FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
MIEIC – 2020/2021

COMPUTER VISION

PROJECT Nº 2

Artwork recognition

(based on the iMet Collection 2019 competition)

Summary

Figure 1 – Examples of images that are part of the dataset
[source: https://sites.google.com/view/fgvc6/competitions/imetcollection-2019]

The broad goal of this project is to develop a system that can perform an automatic classification of artwork shown in images. The images contain artefacts that are part of the collection of The Metropolitan Museum of Art in New York, also known as The Met - some examples of the images are shown in Figure 1. The project relies on the datasets provided by the iMet Collection 2019 competition that was proposed in The Sixth Workshop on Fine-Grained Visual Categorization [https://sites.google.com/view/fgvc6/competitions/imetcollection-2019]. Each image in the complete dataset [https://www.kaggle.com/c/imet-2019-fgvc6] has at least one attribute label from a label set of 1103 attributes. The dataset is divided in training (109274 samples) and test (7443 samples) sets. However, the annotations of the test set are not available and, therefore, in this project the training set should be adequately divided into training/validation and test sets. Also, only a subset of the labels will be used, as explained in the tasks description.

The photographs are often centred for objects, and in the case where the museum artefact is an entire room, the images are scenic. The dimension of each image is normalized such that the shorter dimension is 300 pixels. Each object was annotated by single worker without a verification step (can be noisy). The workers were advised to add multiple labels from an ontology provided by The Met.

The project will be divided into three tasks, with increasing difficulty and a different weight in the final grade:

- **Multiclass classification** (50%)
  The goal of this task is to implement and compare two or more methods/architectures that classify the artwork images (as shown in Figure 1) as one class. The original dataset was filtered to a smaller dataset containing images that only have one of 50 possible labels, as defined in the file `multiclass.csv` available in Moodle. Note that the dataset is highly unbalanced. Some possible approaches include, but are not limited to:
  - Dictionary-based representation (e.g. Bag of Visual Words) [1] + classifier
  - Deep learning methods (e.g. CNN)[1] [2]

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

- **Multi-label classification** (25%)
  This task focuses on implementing and evaluating one method/architecture to predict the labels associated with each image. The prediction for each image should contain at least one of the possible labels. A subset of the original dataset should be used, as defined in the file `multilabel.csv` available in Moodle. An adaptation of the (most successful) method applied in the first task should be considered for this task. To evaluate the results, consider using the F2 metric, that was used in the original competition [https://www.kaggle.com/c/imet-2019-fgvc6/overview/evaluation].

- **Artwork detection** (10%)
  The final task is related to artwork detection, in particular paintings (as shown in Figure 2). There is no dataset readily available for this task and creating a new dataset would require significant effort. Therefore, two options are possible: 1) train a detection model (e.g. YOLO [3]) with a synthetic dataset, generated by combining different backgrounds with randomly distributed paintings on top, or 2) use a regular dataset of paintings and apply class activation maps to make a prediction of the object location, as proposed by Zhou et al. [4]. The second approach is more straightforward and is, therefore, recommended. Consider only a qualitative evaluation in this case.

Figure 2 – Artwork detection, where each painting is identified by a bounding box (in green) [source: Wikipedia]

An objective evaluation should be presented for the classification tasks, based on relevant metrics. Note that the grading of the project will not be defined by the system's performance but rather by the correctness of the adopted methodologies.

Scientific Paper and Delivery

A short report (10%), must be elaborated in the format of a scientific paper (max. 3 pages), including:

- Brief introduction to the problem, including some references to the state of the art;
- Description of the methodology, including evaluation;
- Results of the evaluation using metrics adequate for the specific problem;
- 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 work (paper+code) should be submitted at the Computer Vision page (Moodle U.Porto) until the end of the day **June 11, 2021**. An anonymous version of the paper should be submitted in EasyChair (similarly to the first project) until the same deadline.

After the submission, each student should review one anonymous paper of another group (5%) until the end of the day **June 18, 2021**.

Bibliography

[1] Sampling Strategies for Bag-of-Features Image Classification. E. Nowak, F. Jurie, and B. Triggs. In *Proceedings of the European Conference on Computer Vision*, pp. 490-503. 2016.

[2] 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.

[3] 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.

[4] Learning Deep Features for Discriminative Localization. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2929. 2016.