Computational Sciences and Engineering 5(2) (2025) 267-286

University of Guilan

Computational Sciences and Engineering

journal homepage: https://cse.guilan.ac.ir/

CoGY-Net: An Efficient AI-Powered Shelf OOS Detection System

Mohsen Eshghanmalek [a], Vali Derhami [b,*], Mohammad Ghasemzadeh[b]

[a] Ph.D. Candidate at Department of Computer Engineering, Faculty of Engineering, Yazd University, Yazd, Iran
[b] Professor of AI at Department of Computer Engineering, Faculty of Engineering, Yazd University, Yazd, Iran

ARTICLE INFO

Article history:
Received 27 December 2025
Received in revised form 11 February 2026
Accepted 13 February 2026
Available online 13 February 2026

Keywords:
Contourlet Transform
Golden Eagle Optimizer
YOLO Neural Network
Deep Learning
Out of Stock Detection

ABSTRACT

In modern retail management, there is a high demand for being able to efficiently identify empty shelves. To address the inherent limitations of current monitoring systems in handling dense arrangements and geometrically diverse products, To address this image processing problem, we propose CoGY-Net—a robust, intelligent Out-of-Stock (OOS) detection framework that takes RGB images of retail shelves as input. Our approach significantly enhances the standard YOLO architecture through two primary innovations. First, we integrate the Contourlet Transform as a geometric pre-processor to improve the extraction of curved product features within cluttered backgrounds, leveraging its superior directionality over traditional transforms. Second, we employ the Golden Eagle Optimizer (GEO), a metaheuristic algorithm, to eliminate the inefficiency of manual tuning by autonomously identifying the ideal training hyperparameters and anchor boxes tailored to the specific dataset. Furthermore, to ensure the system remains reliable across varying shelf depths, we implemented a scale-invariant dynamic gap analysis logic to pinpoint empty spaces accurately. In this way we manage to fill the research gap which was the lack of an automated and geometry-aware detection framework capable of handling dense shelf layout and curved products in real environments. The system was evaluated on the Out-Of-Stock-23 dataset. Experimental results demonstrate that CoGY-Net achieves an accuracy 90% and provides a high-precision, automated solution with superior stability, making it highly suitable for seamless integration into real-time smart retail environments and autonomous inventory systems.

1. Introduction

The application of artificial intelligence techniques, including computer vision, is of particular importance, and valuable research is currently being conducted in this field [1-3]. Out-of-Stock

* Corresponding author.
E-mail addresses: vderhami@yazd.ac.ir (V. Derhami)

https://doi.org/10.22124/cse.2026.32660.1150

(OOS) events on store shelves are among the most significant factors contributing to lost sales and decreased customer satisfaction in offline supermarkets. Frequent product shortages not only lead to immediate sales losses but also have significant negative long-term effects on brand reputation and store commercial performance by reducing customer loyalty. Therefore, accurate and timely identification of shelf inventory status and ensuring "On-Shelf Availability" (OSA) is considered a key performance indicator in the retail industry [4]. Furthermore, optimizing logistical processes, including supply, storage, arrangement, and transportation of goods, plays a vital role in increasing supply chain productivity and profitability [5].

Previous studies have shown that OOS events directly impact future customer purchasing behavior and can lead to brand switching or store abandonment [6]. in most stores, OSA monitoring is performed manually by staff, a method that is costly, time-consuming, and prone to human error. In recent years, automated systems based on image processing and machine vision have gained attention as efficient alternatives to manual inspections. In these systems, alerts for product replenishment are issued by analyzing shelf images and detecting the percentage of empty space.

Some studies have proposed using Radio Frequency Identification (RFID) tagging for inventory monitoring [7]; however, high implementation costs and the difficulty of integrating this technology with existing infrastructure limit its practical application [8]. Conversely, methods based on traditional image processing [9] and deep learning [10] have been developed, achieving high object detection accuracy by utilizing networks such as YOLO and Convolutional Neural Networks (CNNs). However, these approaches require extensive labeled training data, whereas public machine vision datasets usually do not include store product labels, and the manual annotation process is costly and time-consuming.

Implementing OSA detection systems in real-world store environments involves numerous technical challenges. These include high product density, severe object occlusion, changes in packaging appearance due to damage or irregular arrangement, and shelf obstruction by customers, staff, or store equipment. Additionally, conventional Background Subtraction methods [11–13], while widely used for detecting visual changes, are unable to distinguish between product removal and shelf replenishment. Consequently, they alone do not constitute a robust and reliable solution for precise OSA monitoring.

To overcome these limitations, this paper presents a novel hybrid architecture titled **Co**ntourlet-optimized **G**olden-eagle **Y**OLO **Net**work (CoGY-Net). By combining advanced image processing, deep learning, and computational intelligence, it provides a comprehensive solution for detecting OOS in store shelves.

The general idea of CoGY-Net is to establish a robust, three-stage hybrid framework that addresses the specific visual challenges of retail environments.

- **Geometric Enhancement:** First, the system overcomes the limitations of traditional image processing by applying the Contourlet Transform. Unlike Wavelets, this step preserves the smooth, curved boundaries of products in dense arrangements, effectively 'cleaning' the input for the neural network.
- **Automated Optimization:** Second, to remove the inefficiency of manual tuning, the Golden Eagle Optimizer (GEO) is employed. This metaheuristic algorithm automatically searches for

the optimal network configuration (Anchor Boxes and Hyperparameters) that best fits the specific store's merchandise.

- **Dynamic Logic:** Finally, instead of relying on fixed pixel thresholds, the system employs a Scale-Invariant Gap Analysis. This logic determines Out-of-Stock (OOS) events by comparing gaps relative to the average product width, making the detection stable against changes in camera distance or zoom.

