

Test Task

The task is to develop a console tool which finds similar images in a given folder and prints similar pairs. We provide you with an [example dataset for development](#). There are three types of similarity:

- duplicate (images which are exactly the same)
- modification (images which differ by size, blur level or noise filters)
- similar (images of the same scene from another angle)

The images are marked in the dataset with words in the file names that correspond to the type of similarity. The minimal acceptable solution should be able to find “duplicates”. The complete solution should handle all three types of similarity.

Also, you are only allowed to use plain python with the standard library and the following libraries: <https://pillow.readthedocs.io/en/stable/> and <https://www.numpy.org/> . You shouldn't use filenames to identify duplicates and be aware that another dataset will be used for assessing solution performance.

Please send the link to git or [gist](#) with your solution here:

https://docs.google.com/forms/u/1/d/1Vcy_Y3x7mQmubatRNjse16tm32rVrcPIZ31vx30ID10.

Deadline is May 31st 23:59 GMT+3.

If you have any questions regarding the test task, you can reach us at dima.lylyk@railsreactor.com

Example of solution interface with [the example dataset](#):

```
# python solution.py
usage: solution.py [-h] --path PATH
solution.py: error: the following arguments are required: --path
```

```
# python solution.py --help
usage: solution.py [-h] --path PATH
```

First test task on images similarity.

optional arguments:

-h, --help	show this help message and exit
--path PATH	folder with images

```
# python solution.py --path ./dev_dataset
4_similar.jpg 4.jpg
```

11_modification.jpg 11.jpg
11_modification.jpg 11_duplicate.jpg
6_similar.jpg 6.jpg
11.jpg 11_duplicate.jpg
15_modification.jpg 15.jpg
1.jpg 1_duplicate.jpg

Good luck!