**Supermarket Shelf Product Identification using Computer Vision**

**1. Project Objectives and Purpose**

**Business Context**

In the retail industry, ensuring that products are correctly placed on supermarket shelves according to contractual agreements is a critical challenge for suppliers. Traditionally, this process requires suppliers to deploy personnel to manually verify product placements, which is time consuming and costly.

This project aims to develop a proof of concept with AI powered system that automatically identifies and highlights the same products on a supermarket shelf using different colors. The goal is to provide suppliers with an automated solution to monitor product placement remotely through images, eliminating the need for physical store visits. By leveraging computer vision, the system enables suppliers to:

* Verify if their products are correctly displayed on supermarket shelves.
* Identify potential issues in product distribution.
* Reduce operational costs by minimizing manual inspections.

**Expected Achievements**

* **Product Detection:** Identify and locate products in images of supermarket shelves.
* **Product Grouping:** Cluster visually similar products using unsupervised learning techniques.
* **Visual Representation:** Highlight detected products in different colors to distinguish between similar items.
* **Remote Monitoring:** Provide suppliers with image-based insights to ensure product placement compliance.

**2. Methodology**

To achieve the objectives, a two-step AI pipeline was developed:

1. Product Detection using YOLOv10
2. Product Grouping using Feature Embeddings and Clustering

**2.1 Object Detection: YOLOv10 Fine-Tuning**

The first stage of the model leverages **YOLOv10**, a state-of-the-art object detection model, to identify and localize products on supermarket shelves. Since product detection is a complex task with numerous small objects appearing in dense arrangements, the model was fine-tuned using the SKU-110K dataset—a widely used benchmark dataset for product detection in retail environments.

**Fine-Tuning Strategy:**

* **Dataset:** SKU-110K (A dataset containing 110,000 images of store shelves).
* **Training Epochs:** 10 epochs (instead of the recommended 100) as an initial proof-of-concept.
* **Loss Function & Optimizer:** Standard YOLO loss function with Adam optimizer.

This approach allowed for a rapid first iteration of the model while maintaining reasonable detection performance.

**2.2 Feature Embeddings for Product Grouping**

Once products were detected, the second stage aimed to group similar products together. Due to the lack of labeled training data, a feature extraction-based approach was used instead of traditional classification models.

Three pre-trained deep learning models were combined to generate product feature embeddings:

* **CLIP:** Extracts semantic relationships between images and text.
* **DINOv2:** A self-supervised vision transformer designed for feature-rich embeddings.
* **ResNet-18:** A lightweight but effective CNN model for capturing visual features.

**Combination Strategy:**

Each model contributed equally to the final embedding using weighted averaging:

* CLIP: 0.33
* DINOv2: 0.33
* ResNet-18: 0.33

These combined embeddings captured diverse feature representations of products, allowing the system to recognize visual similarities effectively.

**2.3 Clustering with HDBSCAN**

With feature embeddings extracted, the next step was to cluster similar products. HDBSCAN was used to group the products without requiring a predefined number of clusters.

* **Why HDBSCAN?** Unlike K-Means, HDBSCAN allows for variable cluster sizes and can detect outliers, which is crucial for supermarket shelves containing both grouped and unique items.

Finally, in each image, similar products were assigned a unique color for easy visual distinction, allowing suppliers to quickly assess the product distribution.

**3. Key Findings**

* **Image Quality Matters:** Clear, high-contrast images significantly improved product identification accuracy.
* **Complex Shelves Pose Challenges:** When shelves are cluttered or contain visually complex arrangements, the model struggled to separate similar-looking products.
* **Embedding Combination Improved Recognition:** The multi-model approach (CLIP + DINOv2 + ResNet-18) helped detect similarities, though fine-tuning could further improve performance.
* **Simple Product Designs Are Easier to Identify:** Products with minimal details or distinctive shapes were easier to detect compared to those with intricate packaging.

**4. Limitations and Next Steps**

**4.1 Limitations**

* **Lack of Labeled Data:** None of the models were fully fine-tuned due to the absence of labeled training data specific to the supplier's products.
* **Limited Clustering Precision:** While HDBSCAN grouped similar products effectively, some clusters contained misclassified items due to visual ambiguity.
* **Challenges in Complex Environments:** Products positioned at extreme angles or partially obscured were harder to detect reliably.

**4.2 Next Steps**

1. **Self-Supervised Learning:** Implement self-supervised learning models to overcome the lack of labeled data, enabling better product representation learning. Using Data augmentation for the self-supervised learning models.
2. **Improved Fine-Tuning:** Collect a small, labeled dataset for targeted fine-tuning of feature extraction models.
3. **Enhanced Clustering Techniques:** Explore advanced techniques such as contrastive learning to improve clustering precision.
4. **User-Friendly Dashboard:** Develop a web-based interface where suppliers can upload images and automatically receive annotated results with product locations.

**5. Conclusion**

This project presents a first-step AI solution for supermarket product monitoring, allowing suppliers to verify product placements remotely. By combining object detection with unsupervised clustering, the system identifies and highlights similar products in shelf images. While the current approach provides valuable insights, further fine-tuning and self-supervised learning methods will enhance accuracy and scalability. This technology has the potential to redefine retail monitoring for suppliers to comply with its shelf auditing, reducing operational costs and ensuring compliance with display agreements in supermarkets.