The main contributions and innovations of this research are summarized as follows:

- **Overcoming Isotropic Limitations via Contourlet Transform:** Unlike traditional Wavelet-based approaches that fail to capture smooth curves due to their isotropic nature, we integrate the Anisotropic Contourlet Transform. This innovation allows for the precise extraction of directional edges in curved products (e.g., bottles and cans) within cluttered backgrounds, significantly reducing false negatives.
- **Automated Anchor Topology Optimization:** Addressing the inefficiency of manual hyperparameter tuning and generic K-Means clustering, we employ the Golden Eagle Optimizer (GEO). This metaheuristic approach autonomously explores the search space to generate adaptive anchor boxes that statistically match the diverse dimensions of the specific retail dataset.
- **Multi-Objective Density-Aware (MODA) Optimization:** We introduce the MODA-Fitness Function, a novel metric that goes beyond simple accuracy. By simultaneously optimizing for spatial precision (mean Average Precision (mAP)), small object sensitivity (Recall), and computational cost, this function guides the model to balance performance constraints in real-world scenarios.
- **Automated Data-Driven Search Space Profiling:** To prevent convergence issues common in metaheuristic algorithms, we present an Automated Data Profiler. This module analyzes dataset statistics to define intelligent search boundaries, ensuring the optimization process starts from a viable region rather than random initialization.
- **Robust Scale-Invariant Decision Logic:** Moving away from fixed pixel thresholds which fail under varying camera zooms, we develop a Dynamic Gap Analysis algorithm. This logic defines "Out-of-Stock" based on local product width ratios, ensuring consistent detection regardless of the camera's distance from the shelf.

The remainder of this paper is organized as follows: Section 2 provides a review of related research on OOS detection and object detection model optimization methods. Section 3 details the proposed CoGY-Net architecture and its technical specifics. Section 4 reports and analyzes the experimental results and performance evaluation of the proposed method. Finally, Section 5 is dedicated to the conclusion and suggestions for future research.

2. Related Work

Out-of-Stock (OOS) detection is a key issue in smart retail, playing a significant role in real-time inventory management and increasing customer satisfaction. In recent years, numerous solutions based on sensors, image processing, deep learning, and Internet of Thing (IoT) have been proposed for shelf status monitoring. However, many of these systems face challenges such as high hardware

installation costs, the need for precise sensor calibration, and low accuracy in determining actual product positions [14].

In early studies, classical image processing methods were used to detect the presence or absence of products on shelves. Moorthy et al. [9] utilized feature extraction and the Speeded Up Robust Features (SURF) algorithm to match product reference images with shelf images. Although this approach yielded acceptable results in controlled environments, its heavy reliance on reference images and high sensitivity to lighting changes, occlusion, and packaging variations limit its application in real-world store environments. A similar approach was proposed by [17] Kejriwal et al. [17] for direct product counting by mobile robots, requiring a large database of product templates [18]; maintaining and updating such a database in dynamic environments is impractical.

With the advent of deep learning, Convolutional Neural Networks (CNNs) have been widely adopted for shelf monitoring. As shown by Yilmazer et al. [7], a shelf auditing framework based on semi-supervised deep learning and YOLO v4 was presented, aiming to increase OSA in grocery stores. Also, Mohammad Noor et al. proposed a Fast Convolutional Neural Network (FCNN) model combined with global thresholding for identifying empty spaces [8]. Although these methods offer higher accuracy than classical methods, their dependence on fixed anchor boxes and manual hyperparameter tuning reduces their generalizability in dense environments.

In retail management, simultaneous monitoring of shelf status and Planogram compliance has garnered attention. Saqib et al. [15] presented the Hybrid Approach for Shelf Monitoring and Planogram Compliance (Hyb-SMPC) framework, a combination of machine vision and deep learning for product detection and planogram compliance checking. This study compared the performance of YOLO v4, YOLO v5, and YOLOR detectors, showing that single-stage detectors are more suitable for store applications in terms of speed and accuracy.

Video-based studies have also been used to monitor shelf changes. Higa et al. [10] used low-resolution Close-Circuit Television (CCTV) videos and employed a combination of background subtraction and CNN to detect product changes. An extended version of this research [16] analyzed customer behavior and high-traffic shelf areas by analyzing sequential images and using the Hungarian algorithm and networks based on CIFAR-10 and CaffeNet. However, the dependence of these methods on video quality and frame rate limits their performance in crowded, real-world environments.

Methods have also been proposed to identify empty areas without using a large database. Rosado et al. [19] used key points extracted by the FAST algorithm [20] to detect empty areas. This method suffers a severe drop in accuracy in the presence of environmental occlusions, such as customers standing in front of the shelf, providing an incorrect estimate of inventory status.

In recent years, newer research has focused on more advanced detectors. The use of improved YOLO versions like YOLO v7 and YOLO v8 for detecting small and dense objects in store environments has been reported [21–23]. Additionally, Transformer-based architectures, with the ability to model long-range dependencies, have been considered for analyzing complex retail scenes [24]. Despite these advancements, challenges such as poor representation of curved product edges, dependence on manual anchor box settings, and sensitivity to scale changes persist [25].

Recent advancements in object detection have shifted towards Anchor-free detectors and Transformer-based architectures. Anchor-free models, such as Fully Convolutional One-Stage

Object Detection (FCOS) [27] and Objects as Points Network (CenterNet) [28], eliminate the need for predefined anchor boxes, viewing object detection as a point-based regression problem. While this reduces hyperparameter complexity, these models often struggle with the extreme occlusion and dense packing typical of retail shelves, where object centers often overlap. Simultaneously, Transformer-based models like DEtection TRansformer (DETR) [29] and Shifted Window Transformer (Swin Transformer) [30] have utilized self-attention mechanisms to model global context, achieving state-of-the-art results in occlusion handling. However, their high computational cost (FLOPs) and slow inference speeds make them less viable for real-time edge computing applications in retail stores compared to CNN-based architectures. CoGY-Net addresses this trade-off by utilizing an optimized YOLO v11 architecture—leveraging its efficient single-stage nature—while employing GEO to autonomously fine-tune internal hyperparameters and anchor dimensions, effectively bridging the gap between the speed of CNNs and the adaptability of newer architectures.

Despite significant progress in using deep networks for shelf monitoring, most existing methods have one or more fundamental limitations: (1) Use of isotropic pre-processing methods like Wavelet, which are inefficient in representing curved product edges. (2) Dependence on fixed anchor boxes or local clustering, which are incompatible with the dimensional diversity of store goods. (3) Lack of a scale-independent OOS detection logic that is robust to changes in camera distance and angle.

These limitations reduce the accuracy and stability of existing systems in real and dense store environments, highlighting the necessity of developing a hybrid, geometry-aware, and optimized framework.

3. The Proposed Methodology

From an operational perspective, product unavailability on shelves reduces the potential impact of advertising and decreases actual demand. From a strategic perspective, OOS affects brand loyalty, promotes competitor brands, and reduces the effectiveness of sales team resources.

The proposed design presents a new architecture named CoGY-Net, consisting of five main modules, each responsible for a specific task in the processing chain. In the following, this architecture and its constituent modules are examined and described in detail.

