

## Programming challenge for Machine Learning / Computer Vision Engineer role

At Kopernikus we work with a lot of image and video data. To organise our datasets, we perform different optimisation techniques. One of them is to remove similar looking images from the database.

Your task is to create a program that will find and remove all similar-looking images in a folder. In the attachment, you can find `imaging_interview.py` script that contains functions for image comparison. You can use them to implement your solution.

To compare images, you need to run `preprocess_image_change_detection` function on comparing images and then run `compare_frames_change_detection` with a pair of them. The result will be a score that will show how much images differ from each other. You can experiment with different images to get a sense of how the value depends on images.

Please use the following dataset for your experiments: <https://drive.google.com/file/d/1xJmNedtO1cB9cjcW0gCuab8HJi8IMfZh/view?usp=sharing> .

Please put your solution on GitHub in an open repo with no view restrictions and send us the link to it.

In addition, please describe:

1. What parameters you decided to use for the provided example dataset
2. How you found these values
3. What amount of duplicates script found with these parameters
4. What you would suggest improving to make data collection of unique cases better
5. Any other comments about `imaging_interview.py` or your solution

We wish you good luck with the challenge!