Internship Report on
Image segregation using Python
In
Ameya Intelligent Machine Labs

Supervised by:

Lalit Vidyasagar
Santhiraj K
Raja Emmadi

Submitted by:

Bhargav Malyala

Intern

Bhargav4402@gmail.com

Problem Statement:

Three folders with images will be provided , all stored with unique names of which some images are similar, and some are not. For this reason , two new folders are to be created , one containing all the unique images in it and the other folder containing the duplicate images.

Scenarios

- In WhatsApp, we frequently get same images forwarded from various groups which is redundant data. Tackle this issue, we can use this code to separate the duplicates from group of images and can delete the duplicates if required.

- While doing backup of our images, we usually backup all the images and do not check for duplicates which may in turn eat up our storage, so this code can be used in such scenarios as well.

How does the code work?

The code takes a folder path containing large group of images, and then using “md5” hashing algorithm, each unique image is assigned a hashing ID and is mapped with the respective image using a dictionary. Then, while loping through the images, a list is modified with each image’s hash number and is then copied to a folder containing unique images, and while looping, if the image’s hash number is already in the list, then the image is directly copied to the folder containing duplicate images.

Test Case:

Test Case ID	Test Case Description	Test Steps	Expected Results	Actual Results	Pass/Fail
1	User inputs image, unique and duplicate folders' paths	1. Enter path of folder containing images 2. Enter unique folder path 3. Enter duplicate folder path	Images will be segregated into unique and duplicate folders	As Expected.	Pass
2	User inputs image folder and unique folder paths	1.Enter path of folder containing images 2.Enter unique folder path	New folder for duplicate images will be created and images will be segregated.	As Expected.	Pass
3	User inputs image folder and duplicate folder paths	1.Enter path of folder containing images 2.Enter duplicate folder path	New folder for unique images will be created and images will be segregated	As Expected.	Pass
4	User inputs path of images folder only	1.Enter images folder path	New folders for storing unique and duplicate image will be created and the images will be segregated	As Expected.	Pass

Limitations:

- This code treats the same image with different resolutions as unique images.
- This code will treat a cropped version of an image as unique.
- This code will treat zoomed version of an image as unique.

Future Enhancement:

- It can be developed to treat two same images with different resolutions as same.
- This code can be further developed to compare an image with its cropped version and duplicate it.