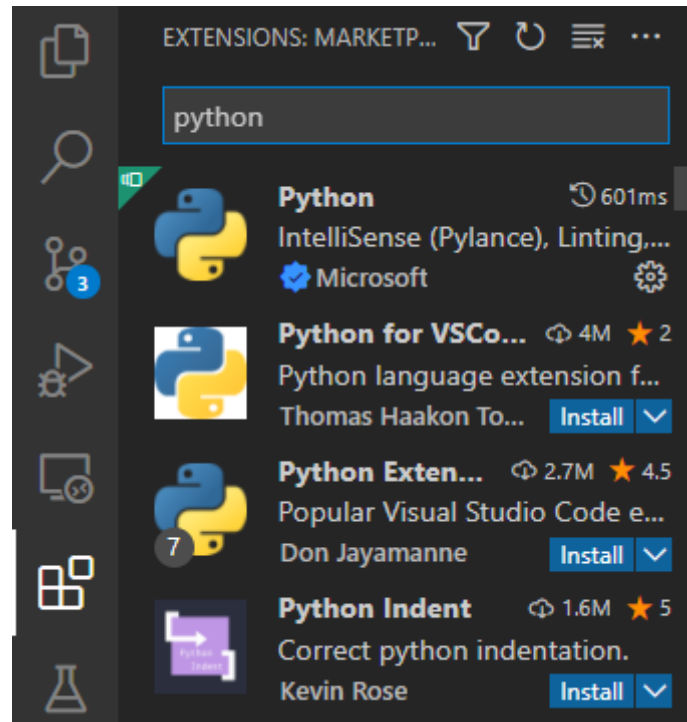
<https://code.visualstudio.com/>



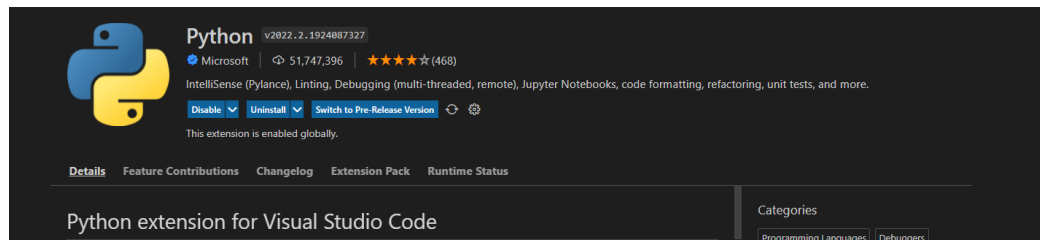
After installing VS Code, please click the “Extension” tab



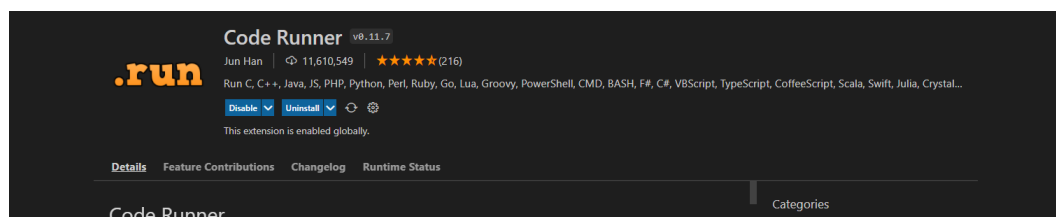
Then, search for Python



Install Python extension on VS Code to have IntelliSense features such as code completion



Please install “Code Runner” extension too



I will be using VS Code throughout the workshop, it would be easier for you to follow along if you install VS Code.

That being said, you can use other code editors for the workshop. Another alternative that I recommend is [PyCharm](#). It has better integration with Python and even more features. You can get the full version of PyCharm by requesting for the Student License.

Tesseract Installation

We will be using PyTesseract module in this project. Since PyTesseract is just a wrapper for Google's Tesseract-OCR Engine, so we need to install the Google Tesseract OCR Engine.

For MacOS users, you can refer to <https://pyimagesearch.com/2018/09/17/opencv-ocr-and-text-recognition-with-tesseract/> to install Tesseract 4.

For Windows, you can download the Windows Installer from <https://github.com/UB-Mannheim/tesseract/wiki>

The installer is endorsed by the Tesseract team

Tesseract installer for Windows

Normally we run Tesseract on Debian GNU Linux, but there was also the need for a Windows version. That's why we have built a Tesseract installer for Windows.

