UC San Diego Extension
Cloud Services for Machine Learning

Summer 2020
Homework#5

Date Given: July 27, 2020 Due Date: Aug 2, 2020

==========

There are 2 problems in this homework assignment.

==========

Problem#1:

In this assignment we will build an image classification model (single label) using the AutoML feature of the GCP Vision API. We will use images of different categories and assign a label to them. We will train our model with these images. Next, we will classify test images (which the model has not seen before) using the model and measure its accuracy.

Image quantity and quality

GCP AutoML feature works better when we have thousands of images of each class (sub-categories). We will download hundreds of images for our assignment. The accuracy of the prediction depends on the number of images used for training. More images we have to train the model, better would be the accuracy.

An image classifier is only as good as the images used during the training period. If the images fed to the model are of poor quality, accuracy of the prediction will also be poor.

Download the images from the web using a browser/extension & Google Search

We need hundreds of images to train our model. If you have access to hundreds of images of different sub-categories (classes), use them. If not, download images of different categories from the web using the image browser extension and Google Search. You must decide on a category and sub-categories (classes) you would like to use in this assignment. There should be at least 3 sub-categories for the category you have chosen. A few possible categories/sub-categories are shown in the table below.

If you do NOT have access to hundreds of images of different categories, scrape images from the web using image browser extension and Google Search. There are many web image-scraping browser extensions available. Do a Google search to find them. One possible browser extension is called “Wayixia”. Install the “Wayixia” extension on Google’s Chrome browser running on your computer. The procedure of using ‘Wayixia’ extension with Google Search to download images from the web is explained in the class notes (Lesson#5.1). If you like some other browser (instead of Google’s Chrome) and some other extension (instead of “Wayixia”), feel free to use them.

After downloading the images remove poor quality images and those images which do not fit the theme of the sub-category.

Categories	Animals	Airplanes	Dress
Sub-categories	Cat	Large planes	Women dress
	Dog	Midsize planes	Men dress
	Rabbit	Single engine planes	Child Dress
	Mouse		
	Fish		

Make sure that each sub-category has more than 100 images. For example, if you choose ‘Animal’ as a category, then download more than 100 images of each sub-category (cat, dog, rabbit, mouse, fish). Save 5 images (or more) of each sub-category as test images.

Problem#2:

Using your GCP account build an Image Classifier using the AutoML feature of GCP Vision API without using Python/Colab. Use the images downloaded in Problem#1 to **train** your model.

Evaluate the model by displaying the confusion matrix. Predict the class of the **test** images using your image classifier. Report the accuracy of your predictions.

==========

The procedure of building an AutoML image classifier on GCP is as follows.

==========

Data Preparation

- Download bunch of images of a few categories using browser and Google Search + browser extension

- Create a file structure on your computer as follows.
    - <parent-folder>
        - <Sub-folder-1> contains images of sub-category#1
        - <Sub-folder-2> contains images of sub-category#2
        - …
        - …
        - <Sub-folder-n> contains images of sub-category#n

    - Create <parent-folder>.zip file

  The name of the <sub-folder> will become the label for all the images of that sub-category.

==========

Model Building

- GCP/Storage: Create a bucket
- GCP/Vision/ Dashboard/AutoML Vision
- GCP/Vision/Dataset: New Dataset/Single label classification
    - Select :”Upload Images from your computer” & specify the <parent-folder>.zip file
    - Destination: the bucket name created before

- After images are stored in a bucket, train the model
- Evaluate the model
- Predict the class of the test images which the model has not seen before.

==========

Building an AutoML image classifier on GCP will cost a certain amount. Please check the GCP charges on your account before and after you complete this assignment. Make sure you do not deplete the $300 credit you have on your account.