Group 11 Readmi file,

Abdullah Albattal & Anjali Narayanan

The scripts used to implement our project are divided into two folders. The scripts in each folder must be executed in order and on the specified machine as indicated in the steps below for the code to run without errors. The 1st folder (Fish4Knowledge\_code) includes the scripts needed to implement our work on the Fish4Knowledge dataset. The scripts are numbered in the order they should be executed. The 2nd folder (Nature\_Conservancy\_dataset\_code) is for the scripts needed to implement our work on the Natural Conservancy dataset.

IMPORTANT: Some of the scripts are meant to be implemented on a local PC, some on the Datahub, and one script can only be executed on Google Colab.

You can either start by implementing the scripts for the Fish4Knowledge Dataset or the National Conservancy dataset.

1) Fish4Knowledge dataset:

In the directory (/Fish4Knowledge\_dataset\_code)

Downloading the data:

Visit the dataset website (<http://groups.inf.ed.ac.uk/f4k/GROUNDTRUTH/RECOG/> ) or download the dataset directly from the link (<http://groups.inf.ed.ac.uk/f4k/GROUNDTRUTH/RECOG/Archive/fishRecognition_GT.tar>). After downloading the dataset in the .tar file, the dataset need to be extracted into the respective folders depending on the class of the fish is under (there will be 23 folders named fish\_01 to fish\_23 and 23 folders named mask\_01 to mask\_23) these should be extracted in the data subdirectory in Fish4Knowledge\_dataset\_code folder (/Fish4Knowledge\_dataset\_code/data/fish\_xx).

1.A) Run the script “script\_01\_local\_pc\_create\_fish\_dataset\_original\_to\_npy\_file.ipynb” on your local machine: This will load the data set into a data.npy (containing the images and the masks) and species.npy (containing the lables of the classes). The script will save these two .npy files in the same directory (/Fish4Knowledge\_dataset\_code).

1.B) Run the script “script\_02\_local\_pc\_fish\_augmentation.m” on your local machine: This will perform data augmentation on the dataset, create a subdirectory called “augmented\_data” that contains the augmented images.

1.C) Run the script “script\_03\_local\_pc\_create\_fish\_dataset\_augmented\_to\_npy\_file.ipynb” on your local machine: This will load the augmented data set into a data\_aug.npy (containing the images and the masks of the augmented dataset) and species\_aug.npy (containing the lables of the classes). The script will save these two .npy files in the same directory (/Fish4Knowledge\_dataset\_code)

1.D) The files need to then be uploaded to the datahub in a subdirectory called (/data) from the directory where you will run the last script which will implement the CNN models and generate the final results (training, prediction, and classification report).

1.E) Run the script “script\_04\_data\_hub\_fish\_classification.ipynb” on the datahub: The code will load the .npy files from the (/data) subdirectory and train and test Rathi’s model and our improved model on the Fish4Knowledge dataset (the augmented dataset). This script will perform training, prediction, and the generation of the confusion matrix and classification report.

2) The Natural Conservancy dataset:

In the directory (/Nature\_Conservancy\_dataset\_code)

Downloading the data:

This dataset is a private dataset and was given to us by our colleagues in the Natural Conservancy for the purpose of using in this class. Please do not share this dataset and if you download it and use to review our submission make sure you delete it afterwards.

The dataset can be downloaded from the following Google Drive Folder: <https://drive.google.com/drive/folders/12lZfMH-OSpcDpR4-2EqukVKzL5bt5-jw?usp=sharing>

Instead of downloading it, please add it to your Google Drive as the 1st script for this part of the project is implemented on Google Colab (<https://colab.research.google.com/notebooks/intro.ipynb>).

2.A) Run the script “script\_01\_GOOGLE\_COOLAB\_fish\_segmentation\_MaskRCNN\_ResNet\_152.ipynb” on your Google Colab: This will load the dataset then apply MASK-RCNN on the images to segment and isolate the fish in them. After that, the masks together with the images will be saved to .npy files. These files should be downloaded to a local machine for the remaining scripts to function (save them to the same directory where the remaining scripts are). This script is based on the repository (<https://github.com/fizyr/keras-maskrcnn>)

IMPORTANT: (before you run the 1st script)

2.A.1) In the 1st script, the directories to the files must be adjusted manually according to the drive folder. The locations in the script where the directory must be changed are marked with the comment “# ------------>”

2.A.2) You need to also use the pre-trained model, which is available in Google Drive above. The pretrained model is saved in the file: (resnet152\_oid\_v1.0.1.h5)

2.B) Run the script “script\_02\_Local\_PC\_preparing\_data\_for\_augmentation.ipynb” on your local machine: This will create sub directories where the images are going to be stored for the augmentation script to perform augmentation.

2.C) Run the script “script\_03\_Local\_PC\_fish\_augmentation\_Natural\_Con\_dataset.m” on your local machine: This will perform data augmentation on the dataset, create a subdirectory called “augmented\_data” that contains the augmented images.

2.D) Run The script “script\_04\_Local\_PC\_create\_fish\_dataset\_to\_npy\_file\_augmented.ipynb” on your local machine: This will load the augmented data set into a data\_aug\_nature.npy (containing the images and the masks of the augmented dataset) and species\_aug\_nature.npy (containing the lables of the classes). The script will save these two .npy files in the same directory (/Nature\_Conservancy\_dataset\_code)

2.E) Run The script “script\_05\_Local\_PC\_create\_fish\_dataset\_to\_npy\_file\_original.ipynb” on your local machine: This will load the original data set into a data.npy (containing the images and the masks of the original images of the dataset) and species.npy (containing the lables of the classes). The script will save these two .npy files in the same directory (/Nature\_Conservancy\_dataset\_code)

2.F) These 4 .npy files need to then be uploaded to the datahub in a subdirectory called (/data) from the directory where you will run the last script which will implement the CNN models and generate the final results (training, prediction, and classification report).

1.E) Run the script “script\_06\_data\_hub\_Fish\_classification\_Nature\_Data.ipynb” on the datahub: The code will load the .npy files from the (/data) subdirectory and train and test Rathi’s model and our improved model on the The Nature Conservancy dataset (the augmented dataset).