WARNING: Tesseract should be either installed in the directory which is suggested during the installation or in a new directory. The uninstaller removes the whole installation directory. If you installed Tesseract in an existing directory, that directory will be removed with all its subdirectories and files.

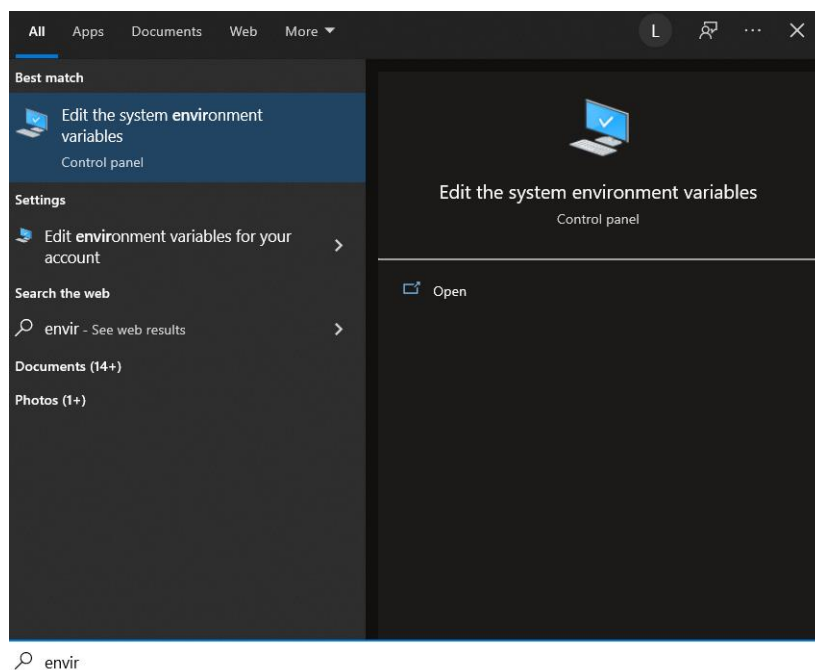
The latest installers can be downloaded here:

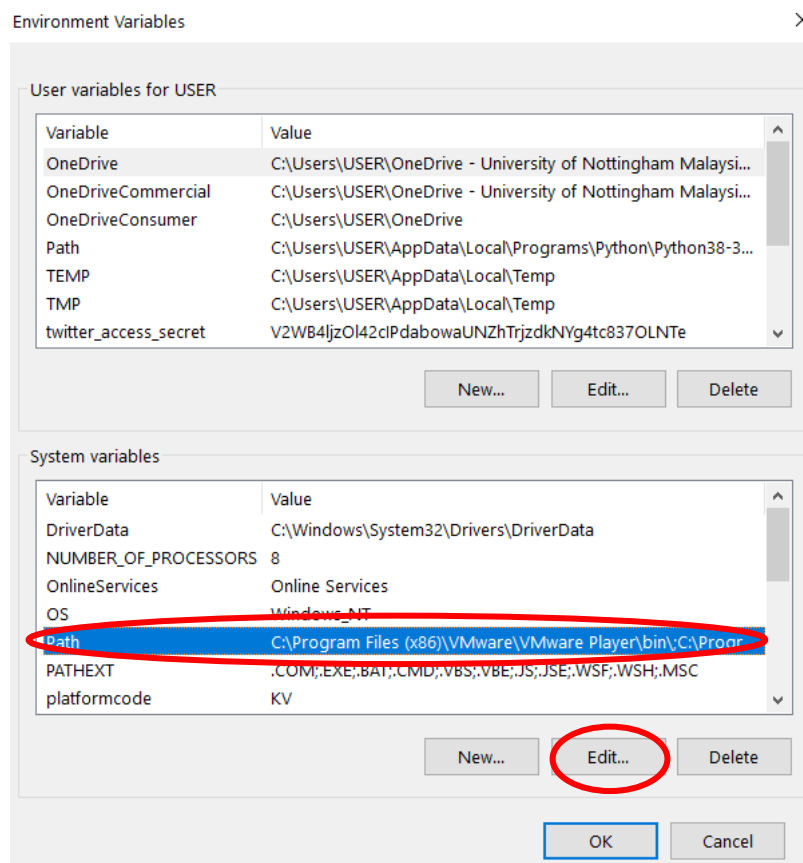
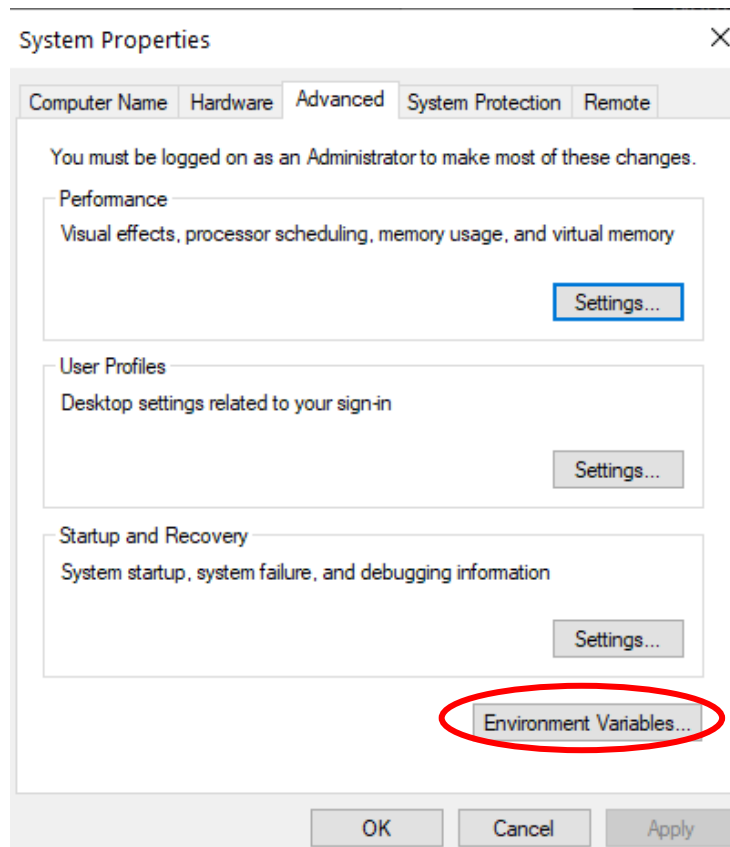
- [tesseract-ocr-w32-setup-v5.0.1.20220118.exe](#) (32 bit) and
- [tesseract-ocr-w64-setup-v5.0.1.20220118.exe](#) (64 bit) resp.

