Object counting and description in a supermarket

Shankho Boron Ghosh, Michele Bellitti, Jacopo Pecchini

I. Introduction

The problem investigated in this project is object counting and description in a supermarket. The proposed pipeline has numerous practical applications in the retail industry, such as monitoring stock levels, optimizing inventory management, and improving customer experience. The development of such a pipeline can contribute to creating more efficient and accurate systems, which could benefit various sectors beyond retail, including logistics, transportation, and security.

II. Literature Survey

A literature review will be conducted to provide context and background in relevant domains and categories, which include Product Recognition Techniques [1] [2] [3] [4], Image Classification Techniques [5], and Transformer Networks [6]. By examining these studies, the proposed pipeline can be contextualized within existing research on object recognition, image classification, and machine learning techniques in the retail industry.

III. Datasets

A preliminary literature review of relevant studies yielded a number of potential datasets, including the Hierarchical Grocery Store Image Dataset with Visual and Semantic Labels [7], the Freiburg Groceries Dataset [8], the Grocery Dataset [2] and the Grozi dataset [5]. The decision of which dataset to employ will be contingent upon a thorough analysis of specific project requirements, such as the characteristics of the objects to be counted and described, the quality and diversity of the images, and the availability of semantic labels.

IV. Methodology

The preliminary proposed pipeline will consist of multiple stages. Firstly, classical image processing operators such as deblurring filters like Gaussian or Median and morphological operators like Hough Transform and Canny Edge Detector will be utilized to preprocess the supermarket scene images into individual shelves. Each shelf would then be passed on to the second stage, which focuses on segmentation and product detection using a CNN-based deep-learning network like ResNeXt, EfficientNet, and Wide ResNet. Subsequently, brand or sub-categorical recognition can be achieved in the third stage using Machine Learning classifiers like SVM, decision trees, or KNN. During the training of combined second and third stages, ensemble methods like voting, bagging, boosting, and stacking can be used to improve the performance of individual classifiers, and further improvements in generalization are possible by using Cross-Validation to avoid making the model optimal against an individual set of validation data. The fourth stage focuses on assimilating the object position relationships in the 3D space of the obtained bounding boxes from the previous stage by using geometric-based homography transforms like Perspective-n-Projection or graph-based spatial relations. Finally, in the final fifth stage, each bounding box has its corresponding relation in the 3D space, which can be stored in appropriate data structures like graphs or trees, which on the query from the user, can generate a natural language description of the scene.

V. Evaluation

To evaluate the performance of the proposed method, we will measure the accuracy and precision of the object counting and description using standard metrics like Precision, Recall and F1-Scores from the confusion matrix, losses and accuracy over the training, validation and test sets can provide insight into improvements with respect to the current state of the art methods including but not limited to hierarchical grocery store image datasets, deep learning pipelines, per-exemplar multi-label image classification, and vision transformers. We expect the results to be consistent across different supermarket settings and lighting conditions; hence, we plan to train and infer our proposed pipeline over multiple datasets. Additionally, we also plan to compare our results against traditional methods such as manual counting and inventory tracking.

References

[1] A. Tonioni, E. Serra and L. Di Stefano, "A deep learning pipeline for product recognition on store shelves," 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), Sophia Antipolis, France, 2018, pp. 25-31.

[2] Varol, Gül, and Rıdvan Salih Kuzu. "Toward retail product recognition on grocery shelves." Sixth International Conference on Graphic and Image Processing (ICGIP 2014). Vol. 9443. SPIE, 2015.

[3] Sinha, Ankit, Soham Banerjee, and Pratik Chattopadhyay. "An improved deep learning approach for product recognition on racks in retail stores." arXiv preprint arXiv:2202.13081 (2022).

[4] J. D. Domingo, R. M. Aparicio and L. M. G. Rodrigo, "Cross Validation Voting for Improving CNN Classification in Grocery Products," in IEEE Access, vol. 10, pp. 20913-20925, 2022

[5] George, Marian, and Christian Floerkemeier. "Recognizing products: A per-exemplar multi-label image classification approach." Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13. Springer International Publishing, 2014.

[6] Liu, Shu, et al. "GRVT: Toward Effective Grocery Recognition via Vision Transformer." Advances in Computer Graphics: 39th Computer Graphics International Conference, CGI 2022, Virtual Event, September 12–16, 2022, Proceedings. Cham: Springer Nature Switzerland, 2023.

[7] M. Klasson, C. Zhang and H. Kjellström, "A Hierarchical Grocery Store Image Dataset With Visual and Semantic Labels," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2019, pp. 491-500

[8] Jund, Philipp, et al. "The freiburg groceries dataset." arXiv preprint arXiv:1611.05799 (2016).