National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Daniel Murphy
Student ID: x20138164

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

Configuration Manual

Daniel Murphy
x20138164

1 Introduction

This Con guration Manual outlines the requirements and prerequisites as well as the stepwise procedure to setup the environment and to reproduce the results as reported in the accompanying research project report. The environment used as part of this work was the Google Colaboratory, and the steps to sign-up and con gure this environment will be presented here rst. This is followed by detailed instructions of how to acquire and store the datasets (although these are also provided as part of the research portfolio), and nally the directions to execute the accompanying code and reproduce the nal results is presented.

2 Hardware & Software Requirements

2.1 Hardware Specification Requirements

This section details the hardware speci cations that were used to access the Google Colaboratory environment. It is important to note that no processing was carried out using the local desktop system presented here, the only requirement on this hardware is the ability to access the Google Drive/Colaboratory environment which is detailed in the next section.

- **CPU:** Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99 GHz
- **RAM:** 8 GB
- **Stystem Type:** 64-bit operating system, x64-based processor

2.2 Software Specification Requirements

This section presents the software used to access the Google Drive/Colaboratory environment and the system speci cations that this cloud o ering provides. Google Colaboratory provides access to GPUs and CPUs located in Google's data centres (*Google Colaboratory*; 2022).
Local Desktop Software Speci cation:

- **Operating System:** Windows 10
- **Browser** Google Chrome Version 104.0.5112.81 (O cial Build) (64-bit)

Google Drive/Colaboratory (Colab) Environment:

- **Colab Subscription:** Colab Pro+
- **GPU:** 2 (unknown specs - depended on runtime)
- **RAM:** 52 GB
- **Disk Storage:** 15 GB

3 Google Environment Configuration

This section provides the instructions for setting up and con guring the Google environment. The instructions include the optional step of creating a new account with Google for users who do not already have access to one, followed by the installation of all required packages and dependencies.

3.1 Google Account Creation

This section outlines how to create a new account with Google. While this is an optional step for those who already have a Google Account, it is still highly recommended that a new account is created as the storage of the datasets and other artefacts requires a large amount of the available disk space in a Google account (in excess of 90%).

To create a new account, navigate to the following link: `https://support.google.com/accounts/answer/27441?hl=en` and select 'For Myself'. This can be seen in the screenshot shown in Fig 1 After lling in all relevant information, and agreeing to the terms and conditions, you showld be greeted with the following home page shown in Fig. From here, navigate to Google Drive by clicking the ellipses in the top right corner of the screen. From there, Colab can be accessed by clicking New {> More {> Google Colaboratory. Note that if the application does not appear on this list at the rst attempt, it may need to be added by clicking 'Connect More Apps' and selecting it from the pop up screen, as shown in Fig 2