3.1. CoGY-Net Architecture

A comprehensive overview of the CoGY-Net framework is illustrated in ***Figure 1***. The architecture follows a sequential pipeline where each module is strategically integrated to optimize the detection process, as detailed in the following subsections.

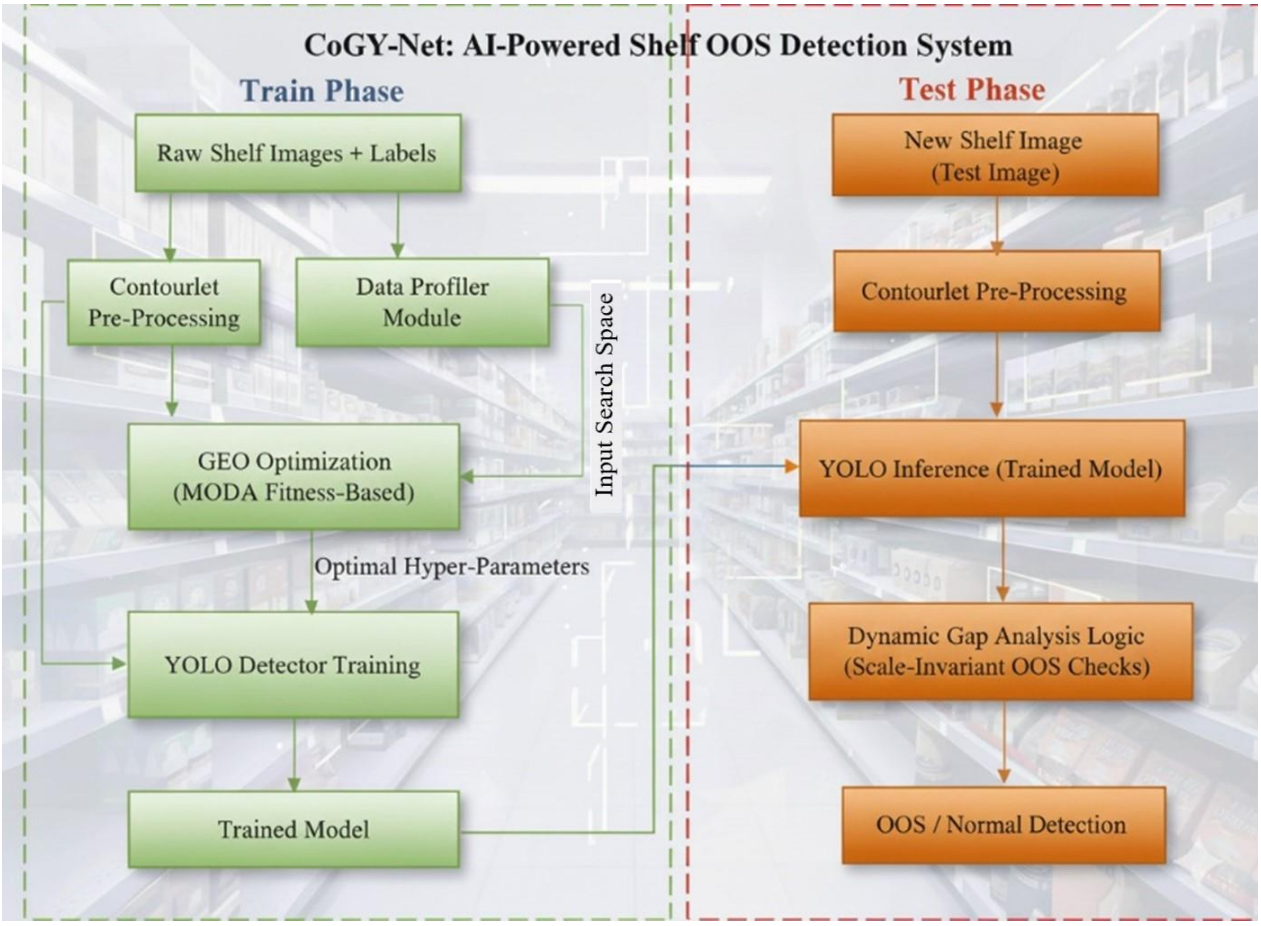

Figure. 1. Block diagram of CoGY-Net architecture in training and testing phases (Generated using Google Banana)

3.1.1. Geometric Pre-processing Module (Contourlet Transform)

In the scenario of smart shelf monitoring, input data quality plays a decisive role in the final performance of the object detection model. Images captured in retail environments usually have inherent visual challenges, including high arrangement density leading to product boundary interference, and noise caused by light reflection on plastic and glossy packaging. Common pre-processing methods, like the Wavelet transform, although successful in noise removal, have serious limitations in preserving the geometric structure of products. This section explains why and how the Contourlet transform is used as a superior alternative solution in the CoGY-Net architecture.

To extract image features in detection processes, two methods can be used: image-based and model-based. In the image-based method, features are extracted directly by applying mathematical transforms such as Fourier, Wavelet, Discrete Cosine, Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA) on images. The problem with these methods is the large dimensions, data redundancy, and sensitivity to image rotation and translation. In the model-based method, a model of the referenced image is built and described by a small set of parameters (e.g., active shape models, active contour models), offering the advantage of describing features in small dimensions and being invariant to image brightness, rotation, size, and translation.

The Wavelet transform views the world as consisting of points and squares and is limited to only three directions (horizontal, vertical, and diagonal) in edge detection. However, in store shelves,

most products like beverage bottles, cans, and chip bags have structures with curved and smooth edges. When trying to represent a curved line (like a bottle edge) with a tool that has a square nature (Wavelet), a large number of coefficients must be used, ultimately leading to jagged edges and loss of product boundary continuity. This discontinuity in edges causes errors in the neural network when separating two adjacent products.

The proposed framework uses the Contourlet transform to overcome this limitation. Unlike Wavelet, Contourlet possesses anisotropy and directionality. This means it can see edges in various directions and with different elongations. The function of Contourlet in this system can be likened to a two-stage process similar to the human visual system:

- **Point Detection Stage:** The image is first decomposed by a pyramidal structure to identify discontinuities and sudden changes (like noise and sharp edges). This stage is similar to Wavelet and captures fine image details.
- **Line Connection Stage:** The main innovation occurs here. The points identified in the previous stage enter a directional filter bank. These filters connect points located along a specific alignment.

The result is that scattered points turn into continuous lines and smooth curves. Therefore, Contourlet can reconstruct the spherical shape of a can or the curvature of a bottle with very high accuracy and without blurring. Using Contourlet in the proposed architecture provides three main advantages: intelligent noise removal, separability in high density, and reduced network processing load.

