Configuration Manual

MSc Research Project - Identification of Pneumothorax in X-ray images using Convolutional neural network

Programme Name - MSc DAD-A 2019-20

Aditya Sharma
Student ID: 18198821

School of Computing
National College of Ireland

Supervisor: Christian Horn

Student Name:	Aditya Sharma…………………………………………………………		
Student ID:	18198821………………………………………………………...……		
Programme:	MSC-DAD-A…………………………………	**Year:**	2019-20………..
Module:	Research Project…………………………………………….………		
Supervisor:	Christian Horn………………………………………………….………		
Submission Due Date:	17-08-2020…………………………………………………..………		
Project Title:	Identification of Pneumothorax in X-ray images using Convolutional neural network.		

Word Count: 890   **Page Count**   5

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: aditya sharma…………………………………………………………………

Date: 16-08-2020……………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Aditya Sharma
Student ID: 18198821

1 Configuration Manual

This document is the configuration manual which will help to run the model and train it too. The first part of this document covers all the steps which are to be followed to load the trained model and run it on images to get the result if the x-ray images are having pneumothorax are not. The second part of the document covers steps which should be followed to load the image data and train the model.

1.1 Loading the trained model and running it on images present in folder to get result

Please follow the below steps to load the model and run it on images to find out if the images are having pneumothorax or not.
1. Download and Install Anaconda (individual edition) from website: https://www.anaconda.com/
2. Open the Anaconda and click on Jupyter Notebook, Jupyter notebook will open in browser.
3. Copy the file named "ThesisResnet50-LoadAndRun.ipynb" in the working directory of Jupyter Notebook. For example: working directory where this project's code is present is 'C:\Users\adity'. To find out working directory, open a blank jupyter notebook and run the commands mentioned below. To open a blank python 3 notebook, click on "new" in jupyter notebook and select "python3".
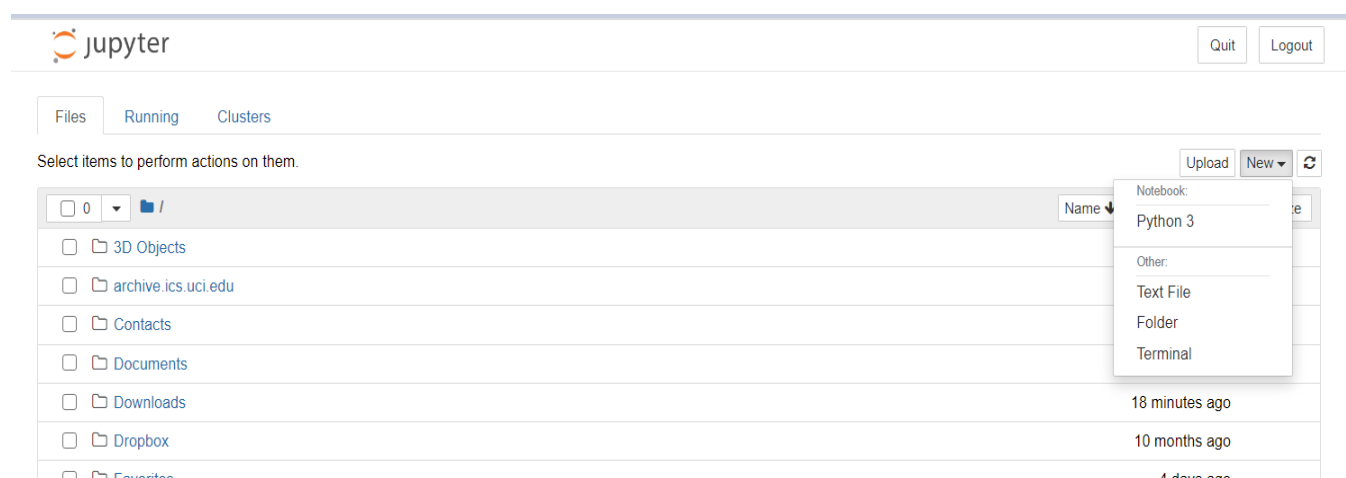