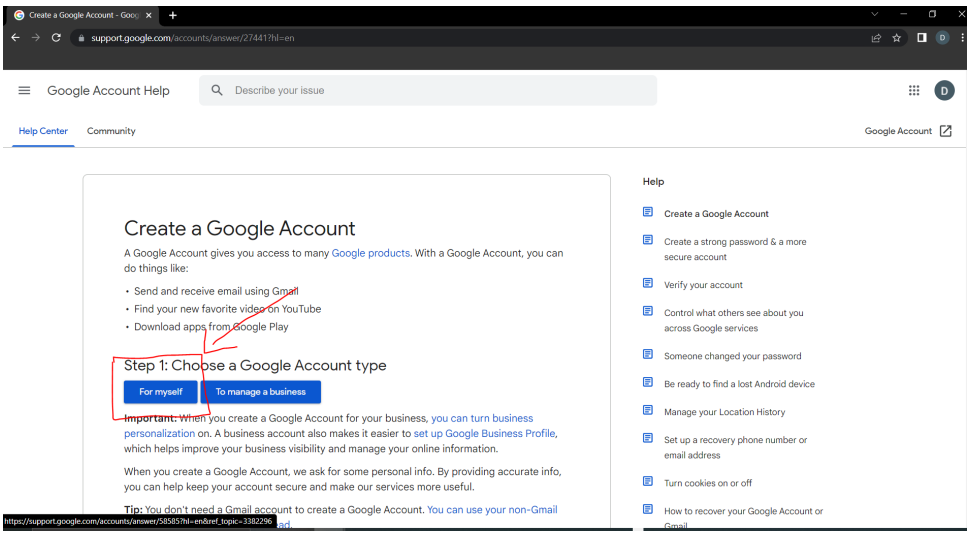

Figure 1: Create Google Account Home Page

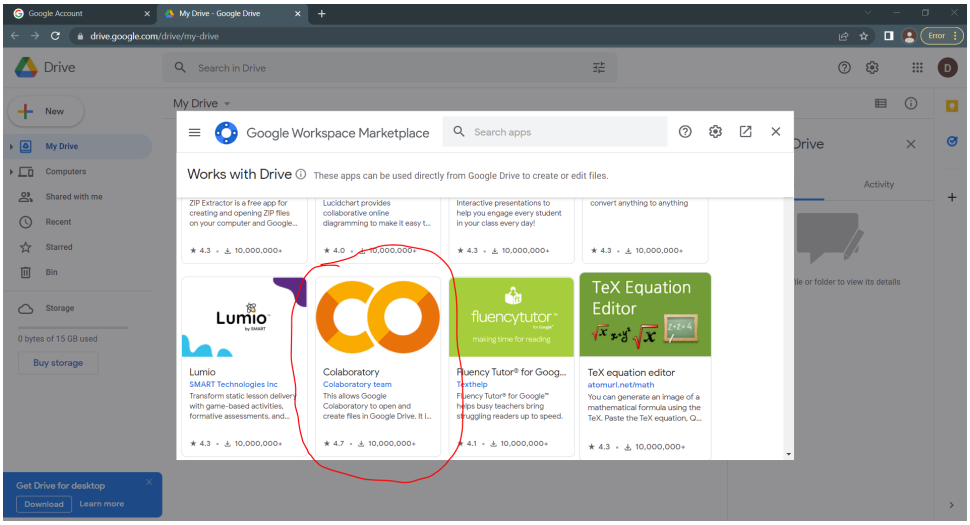

Figure 2: Installing Google Colaboratory in GDrive

3.2 Installing & Importing Required Packages

Installation

Since Google's Colab environment comes with many Python packages pre-installed, there is no need to install every package to be used as they can simply be imported straightaway. There are however, two additional packages which do not come pre-installed and need to be installed manually. These packages are:

- cartopy
- richdem

They can be installed by running the following commands:

- !pip install richdem
- !pip install cartopy

Importation

The full set of packages and how to import then are listed below. In some cases, individual functions are imported from a given package, and imported functions or packages may be aliased to a more convenient name to be referenced in the code.

- import numpy as np
- import pandas as pd
- import matplotlib.pyplot as plt
- import richdem as rd
- import cartopy
- from google.colab
  import drive

- import h5py
- from tensor ow.keras.layers
  import Input, Lambda, Dropout, Reshape, Dense, Flatten, Conv2D, Concatenate
- from tensor ow.keras.models
  import Model
- from tensor ow.keras.applications.vgg19
  import VGG19
- from tensor ow.keras.applications.resnet50
  import preprocess_input
- from tensor ow.keras.preprocessing
  import image
- from tensor ow.keras.preprocessing.image
  import ImageDataGenerator,load_img
- from tensor ow.keras.models
  import Sequential
- from tensor ow.keras.regularizers
  import l2
- from tensor ow.keras.layers
  import UpSampling2D
- from tensor ow.keras.optimizers
  import Adam
- from tensor ow.keras.utils
  import plot_model
- from tensor ow.keras.models
  import load_model
- from glob import glob
- from tensor ow.keras.layers
  import MaxPooling2D
- from tensor ow.keras.losses
  import SparseCategoricalCrossentropy
- from tensor ow.keras.utils
  import plot_model
- from sklearn.metrics
  import confusion_matrix
- from sklearn.metrics
  import plot_confusion_matrix

