231-ICS-619 Project

Video-based Retail Store Shelf Monitoring

Khaled Almutairy

Mamoun Alghaslan

Introduction

- **Seamless**, Saudi Arabia's premier event, highlighted innovations in payments, fintech, **retail**, e-commerce, home delivery, and digital marketing.
- A focus on retails future featured cashier-less stores showcasing **AI-powered shelf monitoring**, addressing a key industry trend.

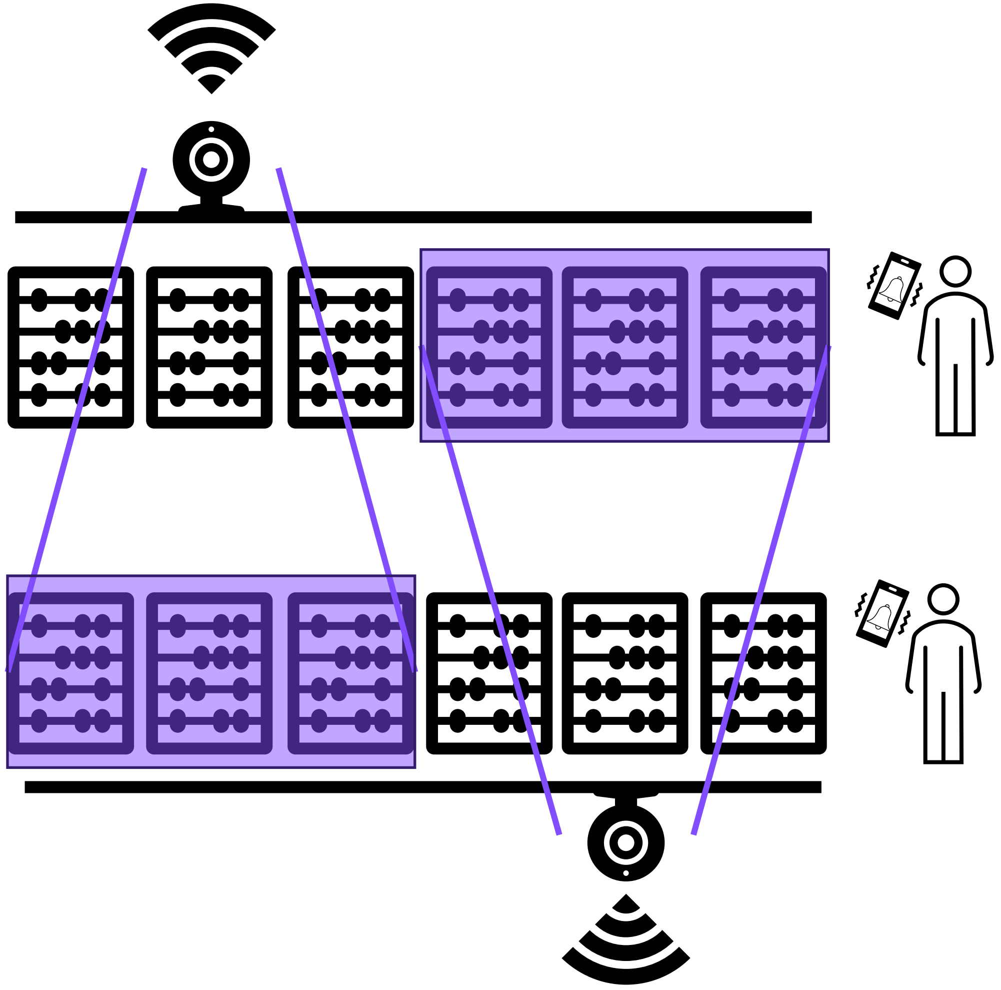
Challenges in Retail Inventory Management

- **Manual Inspection Process**: relying on labor-intensive manual work, with employees counting and correcting products on shelves, preparing OOS inventory, retrieving items from storage, and replenishing shelves.
- **Financial Implications**: High expenditure on inventory management and potential customer loss due to frustration or simply not finding the convenience
- **Limited stock monitoring**: Currently, supermarkets rely on distributers to manually count and sort the shelf products and report stock status.

Challenges in current approaches

- Limitations of **traditional methods** (ML, image processing)
- Deep learning's need for **large annotated datasets**
- Difficulty in achieving high precision for Out-Of-Stock and Misplaced detection due to **shelf layout**, **object shapes**, and **occlusion**.