3.1.2. Data Profiler Module (Intelligent Search Space Extraction)

One of the main challenges in training object detection models is the dependence of final performance on manual and time-consuming hyperparameter tuning, often based on trial and error, which may lead to a suboptimal space selection. To overcome this problem and eliminate human intervention, the CoGY-Net architecture is equipped with an intelligent pre-analyzer module called the Data Profiler.

Before the training process begins, this module acts as a bridge between raw images and the optimization engine. It performs a statistical scan of dataset labels to extract the structural and geometric features of products. The key task of this section is to analyze the distribution of object dimensions and aspect ratios to understand product geometry (e.g., detecting the vertical structure of bottles or horizontal boxes), helping the Golden Eagle algorithm focus on realistic dimensional ranges. Furthermore, by assessing environmental density and calculating the average number of objects per image, the profiler dynamically adjusts data augmentation strategies to prevent excessive model complexity in dense environments. Finally, the output of these analyses is converted into a set of upper and lower bounds, serving as input to limit and target the search space of the Golden Eagle algorithm. This intelligent boundary determination prevents computational resource waste in irrelevant areas and guarantees that the optimization process starts with an informed initialization, directly resulting in increased convergence speed and improved final system accuracy.

3.1.3. Adaptive Optimization Module (Golden Eagle Optimizer)

In standard object detection architectures, the network's final performance heavily depends on the quality of initial anchor boxes. Common methods like K-Means clustering, due to sensitivity to outliers and a tendency to converge to local optima, are often unable to extract anchors that exactly match the diverse geometry of store products. To overcome this challenge, the proposed system utilizes a metaheuristic optimization engine based on the Golden Eagle Optimizer (GEO) algorithm. This module is tasked with intelligently searching the parameter space to extract the best possible configuration for the YOLO network.

The Golden Eagle algorithm is designed based on the intelligent hunting behavior of golden eagles. A distinctive feature of this algorithm is movement in spiral paths, allowing it to explore the search space with high efficiency. The engine's operation relies on balancing two key behaviors:

- **Cruising Behavior:** In the early stages of optimization, eagles tend to move in large circular paths to examine different areas of the search space. This mechanism ensures the system does not get trapped in local settings and tests diverse configurations of anchor box dimensions [18].
- **Attack Behavior:** As the process progresses and promising areas are found, the eagles' behavior changes, and their movement converges linearly and focused toward the best prey (the best settings found). This stage allows for fine-tuning parameters with very high precision [18].

In the proposed system, each "eagle" in the population represents a vector of YOLO network configuration settings. This vector includes precise anchor box dimensions (width and height for different scales) as well as vital training hyperparameters such as learning rate and optimizer momentum, which accelerate the training process.

3.1.3.1 Multi-Objective Density-Aware Fitness Function (MODA)

The most important component in any optimization system is the metric by which solution quality is measured. In conventional applications, usually only Mean Average Precision (mAP) is used. However, in shelf monitoring scenarios, merely high accuracy is not enough; the system must be able to see very small objects and have high processing speed. To this end, we designed a novel fitness function named MODA that evaluates the fitness of each eagle based on three simultaneous components:

- **Strict Spatial Accuracy:** Ensure predicted boxes have very precise overlap with real goods, vital for separating adjacent goods in dense shelves.
- **Sensitivity to Small Objects**: Given that many store items are small, this component rewards models that successfully detect small objects, preventing them from being ignored.
- **Computational Efficiency:** Acts as a penalty mechanism. If settings proposed by an eagle lead to a model exceeding hardware processing time limits, its fitness score is reduced.

In summary, the GEO algorithm uses this fitness function to guide the learning process toward a model that not only has the most accurate anchor boxes but also establishes an optimal balance between small object detection power and execution speed.

In the proposed architecture, each eagle (E_i) in the GEO algorithm population represents a complete answer vector for the YOLO configuration. This vector includes not only geometric anchor box

dimensions but also sensitive training hyperparameters. For a YOLO network with k anchor boxes at s different scales, the position vector of the i-th eagle at iteration t is defined by ***Eq. (1)***. In this equation, w_j and h_j are the normalized width and height for the j-th anchor box. Additionally, η is the initial learning rate controlling network convergence speed, and μ is the optimizer momentum accelerating the network learning speed.

$$\vec{X}_i^t = [w_1, h_1, \dots, w_N, h_N, \eta, \mu] \quad \textbf{(1)}$$

Fitness functions in prior works have two fundamental weaknesses for store applications: firstly, insufficient sensitivity to very small and dense objects, and secondly, ignoring the model's computational cost. To address these limitations, we designed the novel multi-objective MODA fitness function. The fitness function relationship is presented in ***Eq. (2)***. In this relationship, $\mathcal{A}(\vec{X})$ relates to strict spatial accuracy. $\mathcal{R}_{small}(\vec{X})$ is an innovative metric where the algorithm, by maximizing it, imposes a heavy penalty on eagles whose anchor boxes cannot cover small goods (size less than 32x32 pixels). The parameter $\Psi(\vec{X})$ is a penalty metric used to control computational cost. Finally, the output determines weight coefficients ω_1, ω_2, and ω_3 ensuring the final optimal vector $\vec{X}_{opt}$ results in a model that is accurate, efficient in dense environments, and respects real-time constraints.

$$F_{MODA}(\vec{X}) = \omega_1 . \mathcal{A}(\vec{X}) + \omega_2 . \mathcal{R}_{small}(\vec{X}) + \omega_3 . \Psi(\vec{X}) \quad \textbf{(2)}$$

3.1.4. Object Detection and Occlusion Removal Module (YOLO Network)

According to the CoGY-Net block diagram (***Figure 1)***, the final stage of the training phase involves the YOLO neural network, tasked with high-level feature extraction, precise product localization, and removing occlusion caused by people in front of shelves. This architecture was chosen for its single-stage nature, establishing an optimal balance between detection accuracy and real-time processing speed. Our proposed system utilizes YOLO v11, not in its standard form, but as an inference engine fed by Contourlet inputs and tuned with GEO parameters.

Figure 2 illustrates the YOLO v11 network architecture. The YOLO structure in CoGY-Net consists of three main parts:

- **Backbone:** Since the input here consists of Contourlet pre-processed images (noise removed, edges highlighted), initial layers converge faster and can focus on extracting semantic patterns (like brand or packaging type) [26].
- **Neck:** This feature pyramid structure combines deep layer semantic information with shallow layer spatial information. This is vital for detecting small objects (like small cans or chocolates) common in store shelves [26].
- **Head:** Modern YOLO uses a decoupled head structure; the classification branch is separated from the box regression branch. This separation allows the network to distinguish product type more accurately while adjusting the bounding box with pixel precision [26].

