FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
MIEIC - 2018/2019

COMPUTER VISION

PROJECT Nº 2
Porto's Landmarks

Summary

The main goal of this project is to develop a system that can recognize some of Porto's landmarks in photographs taken in the wild (different cameras, illumination conditions, perspectives, etc.). The system developed for this project should detect the 5 landmarks presented in Figures 1-5.

Figure 1 – Torre dos Clérigos

Figure 2 – Casa de Serralves

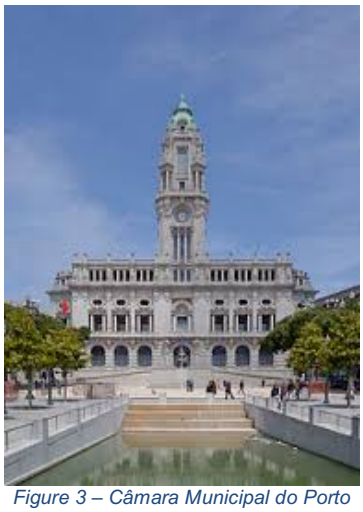

Figure 3 – Câmara Municipal do Porto

Figure 4 – Casa da Música

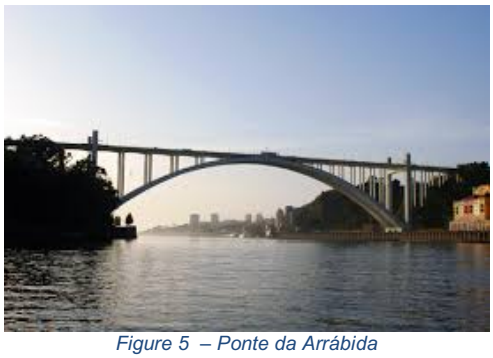

Figure 5 – Ponte da Arrábida

There is no restriction on how the system should be built. Some approaches to create such a system include (but are not limited to):

- Feature detectors and descriptors (e.g. SIFT, color histograms) + classifier
- Dictionary-based representation (e.g. Bag of Visual Words) + classifier
- Deep learning methods (e.g. CNN)[1]

Consider nevertheless the following incremental levels of complexity for the system (and corresponding weight in the grading):

- Tier 1 (75%): Classify images containing the 5 classes of Porto's landmarks; the system should also be able to reject an image that doesn't contain any of the landmarks.

[1] Take into consideration that the training time will take a significant amount of time if GPUs are not used.

- Tier 2 (20%): Detect where in the image the landmark is located; the output should be a bounding box, defined by position (x,y) and size (width, height), such as the one shown in Figure 6.
- Tier 3 (5%): Detect the viewpoint of the photograph; consider only one landmark and three different viewpoints.

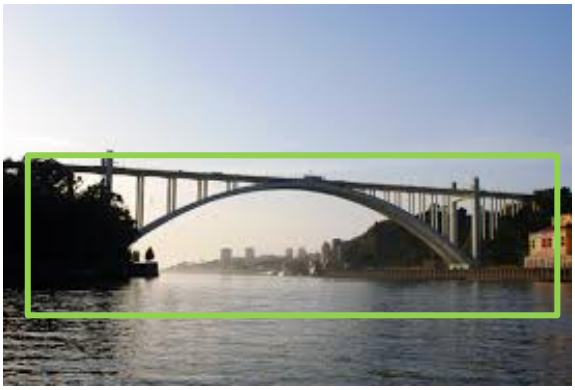

Figure 6 – Detection of a landmark, identified by a bounding box

An objective evaluation of the system (tier 1 and optionally tier 2) should also be presented, based on its accuracy, confusion matrices, or other relevant metrics. Note that the grading of the project will not be defined by the system's performance but by its complexity and methodologies.

In order to train and evaluate the system, a dataset with annotated examples needs to be created. Suggestion: crowdsource this task among all groups, and those that make significant contributions to the dataset will have bonus points.

Scientific Paper and Delivery

A short report must be elaborated in the format of a scientific paper (max. 3 pages), including:

- Brief introduction to the problem, including references about the state of the art;
- Description of the developed system;
- Possible additional specifications or improvements;
- Results of the system, namely percentage of categories correctly identified and other measures considered relevant;
- Discussion about the overall performance of the system and possible situations where it fails;
- Conclusions and future improvements.

The paper can be written in English or Portuguese and should be based on the model available in Moodle. The code, with meaningful comments, should be presented in annex.

The work must be submitted at the Computer Vision page, in the UP Moodle site, until the end of the day January 4, 2019.

Bibliography and other support material

- Sampling Strategies for Bag-of-Features Image Classification. E. Nowak, F. Jurie, and B. Triggs. ECCV 2006.
- Imagenet classification with deep convolutional neural networks, A. Krizhevsky, I. Sutskever, and G. E. Hinton. *Advances in neural information processing systems*, pp. 1097-1105. 2012.
- Faster R-CNN: Towards real-time object detection with region proposal networks . S. Ren, K. He, R. Girshick, and J. Sun. In *Advances in neural information processing systems*, pp. 91-99. 2015.
- You only look once: Unified, real-time object detection. J. Redmon, D. Santosh, R . Girshick, and A. Farhadi.. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788. 2016.