Our solution



Our proposed approach

- **Collecting** images directly from shelves
- **Labelling** products and empty spaces with bounding boxes
- **Fine-tuning** a pre-trained object detection model on a supermarket dataset
- **Building** an efficient system around the detection model

BACKGROUND & EXISTING SOLUTIONS

Background

Object Detection and Recognition

- Detection involves locating and defining the best bounding box
- Recognition refers to classification

Model Implementation Options

- Traditional Image Processing such HOG, SIFT, SURF, BRISK and FAST
- Deep Learning Networks

Metrics

Object Detection

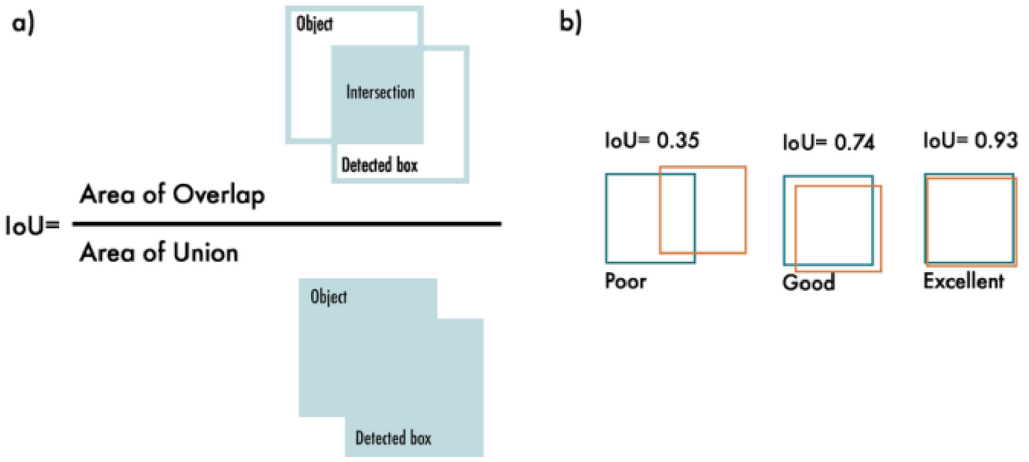


Figure 1.3: a) Intersection over Union b) Examples of different IoU [18]

Object Recognition

Table 1.2: Evaluation Metrics

Metric	Equation	Explanation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Measures overall correctness of predictions.
Precision	$\frac{TP}{TP+FP}$	Measures how many of the positive predictions are accurate.
Recall	$\frac{TP}{TP+FN}$	Measures how many of the actual positives are correctly predicted.
Mean Average Precision	$\frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i$	A comprehensive metric that combines precision and recall for all classes. *AP* is a precision-recall trade-off measurement.

Existing Solutions

Deep Learning methods

- Faster R-CNN
  - A two-stage architecture, first generating region proposals and then predicting class labels, and bounding box coordinates
  - Generally slower due to its two-stage process
- YOLO
  - A single-stage architecture, simultaneously predicting class probabilities and bounding box coordinates in one pass through the network.
  - Known for its real-time object detection capabilities, achieving faster inference speeds.

Other methods

- Classical Methods
  - Do not generalize well when applied to new environments.
- Depth Estimation Method
  - Fail to detect product misplacement

DATA COLLECTION & EXPERIMENTS

Data Collection

Danube Supermarket

- Focused on five shelves ~150 products
- Each product was removed to simulate OOS, and 32 product were moved around five times to simulate MIS
- A total of 310 images were collected and augmented
- Data set was split 90/5/5%, with 15 samples of OOS and MIS in val./test
- Roboflow for annotation

Figure 3.1: Example of a MIS sample (top row) with a bounding box in yellow

Experiments

YOLOv5

Class	Images	Instances	P	R	mAP50	mAP
all	40	6020	0.362	0.408	0.55	0.416
MIS	40	25	0.356	0.16	0.246	0.139
OOS	40	15	0.0891	0.0667	0.409	0.255
Product	40	5980	0.64	0.999	0.994	0.853

Table 3.1: YOLOv5 performance metrics on test set

YOLOv7

Class	Images	Instances	P	R	mAP50	mAP
all	40	6020	0.916	0.893	0.918	0.672
MIS	40	25	1	0.68	0.759	0.42
OOS	40	15	0.753	1	0.995	0.752
Product	40	5980	0.996	0.999	0.999	0.844

Table 3.2: YOLOv7 performance metrics on test set

Paperspace Jupyter Notebook, Quadro-P6000 GPU, 5- and 2.5-hours training time respectively