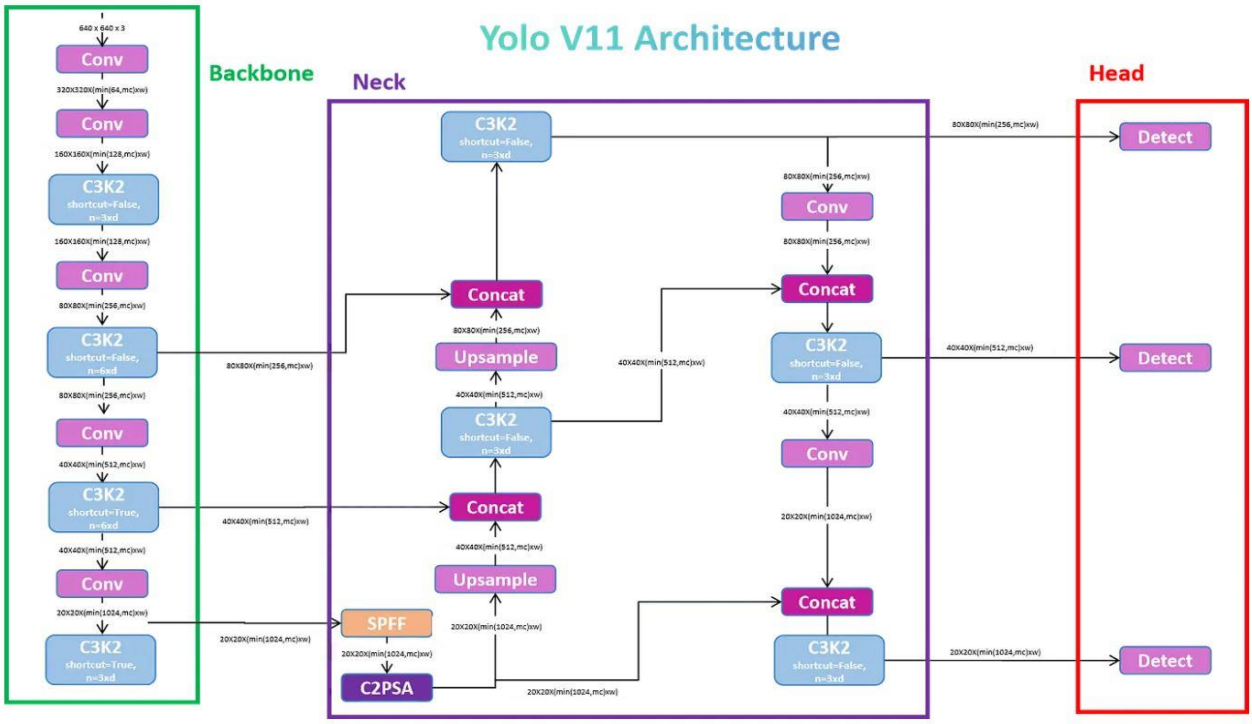

Figure. 2. YOLO v11 network architecture with a backbone for feature extraction, a feature pyramid neck for multi-scale fusion, and a decoupled head for accurate detection [26]

The distinction of CoGY-Net lies in its configuration. Normally, YOLO uses default anchor boxes and fixed parameters. In the proposed architecture:

- **Adaptive Anchors**: Dimensions and geometric ratios for initial box prediction are received directly from the optimal GEO output. The network knows beforehand to look for rectangles with specific dimensions (e.g., bottle dimensions in that store).
- **Training Hyperparameters:** Learning rate and momentum are also determined by GEO to prevent getting stock in local minima.

The output is a set of high-precision bounding boxes sent to the OOS detection logic unit.

3.1.5. OOS Detection Logic Subsystem

The YOLO output is a set of bounding boxes for existing products. However, OOS detection essentially means identifying the absence of objects, not their presence. Common methods either focus on training a separate class for empty space (highly dependent on shelf background texture) or use fixed pixel thresholds for distance measurement. The challenge with the latter is camera distance variation; a 50-pixel distance might indicate a small space in zoom but a large hole in a wide view. To overcome this, CoGY-Net uses a novel post-processing algorithm called the Scale-Independent Gap Analyzer. The process includes:

- **Spatial Organization:** Detected boxes are separated into distinct rows based on vertical coordinates and then sorted horizontally (left to right).
- **Dynamic Thresholding:** The empty space detection threshold $T_{dynamic}$ (***Eq. (3)***) is calculated as a function of the average width of existing products in the same image, rather than a fixed number. In this equation, λ is the system sensitivity coefficient (usually $\lambda \approx 0.8$), and $\bar{W}_{local}$ is the average width of products in that image.

$$T_{dynamic} = \lambda . \overline{W}_{local} \quad \textbf{(3)}$$

- **Decision Making:** The distance between every pair of neighboring products is calculated and compared with the dynamic threshold. Since both parameters (gap distance and product width) change proportionally with camera zoom, their ratio remains constant, making OOS detection stable without needing recalibration.

3.2. Multi-Level Feature Representation in CoGY-Net

To provide a definition of the feature spaces utilized in CoGY-Net, we formulate the feature vectors for each processing stage. The system transforms data through a cascade of five distinct feature domains:

3.2.1. Directional Frequency Coefficients (Contourlet Module)

The input image I is decomposed into J scales and 2^l directions. The feature vector at this stage consists of the coefficients representing edge intensity and directionality as defined in ***Eq. (4)***.

$$V_{CT} = \{C_{j,k}^{(l)}\} \quad \textbf{(4)}$$

Where C denotes the Contourlet coefficient at scale j, direction l, and spatial location k. This vector encodes the "curvature" and "orientation" of product boundaries.

3.2.2. Statistical Dataset Descriptor (Data Profiler Module)

This module views the dataset as a statistical distribution. The feature vector is a compact descriptor defining the geometric bounds as defined in ***Eq. (5)***.

$$V_{stat} = [W_{min}, W_{max}, H_{min}, H_{max}, \rho_{density}] \quad \textbf{(5)}$$

Where W and H are the 1st and 99th percentiles of object dimensions, and ρ is the average object density per image. This vector is used to constrain the search space (Ω).

