Machine Learning Engineer

Task Challenge

This task is to gauge your machine learning knowledge, logic, understanding of the problem and solving it, ability to debug your code.

Problem Statement

You are provided with a dataset of ~5k 512x512 images, your program should accept an 512x512 input image and return N images from the provided dataset similar to the input image. **To solve this problem, building an AutoEncoder model is recommended**.

Link to the dataset

https://drive.google.com/file/d/1VT-8w1rTT2GCE5IE5zFJPMzv7bqca-Ri/view?usp=sharing

Evaluation Method

- Your code submission will be evaluated based code quality and on how accurate it is able to find similar images
  simple score of C/N
  C = no. of similar images returned
  N = no. requested images
- (Optional / Bonus) Clustering the provided dataset into K groups
- Quality of Code based on Modularity, Reusability, Maintainability, Readability

Note

- **Use of pre-trained models are not accepted.**
- Work should be done in Google Colaboratory.
- While submitting the test, provide links to the Colab Notebook(s) and the saved checkpoint Google Drive folder(s) individually or a Single link to the Google Drive folder containing all the Colab Notebook(s) and saved checkpoints.
- Colab Notebook should contain a test function to read in test images from a google drive folder and show the results corresponding to each image available in the folder.
- Test images will be from outside the above provided dataset. For evaluation a test image outside of the dataset would be given and N similar images should be found within the given ~5K dataset.
- For Clustering, K needs to be determined by you.
- Should not manually sort the dataset into categories.
- Codes should be well commented.

Submission

You are expected to send your submission to us within 7 days of receiving the task. You should provide unrestricted accessible links to your test solutions along with all supplementary material in the following link: https://forms.gle/kNmzb53VyHcYYeUM7

While submitting make sure you include the following details:

- Full Name
- Mail ID
- Link(s) to repository for Colab Notebook(s)
- Link(s) to Google Drive folder(s) with saved checkpoints
- Link(s) to output / other supplementary materials uploads

If you have any questions regarding the assignment or need time extension, you can reach us at machinelearning@avantari.org

Sample solution to the above mentioned task

In each box, the 1st row image is the input and 2nd-4th row images are the results.