Continental
The Future in Motion

HACKATHON

DATA SCIENCE

CHALLENGE INFORMATION

NEB-FEUP/ICBAS
núcleo de estudantes de bioengenharia

IEEE
University of Porto - Student Branch

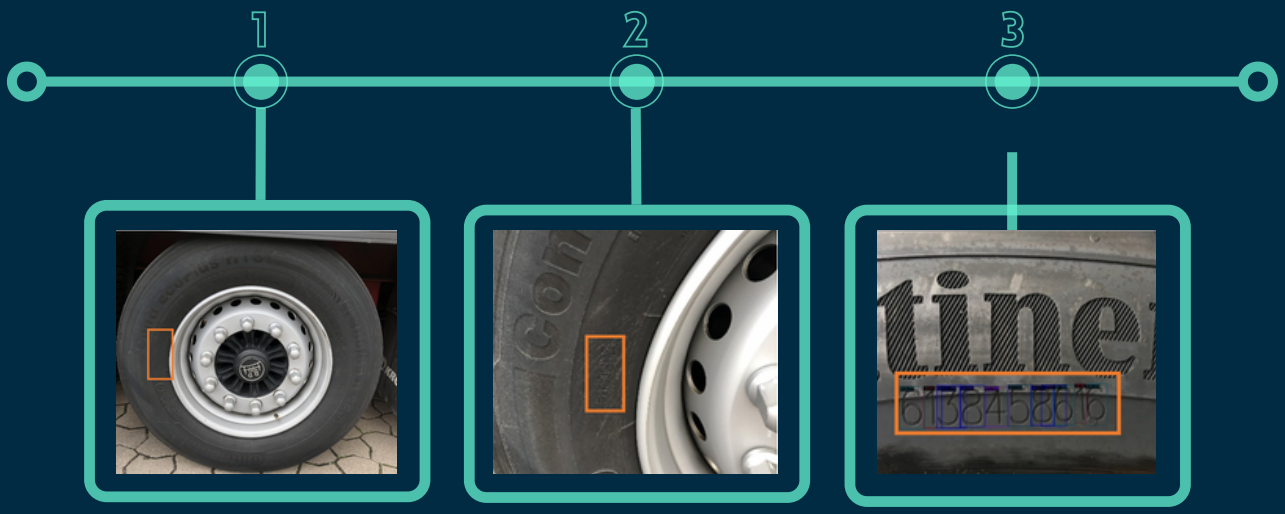
ni
aefeup

PROBLEM

The problem consists in the identification of tires through images.

GOAL

The goal is the Optical Character Recognition (OCR) of serial tire identification number (serial number) on Commercial Specialty Tires (CST)



DATASET

The dataset has 369 images of Commercial Vehicle Tires (CVT), cropped in the serial number region, in the data folder. The annotations are in COCO format:

- labels_train.json: 295 train set annotations (80%)
- labels_test.json: 74 test set annotations (20%)

The test data will be made available for 15 minutes on sunday morning. The teams have 15 minutes to run the inference and submit the predictions in the expected format.

EVALUATION

The Levenshtein distance is a string metric for measuring the difference between two strings. It is calculated as the minimum number of single-character edits necessary to transform one string into another.
Participants should provide a csv file for the test set in the following format:

image_name	string_1_prediction	string_2_prediction
img1	1234567	
img2	122345	132456
img3	12345678	

A script for exporting predictions in this format, given a directory of test images, is given at github

Teams will also be evaluated with additional criteria:

- Pitch
- Technical solution
- Collected data
- Discussion after pitch
- Quality of deliverables

To gain extra points, you can go outside and increase your dataset: acquire more CST tire images and evaluate them on your model! You will be evaluated on how many images you acquire and annotate.

RECOMENDATIONS

Create a new google account for the team, as it will make it easy to work together with Colab. This account will also be used to put all the documents in its drive. This drive must then be shared with us by email.

COMUNICATION

Discord channel: https://discord.gg/BFTgYu4D
Google Colab: https://colab.research.google.com/?utm_source=scs-index

CONFIDENTIALITY

The Dataset is confidential: the participants take responsibility not to distribute, publish or delete the content.

DELIVERABLES

Participants are expected to share their GoogleDrive's link with the name [TEAM_NAME].zip to our email.

FILE STRUCTURE

- *New_data:* containing additional images collected and annotations in COCO format
- *Code:* containing all the scripts used for the solution
- *Pitch presentation*

USEFUL LINKS

- https://imglab.in/
- https://voxel51.com/docs/fiftyone/tutorials/open_images.html
- https://www.kaggle.com/code/artgor/object-detection-with-pytorchlightning/notebook