3.2.3. Configuration Vector (GEO Optimizer Module)

In the evolutionary context, each "eagle" represents a solution vector encoding the network topology and training dynamics as defined in ***Eq. (6)***.

$$V_{GEO} = [w_{a1}, h_{a1}, \dots, w_{aN}, h_{aN}, \eta, \mu] \quad \textbf{(6)}$$

Here, w and h are the optimized anchor dimensions, η is the learning rate, and μ is the momentum. This vector adapts the network architecture to the specific dataset geometry defined by V_{stat}.

3.2.4. Deep Semantic Feature Tensors (YOLO Network Module)

The core detection engine extracts hierarchical deep features from the backbone. These are not 1D vectors but 3D tensors at three Feature Pyramid Network (FPN) scales as defined in ***Eq. (7)***.

$$V_{Deep} = \{f_{P3}, f_{P4}, f_{P5}\} \quad \textbf{(7)}$$

Where $f_{Pi} \in \mathbb{R}^{H_i \times W_i \times C_i}$ represents the feature map at stride i. These tensors encode semantic class information (e.g., "bottle", "can") robust to occlusion.

3.2.5. Spatial Geometry Vector (OOS Logic Module)

For the final Out of Stock (OOS) decision, we abstract away from pixels to spatial relationships. The feature vector is a sequence of gap distances between sorted objects in a row as defined in ***Eq.(8)***.

$$V_{Gap} = \{d_1, d_2, \dots, d_{n-1}\} \text{ and } d_i = x_{start}^{(i+1)} - x_{end}^{(i)} \quad \textbf{(8)}$$

Where d_i is the distance between object i and $i+1$. This vector is compared against a dynamic threshold ($T_{dynamic}$) as defined in ***Eq. (3)*** to classify the shelf status (0 for Normal, 1 for OOS).

4. Experiments and Results

In this section, to evaluate the proposed system for shelf OOS detection, the **Out-of-Stock-23 dataset**[*], containing images captured by cameras stationed on store shelves, is used.

4.1. Dataset

The Out-of-Stock-23 dataset is specifically collected and curated for challenges of "dense object detection" and "On-Shelf Availability" (OSA) monitoring. Images include various store products such as beverage bottles, cans, cartons, and chip bags arranged on shelves with varying fill levels (from fully stocked to semi-empty). The dataset contains 4470 color images (RGB), all labeled in standard YOLO format. Data was split into three subsets: Training (70%), Validation (20%), and Testing (10%) to prevent overfitting (***Table 1***). Analysis of object size distribution shows that a percentage of goods have an area less than 32x32 pixels, categorizing them as "small objects" and justifying the use of the MODA fitness function.

Table 1. Dataset Information

Dataset Split	Num of Images	Small	Medium	Large
Training	3129	1.3 %	57.1 %	41.6 %
Validation	894	1.1 %	45.6 %	53.3 %
Test	447	1.2 %	34.3 %	64.5 %
Total number of images in the dataset: 4,470				
Total number of products in the dataset: 23,775				
Average aspect ratio (height/width): 1.14				

This dataset is distinct due to environmental challenges: high density and overlapping products (highlighting the need for Contourlet), geometric diversity and curved surfaces, light reflections causing noise, and scale variations due to camera distance.

[*] https://universe.roboflow.com/thesis-dpgss/out-of-stock-tnjj1/dataset/23

Figure. 3. Sample images from the dataset

Figure 3 presents representative sample images illustrating key challenges in real-world retail environments, including high product density (products touching each other), occlusion (customers partially blocking shelves), small objects, and variations in camera distance. These visual examples highlight the limitations of conventional methods and justify the necessity of employing the Contourlet transform in the proposed approach. Following the Contourlet transformation, a distance-independent logical system is introduced to enable accurate out-of-stock detection across varying camera-to-shelf distances.

4.2. Evaluation and Results

In this section, the performance of CoGY-Net was analyzed from three aspects: (1) Efficiency of the GEO optimization algorithm, (2) Statistical accuracy compared to the state-of-the-art methods, and (3) Qualitative and visual analysis. Implementation was done using Python in Google Colab with an NVIDIA Tesla T4 GPU.

4.2.1. Convergence Analysis of the Optimization Fitness Function

Figure 4 shows the comparative convergence graph of the MODA fitness function versus the general GEO fitness function. The graph shows a stable upward trend. In initial iterations (exploration phase), fluctuations occur, but over time (exploitation phase), the algorithm converges to the best configuration. This demonstrates GEO's ability to find anchor boxes with high geometric correspondence to the dataset.

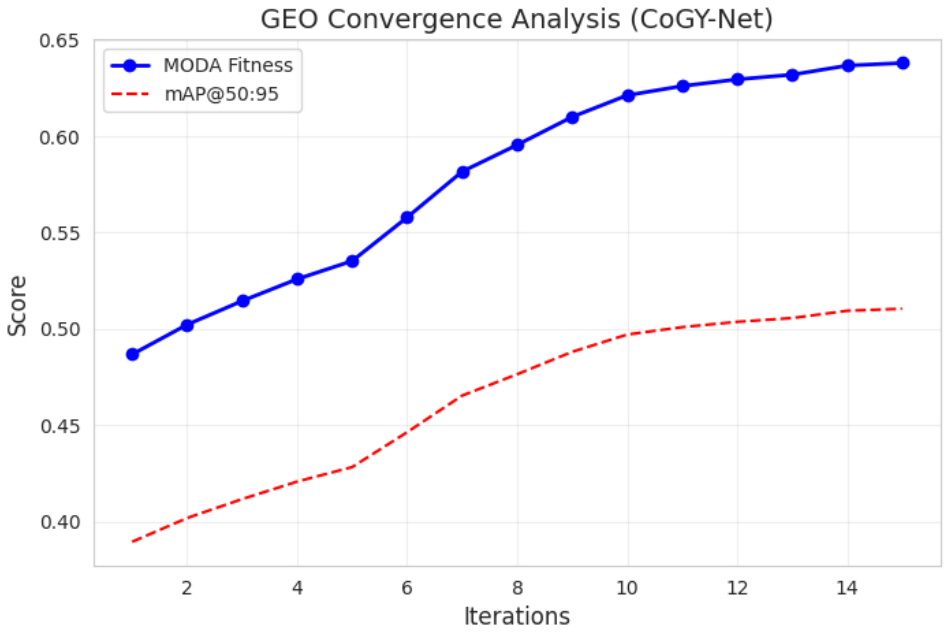

Figure. 4. GEO convergence analysis in CoGY-Net, illustrating the progressive improvement of the proposed MODA fitness compared to mAP@50–95, demonstrating fast convergence and stable optimization behavior.