- from sklearn.metrics
  import accuracy_score
- from sklearn.metrics
  import recall_score
- from sklearn.metrics
  import precision_score
- from sklearn.metrics
  import f1_score

4 Data Acquisition

Importantly, all of the data used as part of the research project is included in the accompanying code submission, and stored in the appropriate file location. If for any reason the data needs to be acquired directly from source, the instructions to do so are provided here. There are two sources to gather the data, the first is the DEM image data and the second is the ground-truth crater labels (Robbins and Hynek; 2012).

Mars DEM Image Data:

1. Navigate to the following link: `https://astrogeology.usgs.gov/search/map/Mars/Topography/HRSC_MOLA_Blend/Mars_HRSC_MOLA_BlendDEM_Global_200mp_v2`
2. In the upper left corner, select download. Make sure to select the full dataset and not an example. The dataset is large ($>$ 10GB) so this may take several minutes.
3. Once the download is complete, the dataset will need to be uploaded to GDrive to be accessible by the Colab environment. Upload the file to the file path: 'My Drive/ MSc Research Project/Raw Source Data/'.

Ground-Truth Craters Database:

- Navigate to the following link: `https://craters.sjrdesign.net/`.
- Use the download link on the page to download a zip file of the Robbins and Hynek (2012) crater database.
- Once the download it complete, unzip the file.
- Again, the file will need to be uploaded to the GDrive location: 'My Drive/ MSc Research Project/Raw Source Data/'.

5 Code Execution

All of the code executed as part of the research project is contained in Colab notebook files which are all stored in the location: 'My Drive/MSc Research Project/Scripts'. There are a total of 6 scripts, 5 of which need to be executed to run the entire code base (the 6th file is a .py file which contains some functions that are imported during the execution of the colab notebooks). To run the colab notebook files, open the file

and click Runtime {> Run All. Again, it should be noted that these notebooks are set to High-RAM (52GB) runtimes. Without a High-RAM runtime, the execution may fail due to hardware constraints, which is why this requirement was particularly stressed in the Hardware & Software Requirements section. To change the runtime type, click Runtime {> Change runtime type. To run the code, follow this procedure and execute the following scripts in the given order:

1. **Create Dataset.ipynb:** This script samples the raw source global dataset and creates the dataset to be used as part of the research project.

2. **Data Exploration.ipynb:** This script combines the image, crater and terrain elements and performs the exploratory analysis of the nal dataset.

3. **Build and Train Models.ipynb:** This script pre-processes the data, builds the models and trains them on the processed data.

4. **Results_Iterate_Over_Params.ipynb:** This script tests the models on the crater detection task, and tunes parameter values to generate the nal results.

5. **Terrain_Classification_Results.ipynb:** This script tests the models on the terrain classi cation task and saves the results.

6 Results & Outputs

All of the outputs from executing the code are stored in the location: 'My Drive/MSc Research Project/Outputs'. This folder contains two folders, as follows:

1. **Results**
   This folder contains the various plots, charts and gures produced.

2. **Saved Models**
   This folder contains the saved checkpoints of all of the 5 developed models. Each model is stored in its own folder with descriptive names. Models can be loaded from these folders at any time. The model folders also contain the saved parameter values and model performance metrics for each associated model.

References

Google Colaboratory (2022).
URL: *https://colab.research.google.com/notebooks/intro.ipynb*

Robbins, S. J. and Hynek, B. M. (2012). A new global database of Mars impact craters 1 km: 1. Database creation, properties, and parameters, *Journal of Geophysical Research: Planets* **117**(E5). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2011JE003966.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1029/2011JE003966*