SYSTEM DESIGN

Model fine-tuning

YOLOv8s

Class	Images	Instances	P	R	mAP50	mAP
all	30	4523	0.987	0.909	0.957	0.767
MIS	30	15	1	0.727	0.882	0.648
OOS	30	15	0.962	1	0.995	0.78
Product	30	4493	0.998	1	0.995	0.873

Table 4.1: YOLOv8 performance metrics on test set

Results

System Design

Components

1. Detection model
2. Model API
3. Web Server
    1. Web-site server
    2. Web worker
    3. Broadcasting service
4. IP Cameras

Diagram

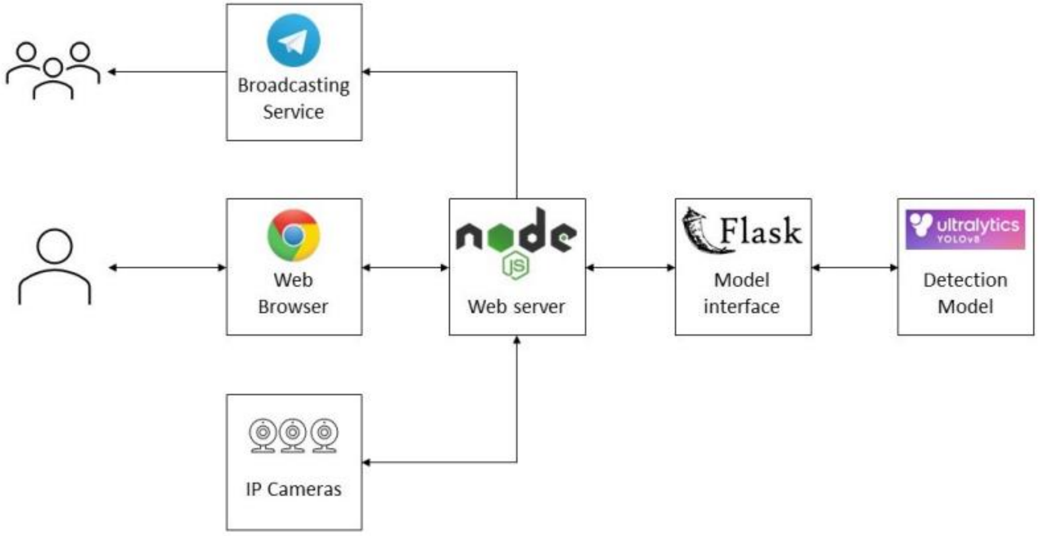


Figure 4.2: System high-level design

Prototype

Design

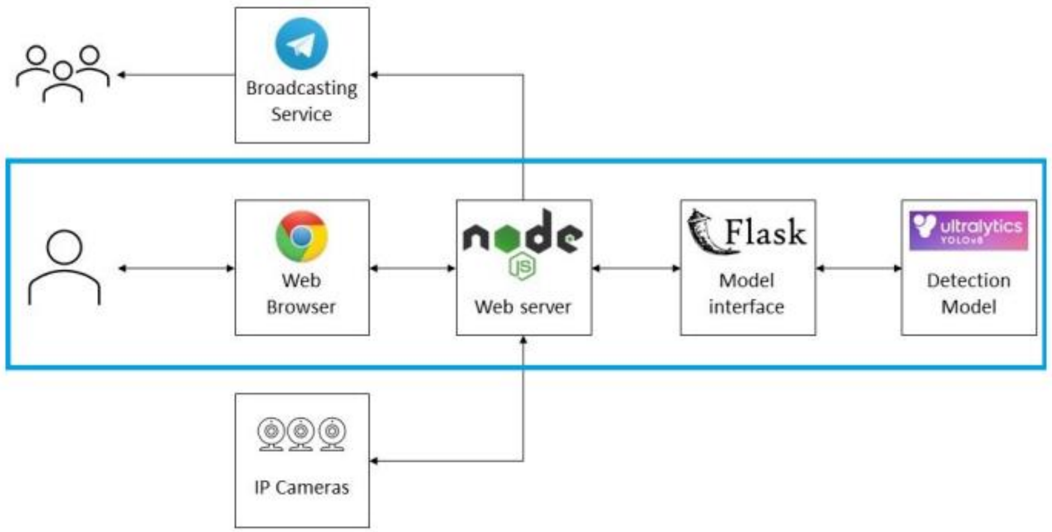
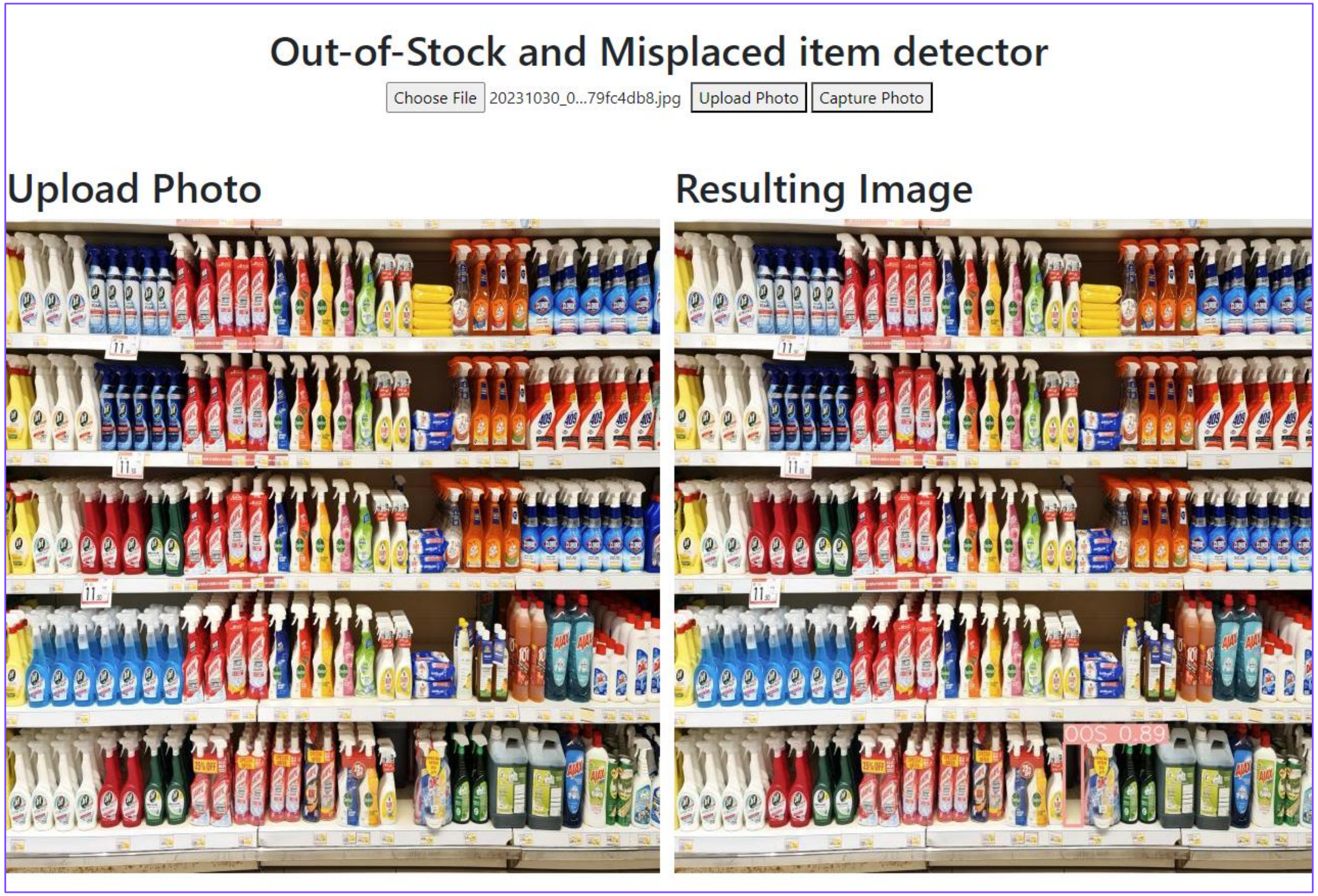


Figure 5.1: System high-level design with the preliminary implemented components boxed

Implemented Components

- **Web server**: Node.js + Express
- **Model server**: Python + Flask

Prototype


Out-of-Stock and Misplaced item detector
Choose File
20231030_0...79fc4db8.jpg
Upload Photo
Capture Photo
Upload Photo
Resulting Image
OOS 0.89

Conclusion

- Developed a Video-based Retail Store Shelf Monitoring System (SMS) using deep learning.
- Created dataset from Danube supermarket for Out-Of-Stock (OOS) and product misplacement (MIS).
- Achieved mean Average Precision (mAP) scores of 0.882 for MIS and 0.995 for OOS.
- Developed a preliminary prototype as a proof-of-concept for feasibility.
- Future work includes real-world testing, refining accuracy, scalability, exploring additional functionalities like shelf restocking

2023

THANK YOU

Mamoun Alghaslan, Khaled Almutairy