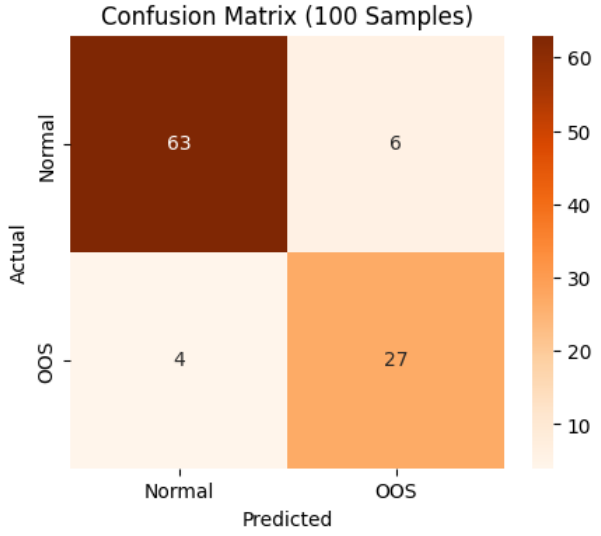

Figure. 5. Confusion matrix of CoGY-Net evaluated on 100 randomly selected test samples, illustrating the classification performance for normal and OOS shelf conditions.

4.2.2. Classification Performance Evaluation

A confusion matrix was calculated on 100 random images (***Figure 4***). Two classes were considered: Normal (full or reduced space) and OOS (critical empty space). CoGY-Net successfully detected the vast majority of OOS samples. The low False Negative rate indicates the system rarely mistakes an empty shelf for a full one, which is vital for preventing sales loss.

4.2.3. Comparative Analysis

To evaluate the effectiveness of the proposed system in detecting OOS conditions on retail shelves, the method was compared with four representative object detection–based approaches. The performance of each method was assessed using Accuracy, Precision, Recall, and F-measure, and the corresponding results are summarized in ***Table 2***.

Table 2. Performance comparison of the proposed algorithm with the state-of-the-art methods

Methods	Accuracy	Precision	Recall	F-measure
Faster R-CNN	0.80	0.78	0.83	0.80
YOLO v8	0.82	0.80	0.77	0.78
YOLO v11	0.85	0.84	0.82	0.83
Wavelet + YOLO	0.87	**0.87**	0.85	**0.86**
CoGY-Net (ours)	**0.90**	0.82	**0.87**	0.84
The standard deviation for all reported metrics ranges between ±0.01 and ±0.03.				

Table 2 presents the performance of the proposed CoGY-Net method in comparison with several representative object detection–based approaches for the OOS detection task. All results are reported as the mean and standard deviation over 10 independent runs conducted on 100 randomly test samples, allowing both accuracy and robustness to be jointly evaluated. CoGY-Net achieved the highest overall accuracy and a good balance between precision and recall. Low standard deviation indicates high stability.

Table 3. Performance comparison of CoGY-Net with baseline and hybrid YOLO-based methods for OOS detection

Methods	Accuracy	Precision	Recall	F-measure
YOLO v11 (Y)	0.85	0.84	0.82	0.83
Y+ Contourlet (C)	0.87	0.83	0.85	0.84
Y+C+ GEO	0.88	0.84	0.86	0.85
Y+C+ GEO (+MODA)	0.89	**0.85**	0.86	**0.86**
CoGY-Net (ours)	**0.90**	0.82	**0.87**	0.84
The standard deviation for all reported metrics ranges between ±0.01 and ±0.03.				

Table 3 presents a comparative performance evaluation of the proposed CoGY-Net and several baseline and hybrid methods. The results indicate that the original YOLO v11 model provides a solid baseline; however, its performance is constrained in complex retail environments. By integrating the Contourlet transform with YOLO v11, an improvement in recall is observed, demonstrating enhanced sensitivity to fine-grained structures and small objects. The addition of the GEO module further improves both accuracy and precision by incorporating geometric consistency into the detection process. When the MODA fitness function is employed, a more balanced optimization between precision and recall is achieved. Despite these incremental improvements, CoGY-Net consistently outperforms all comparative methods, achieving the highest overall accuracy and recall. This performance gain highlights the effectiveness of the proposed end-to-end framework in handling challenges such as high product density, occlusion, scale variation, and varying camera-to-shelf distances, which are critical for reliable out-of-stock detection.

4.2.4. Qualitative and Visual Results

This section also presents representative examples of the results obtained by applying the CoGY-Net system to images that were not seen during the training phase. ***Figures 6***, ***7***, and ***8*** illustrate normal shelf conditions, where both high product density and occlusion are present. As shown, the system successfully identifies empty spaces while correctly reporting a normal shelf status. In ***Figure 9***, the system issues an OOS alert in response to a clearly empty shelf region. ***Figure 10*** presents another example in which the system accurately detects empty or depleted areas and correctly triggers an OOS warning in accordance with the shelf condition. ***Figures 11*** and ***12*** depict cases in which the system fails to correctly identify the shelf status. (In all figures 6-12: Left to right: 1) Raw input image, 2) Contourlet edges, 3) Pre-processed input image, 4) System Output.)

Figure. 6. System output for a normal shelf example, showing accurate detection of empty or reduced spaces and correct identification of the overall shelf status as normal.

Figure. 7. System output for a normal shelf with high product density, demonstrating accurate identification of reduced spaces and correct classification of the overall shelf condition as normal.

Figure. 8. System performance under occlusion: Accurate detection of reduced spaces and normal shelf status despite partial human obstruction.

Figure. 9. System output for an empty shelf (OOS) example, demonstrating accurate detection of shelf emptiness and correct triggering of the OOS alert.

Figure. 10. System output for a partially empty shelf (OOS) example, showing near-accurate identification of empty and reduced spaces and correct classification of the overall shelf status as OOS.

As shown in ***Figure 10***, the shelf image contains multiple empty and reduced regions. The proposed system is able to identify these gaps with high accuracy and, due to the extensive level of shelf space reduction, correctly issues an OOS alert indicating the absence of products. Notably, the detection results demonstrate that variations in illumination intensity have minimal impact on system performance, highlighting the robustness of the proposed approach against lighting changes.

Figure. 11. Output error of the system for an OOS shelf sample, where despite an incorrect final shelf classification, empty spaces are partially detected.

