AUTOMATED PRODUCT DETECTION AND RECOGNITION ON RETAIL SHELVES USING DEEP LEARNING

PROJECT PROPOSAL

Chien-Wei Weng & Ke Chen
Yingzhou Fang & Yuhong Li
CentraleSupélec, Université Paris-Saclay
MSc Data Sciences and Business Analytics
Email: chien-wei.weng@student-cs.fr
ke.chen@student-cs.fr
yingzhou.fang@student-cs.fr
yuhong.li@student-cs.fr

1 MOTIVATION

With the rise of new retail formats, traditional shelf management methods relying on manual inspection have become inefficient and error-prone. For large supermarkets and warehouse stores, manual shelf audits are time-consuming, labor-intensive, and frequently result in missed stockouts or delayed restocking, ultimately degrading the customer experience.

Moreover, misplaced or disorderly shelving is common in practice. Without automated systems to identify what products are on shelves and whether they are correctly positioned, retailers struggle to maintain inventory accuracy and display standardization.

To address these challenges, we propose a deep learning-based automated product recognition system. Our two-stage pipeline performs object detection to localize products and classification to identify them. This system provides the foundational capability for downstream retail applications, including:

- **Stockout detection:** Comparing detected products against expected planograms to identify empty shelf spaces
- **Misplacement detection:** Verifying products are in their correct shelf locations

While our current work focuses on accurate detection and identification, it establishes the technical infrastructure necessary for these advanced inventory management tasks as future work (5).

2 METHODOLOGY

We implement a two-stage deep learning pipeline for automated product recognition on retail shelves. Stage 1 performs object detection to localize products, and Stage 2 classifies each detected product.

2.1 STAGE 1: OBJECT DETECTION

- **Dataset:** For object detection, we use the SKU-110K dataset (1), containing 11,762 densely-packed retail shelf images with bounding box annotations for product localization
- **Model:** YOLOv5 (2) pretrained on COCO dataset, fine-tuned on our retail dataset
- **Output:** Bounding boxes (x, y, w, h) for each product instance on the shelf

2.2 STAGE 2: PRODUCT RECOGNITION

- **Dataset:** Grocery Store Dataset (3) with 876/111/811 train/val/test images across 43 fine-grained product classes

- **Model:** ResNet-18 (4) pretrained on ImageNet, fine-tuned on our retail dataset
- **Training:** Data augmentation (Gaussian blur, random crop, brightness/saturation variations) to bridge gap between reference and shelf images
- **Output:** Product class labels for each detected region

2.3 Integration

Detected bounding boxes from Stage 1 are cropped from the shelf image and fed to Stage 2 classifier. Final output consists of shelf images with labeled products showing both location and identity.

3 Evaluation Strategy

Due to dataset mismatch (SKU-110K lacks product labels, Grocery Store lacks shelf scenes), we evaluate stages separately and demonstrate integration qualitatively:

- **Stage 1 (Detection):** Mean Average Precision (mAP) at IoU threshold 0.5 on SKU-110K test set (2,941 images)
- **Stage 2 (Recognition):** We compare fine-tuned ResNet-18 against the pretrained model without fine-tuning to demonstrate improvement on our domain. The evaluation metrics: Top-1 Accuracy (main metric), Precision/Recall/F1 per class, confusion matrix for 43 classes on Grocery Store test set (811 images)
- **Demonstration:** Qualitative evaluation on 5–10 sample shelf images showing input/output comparisons with detected boxes and classified product labels

References

[1] Goldman, E., Herzig, R., Eisenschtat, A., Goldberger, J., & Hassner, T. (2019). Precise detection in densely packed scenes. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5227–5236.

[2] Jocher, G. (2020). YOLOv5 by Ultralytics (Version 7.0) [Computer software]. `https://github.com/ultralytics/yolov5`

[3] Klasson, M., Zhang, C., & Kjellström, H. (2019). A hierarchical grocery store image dataset with visual and semantic labels. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 491–500.

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

[5] Šikić, F., Kalafatić, Z., Subašić, M., & Lončarić, S. (2024). Enhanced out-of-stock detection in retail shelf images based on deep learning. *Sensors*, 24(2), 693.