Fig 1: Click on 'new' and select 'python 3' to open new notebook

4. Working directory can be found out writing two simple commands and press Cntrl+Enter. Please see in the screenshot mentioned below.
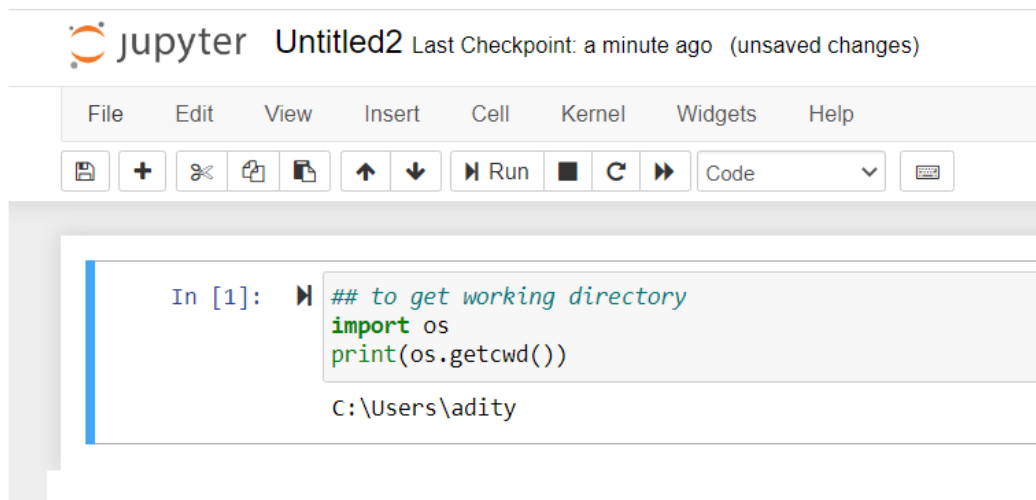

Fig 2: Command to get current working directory of python

5. After finding out the working directory, copy the file named "ThesisResnet50-LoadAndRun.ipynb" in the working directory, after copying it should be visible in the jupyter notebook as mentioned below in image.

Fig 3: File named "ThesisResnet50-LoadAndRun.ipynb" present in working directory

6. Click on the file named "ThesisResnet50-LoadAndRun.ipynb" to open it and then install the libraries if using jupyter notebook for the first time. The libraries should be installed before importing them and using them in code. Three libraries should be installed - numpy, keras, opencv-python. They can be installed one-by-one using command "pip install numpy" like mentioned in below screenshot. Write "pip install numpy" and press Cntrl+Enter to run the cell. Then replace "numpy" with "keras" to install keras library(pip install keras), similarly install "opencv-python" library.

2

```
In [1]:  ▶  pip install numpy

            Requirement already satisfied: numpy in d:\users\aditya\anaconda3\lib\site-packages (1.18.4)
            Note: you may need to restart the kernel to use updated packages.
```

Fig 4: Command to install numpy library

7. After the three libraries are installed, copy the trained model file named "Trained_model.keras" in one folder and give the path of the trained model file in variable "model_path = 'D:/ResearchProject/progress/Trained_model.keras'". Similarly, copy the x-ray images in one folder which should be used by the model to get results if the image is having pneumothorax or not. Give the path of the folder where images are present in variable "image_path = 'D:/ResearchProject/data/siim-acr-pneumothorax/predict'". Any number of images can be copied here. If there is only one image, then there will be one result and if multiple images are copied in the folder, then multiple results will be there.

```
In [6]:  ▶  ## Please change the path, where the model and images are stored

            ## Please give the path where the Trained_model.keras file is present in quotes in variable model_path
            model_path = 'D:/ResearchProject/progress/Trained_model.keras'

            ## Please copy the images in one folder and give the path of that folder in variable image_path
            image_path = 'D:/ResearchProject/data/siim-acr-pneumothorax/predict'
```

