TRIFORK.

Problem description:

We're building a model for object detection. We want to detect different kinds of vegetables. We already have a dataset but we just realized that it's in the wrong format, this is why we want to convert the dataset to our preferred format.

The task that you are assigned is to prepare the dataset for training, this includes converting the annotations to the chosen format, resizing the images to (1024, 768) and splitting the dataset into train/test/validation.

Requirements:

The script should accept 3 arguments as inputs:

- The path to the COCO annotations
- The path to the folder of the input images
- The path where to save the results

The annotations are found in COCO format. It means that there will be a folder with the images and a json with the annotations. We want the data in a YOLO format.

The results should be saved in the following format, inside the path specified for saving the results:

- A folder *images* should contain three different subfolders: train/test/validation which will contain the corresponding split of resized images.

- A folder *annotations* should contain three different subfolders: train/test/validation which will contain the corresponding annotation files, matching the names on the corresponding image subfolder.

Additional information

In this folder you will find 138 images in .jpg format and a json file with the COCO annotations for the images

Submission

You will have 1 week to submit the assignment from the moment you receive it.
This task can be solved in much shorter time, but we want to extend the period of submissions so you can improve it as much as you want.

Please remember that this assignment is what will give us insights on how you reason, and how you code, so it's also a way for showcasing your skills.

Best of luck and let the coding begin!
PS. For the submission we prefer if you could share a GitHub repo, instead of sending the script.