**Readme File**

**IIVP Course Project**

*BTech, IT Department, IIIT Allahabad*

*IIT2018156(Kartik Nema), IIT2018163(Bhupendra), IIT2018172(Prakhar Srivastava),*

*IIT2018175(Ashish Patel), IIT2018177(Shubham Soni)*

# **Detection of Image Splicing using Image Forgery Techniques**

**Pre-requisite details for s/w & h/w,**

The code was developed (in addition to Google Colab) on local system having the following specifications :-​

RAM :- 8 GB​

Intel Core i5 processor​

1 TB hard disk drive​

Operating Systems :- Windows 10 (64 bit), Linux Ubuntu (LTS 18.04)​

In addition to these requirements the system would need a good connection to the internet to run code on Google Colab. Still running in even this setup does give memory errors, if we try to pass the entire dataset all at ones to the PrepProcessing.py code, hence we have run the code for 20 images at a time, by altering the range of iteration for both spliced and authentic image dictionaries. We obtained around 16 pickle files, after running the code for all the images. This amounts to upto 1.2 - 1.5 GB of space, hence the computer must have enough memory to store the pickled dataset.

**To run the project follow the following steps-**

​

**1.** Unzip the file "iivp\_group\_7\_final\_project"

​

**2.** First of all since data present in CUISDE dataset cannot be directly used for training and classification, therefore we first preprocess the data, for this run the file PreProcessing.py:

The goal is to convert the data to an appropriate form and store it in pickle files.

Due to memory Errors it is not possible to run through all images at once, thus this needs to be done in batches of size 10

To create different batches the ranges r1 and r2 are altered as mentioned in the comments.

Total 16 batches (pickle files) are made.

We have already created it and stored in a drive folder link:-

https://drive.google.com/drive/folders/1VFmfk1S5qbzI5y853doqQNrq1Qx58Pt3?usp=sharing

**3.** Classify.ipynb:

To directly load the dataset into the drive, first go to the shared link (it's a folder containing all pickle files) right click on it and select add shortcut.

Now go to collab on mount the drive (code for which is provided in the notebook itself)

Run the cells in the given order.

The last cell will print the accuracy of the model.

The accuracy comes out to be 68.5%

**4.** To get the output for a single test image there are 2 options :-

(i) Run the sample\_output.py file (could be done using terminal python sample\_output.py), it will provide you a window (GUI)

Preconditions :- You need to first download the dataset all 16 pickle files (depending on system, our systems were not able to load more than 7 pickle files) from drive (size > 1 GB), then adjust the paths in the file accordingly (path to ds).

Once that is done run the file, click on the load dataset button and wait for some time while the dataset is loaded, once done a success notification will be provided.

Next, click on the browse files button, to select the target image.

Next, click on the classify image button, again wait for some time before the result is loaded (spliced or authentic)

(ii) Since, the above file uses only a fraction of entire dataset, and runs on the Random Forest Classifier, thus the accuracy is lower, for greater accuracy, we could upload the pickle file generated for an image, in the directory where you have place the sample\_output.py file, now upload this file to the drive link shared above,

Now go into the Classify notebook (step 3), and in the last cell alter the google drive file link to the one you uploaded (before that run all cells in order, except the last one). Run the code and the model will provide it's verdict.

Our model takes an image as input, preprocesses it to extract the features and passes these features to our Machine learning model for classification. The output is binary

(0/1) spliced or not spliced (authentic).