Fig 4: Path of trained model and images

8. After you have copied the image files and trained model and given its path in the code as mentioned above in figure 4, run each cell one by one, click inside rectangular cell to select it and press Cntrl+Enter(in windows) to run it. The "*" in the top left "In [*]" implies that the cell is running, wait for it to complete then run next cell. Below mentioned screenshot is of a cell which is running.

```
In [*]:  ▶  model = load_model(model_path)
            def load_images_from_folder(folder):
                images = []
                for filename in os.listdir(folder):
                    img = cv2.imread(os.path.join(folder,filename))
                    if img is not None:
                        #img = np.array(cv2.resize(img, (224,224)))
                        img = np.array(cv2.resize(img, (64,64)))
                        img = img.astype('float32')
                        img = img / 255.0
                        images.append(img)
                return np.array(images)
            images = load_images_from_folder(image_path)
            print(type(images),images.shape)
```

Fig 5: Cell is running, start in "[*]" means cell is running

9. Run the cells from top to bottom one by one. Let the execution of one step gets finished before running the next cell to get the result.

1.2   Training the model and importing the images

Please follow below steps to load the image dataset and to train a model.

3

1. Anaconda should be installed like mentioned in step 1 of section "Loading the trained model and running it on images present in folder to get result." After that steps 2, 3 and 4 of section "Loading the trained model and running it on images present in folder to get result" should be followed to open the jupyter notebook and find out the working directory.
2. Copy the file name "ThesisResnet50.ipynb" in the working directory as mentioned in step 5, of section "Loading the trained model and running it on images present in folder to get result". Open the file "ThesisResnet50.ipynb".
3. Install the libraries (if not installed) - keras, numpy, pandas, scipy, opencv-python like mentioned in step 6 of section "Loading the trained model and running it on images present in folder to get result".
4. Give the paths of the folders where the testing and training images are present, in variables "folder_train", "folder_test" as mentioned below in screenshot.

```
In [2]: ⏭  ## This cell loads the images present in folders, test and train images are place
            ## the folders is given in below in variables folder_train, folder_test.

            folder_train = 'D:/ResearchProject/data/siim-acr-pneumothorax/train'
            folder_test = 'D:/ResearchProject/data/siim-acr-pneumothorax/test'
            def load_images_from_folder(folder):
                images = []
                for filename in os.listdir(folder):
                    img = cv2.imread(os.path.join(folder,filename))
                    if img is not None:
                        #img = np.array(cv2.resize(img, (224,224)))
                        img = np.array(cv2.resize(img, (64,64)))
                        img = img.astype('float32')
                        img = img / 255.0
                        images.append(img)
                return images
            images_test = load_images_from_folder(folder_test)
            images_train = load_images_from_folder(folder_train)
```

Fig 6: Cell in which paths of folders where images are present are to be given

5. Give the paths where the excel sheet, named "stage_1_test_images.csv" and "stage_1_train_images.csv" having the labels of images (both test and training) are present, like mentioned in the screenshot below.

```
In [4]: ⏭  # read the labels of the images

            path_test = 'D:/ResearchProject/data/siim-acr-pneumothorax/stage_1_test_images.csv'
            label_test = pd.read_csv(path_test, usecols=[2], skiprows = 1,header = None)

            path_train = 'D:/ResearchProject/data/siim-acr-pneumothorax/stage_1_train_images.csv'
            label_train = pd.read_csv(path_train, usecols=[2], skiprows = 1,header = None)

            Y_train, Y_valid = label_train, label_test
            Y_train, Y_valid = np.array(Y_train), np.array(Y_valid)
            print(Y_train.shape,type(Y_train), Y_valid.shape, type(Y_valid))

            (10675, 1) <class 'numpy.ndarray'> (1372, 1) <class 'numpy.ndarray'>
```

Fig 7: Cell in which paths of folders where excel sheets of labels are present, are to be given

6. Give the paths where the model should be saved after it is trained as mentioned in below image.

```
In [14]:  ▶| model.save('D:/ResearchProject/progress/12.keras')
             model.save_weights("D:/ResearchProject/progress/12.h5")
```

Fig 8: Cell in which paths of folders where model needs to be saved, are to be given

7. After the libraries are installed and paths of image folders, excel sheets and where model
   should be saved, are given, run each cell one-by one, by first clicking inside the cell to select
   it and then press Cntrl+Enter to run the cell. The cell in which images are to be imported will
   take time to complete, wait for each cell to complete and then run next cell. The "*"
   mentioned in top-left implies that the cell is running, as explained in step 8 of section
   "Loading the trained model and running it on images present in folder to get result"