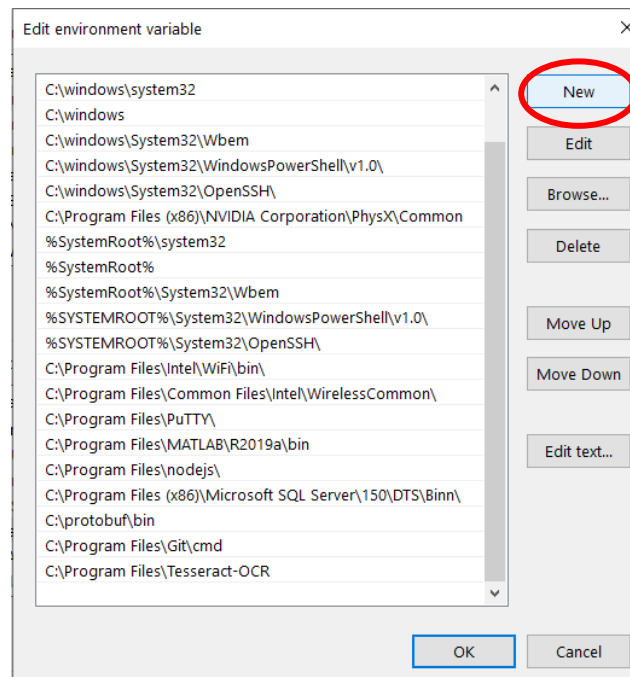
The download speed for the file might be a little slow since the file is hosted by the person who coded the installer.

After that, just your usual, next, next, agree to EULA, next, next, finish XD.

Then, Windows search for "Edit the system environment variables"







Then paste in the file path where you install Tesseract OCR.

Usually it's C:\Program Files\Tesseract-OCR by default

Finally, open command prompt to confirm if the Tesseract version

```
Command Prompt
Microsoft Windows [Version 10.0.19043.1526]
(c) Microsoft Corporation. All rights reserved.

C:\Users\USER>tesseract -V
Usage:
  tesseract --help | --help-extra | --version
  tesseract --list-langs
  tesseract imagename outputbase [options...] [configfile...]

OCR options:
  -l LANG[+LANG]      Specify language(s) used for OCR.
NOTE: These options must occur before any configfile.

Single options:
  --help              Show this help message.
  --help-extra        Show extra help for advanced users.
  --version           Show version information.
  --list-langs        List available languages for tesseract engine.

C:\Users\USER>DS_
```