Project Milestone

**From Object detection to Image Captioning in Retail Environment**

***Authors***: Bucciarelli Davide <271102@studenti.unimore.it>,

Fino Alessandro <313424@studenti.unimore.it>,

Pavone Cosimo <272636@studenti.unimore.it>

***Introduction:***We will study the problem of object detection and description at a supermarket using computer vision and natural language processing methods. Our plan for supermarket object detection and description consists of detecting the objects in the scene and provide a description of them. This approach can benefit from a combination of traditional image processing and deep learning-based techniques to increase the effectiveness and precision of inventory tracking and management in retail settings.

***Problem statement:***  
For our pipeline, we will use publicly available datasets such as the Freiburg groceries dataset [1], SKU110K [2], RP2K [3] and Grocery Products dataset [4] for object detection and recognition. In the first phase of the project, we started working on the last-mentioned dataset using traditional image processing techniques like Canny to detect the edges of the objects we want to identify in the scene. The dataset contains 8350 training images of grocery products, organized in 80 hierarchical classes, and 680 annotated test images of supermarket shelves. In general, our overall plan can be described as:

* The images will first be pre-processed using traditional image processing techniques like Canny to detect the edges of the objects we want to identify in the scene.
* Then we will find the patches on the image that contain the objects to feed them to a CNN to classify the single objects. We will also use a retrieval component to enhance the performance of our network.
* After that, we will identify the parts of the image that belong to the classified object, to then perform some graph based global reasoning to find relationships between the objects.
* Finally, we will create natural language descriptions of the scene using rule-based template filling.

To evaluate our results, we will use metrics such as IoU, precision, recall, and F1 score for object detection and recognition, and BLEU or ROUGE scores for natural language generation and image captioning.

***Technical approach:***   
Regarding multiple object detection for finding items on retail shelves, we decided to use an approach based on deep learning since we found out that classical methods are not feasible for this use case, due to the complexity of the scene, in which we have many objects and borders that we don’t want to detect.

On the other hand, for single object detection, classical methods are good enough. In particular:

* A gaussian filter to smooth out the image and remove some of the noise.
* The canny edge detection algorithm to detect borders in each of the R, G, B channels of the image to find most of the edges. The edge maps are then merged and binarized using thresholding.
* Dilation and erosion to connect into contours edges that do not form a perfectly closed line.
* Contours are then detected from the closed edges.
* We then find the biggest contour and draw a bounding box around it.

Intermediate/Preliminary result: (TODO)

[1]: <https://github.com/PhilJd/freiburg_groceries_dataset>

[2]: <https://github.com/eg4000/SKU110K_CVPR19>

[3]: <https://www.pinlandata.com/rp2k_dataset/>

[4]: Marian George, Christian Floerkemeier, "Recognizing Products: A Per-Exemplar Multi-Label Image Classification Approach", ECCV 2014 <http://www.vs.inf.ethz.ch/publ/papers/mageorge_products_eccv2014.pdf>

[5]: <https://arxiv.org/pdf/1810.01733.pdf>

[6]: <https://arxiv.org/pdf/1904.00853.pdf>

[7]: <https://arxiv.org/pdf/2006.12634.pdf>

[8]: <http://ais.informatik.uni-freiburg.de/publications/papers/jund16groceries.pdf>

[9]: <https://openaccess.thecvf.com/content_CVPR_2019/papers/Chen_Graph-Based_Global_Reasoning_Networks_CVPR_2019_paper.pdf>  
[10]: J. Canny, “A computational approach to edge detection,” IEEE Trans.Pattern Anal. Mach. Intell., vol. 8, no. 6, p. 679–698, 1986  
https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.420.3300&rep=rep1&type=pdf