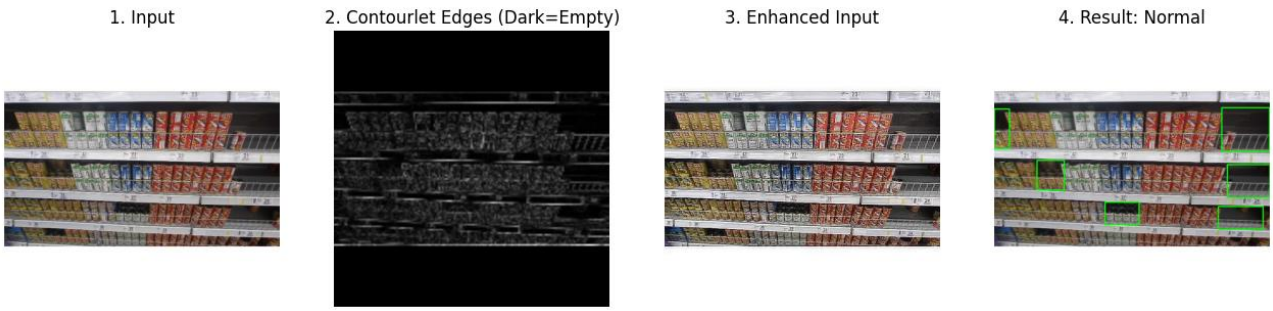

Figure. 12. System error example: Despite the accurate detection of empty and reduced spaces, the final output incorrectly classifies the shelf status as Normal.

In ***Figures 11*** and ***12***, although the system incorrectly classifies the shelf status as normal, it is still able to identify the empty or reduced spaces with a high degree of accuracy.

Overall, based on the statistical, analytical, and visual results presented in this section, the capability of the CoGY-Net system in detecting product unavailability on retail shelves is clearly demonstrated. Furthermore, the system's accuracy in identifying empty spaces among densely packed products confirms the effectiveness of the proposed architecture in real-world retail environments.

5. Conclusion

The development of CoGY-Net demonstrates that integrating advanced geometric pre-processing with metaheuristic optimization significantly enhances Out-of-Stock (OOS) detection in complex retail environments. By leveraging the Contourlet Transform, the system effectively isolated the curved boundaries of dense products, surpassing the capabilities of traditional isotropic transforms. Furthermore, the implementation of the Golden Eagle Optimizer (GEO) successfully automated the selection of anchor boxes and hyperparameters, resulting in an overall accuracy of 90% and superior recall for small items. This framework remains robust across varying camera perspectives due to its scale-invariant dynamic gap analysis logic. From a managerial perspective, CoGY-Net provides an automated, low-cost solution to maximize on-shelf availability and prevent sales loss, while technically establishing a precedent for self-tuning AI systems that eliminate manual configuration in industrial vision tasks.

Future research will focus on enhancing the preprocessing module through adaptive and learnable transforms, while also investigating anchor-free architectures to improve detection flexibility. Additionally, the system will transition from binary OOS detection to multi-level shelf availability assessment, incorporating temporal modeling and semi-supervised learning to predict shortages and further reduce data annotation costs. Finally it is worth to mention that the proposed algorithm can perhaps be improved to be able to recognize objects in the shelves when the empty space is behind of another object.

References

[1] Arash, S., Jamal, G., & Seyyed Ali, Z. (2025). Advanced Deep Learning Approaches for Accurate and Efficient Suspicious Behavior Detection in Surveillance Videos. *Computational Sciences and Engineering*, 4(2), 203-216.

[2] Christos, M., Eleni, V., & George, A. P. (2025). Annotation tools for computer vision tasks. *Seventeenth International Conference on Machine Vision (ICMV 2024), Edinburg, United Kingdom.*

[3] Patel, S., & Niani, R. (2025). Advanced Computer Vision-Based Retail Analytics: A Novel Approach to Customer Behavior and Product Movement Tracking. *Proceedings, IISE Annual Conference,* 1-6.

[4] García-Arca, J., Prado-Prado, J. C., & González-Portela Garrido, A. T. (2020). On-shelf availability and logistics rationalization. A participative methodology for supply chain improvement. *Journal of Retailing and Consumer Services, Elsevier,* 52(C).

[5] Lin, M. H., Sarwar, M. A., Daraghmi, Y. A., & İk, T. U. (2022). On-shelf load cell calibration for positioning and weighing assisted by activity detection: Smart store scenario. *IEEE Sensors Journal, 22*(4), 3455–3463.

[6] Araya Lobos, M. (2021). Detecting out of shelf supermarket products. *Master's thesis, Universitat Politècnica de Catalunya*.

[7] Yilmazer, R., & Birant, D. (2021). Shelf auditing based on image classification using semi-supervised deep learning to increase on-shelf availability in grocery stores. *Sensors*, 21(2).

[8] Mohammad Noor, M. Ch., Usha, K. A., Sampath, D. M. B., Omkar, J. D., & Saikumar, K. (2020). Implementation of online and offline product selection system using FCNN deep learning: Product analysis. *Materials Today Proceedings*, 45.

[9] Moorthy, R., Behera, S., Verma, S., Bhargave, S., & Ramanathan, P. (2015). Applying image processing for detecting on-shelf availability and product positioning in retail stores. *In Proceedings of the ACM International Conference Proceeding Ser, Kochi, India*, 451–457.

[10] Higa, K., & Iwamoto, K. (2018). Robust estimation of product amount on store shelves from a surveillance camera for improving on-shelf availability. *In Proceedings of the IST 2018—IEEE International Conference Imaging Systems and Techniques, Kraków, Poland*, 1–6.

[11] Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. *In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK*, 2, 28–31.

[12] KaewTraKulPong, P., & Bowden, R. (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. *In Proceeding Video-Based Surveillance Systems, Boston*, 135–144.

[13] Godbehere, A. B., Matsukawa, A., & Goldberg, K. (2012). Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation. *In Proceedings of the American Control Conference, Montréal, QC, Canada*, 4305–4312.

[14] Andaur, J. M. R., Ruz, G. A., & Goycoolea, M. (2021). Predicting out-of-stock using machine learning: An application in a retail packaged foods manufacturing company. *Electronics*, 10(22), 2787.

[15] Saqlain, M., Rubab, S., Khan, M. M., Ali, N., & Ali, S. (2022). Hybrid approach for shelf monitoring and planogram compliance (HYB-SMPC) in retails using deep learning and computer vision. *Mathematical Problems in Engineering*.

[16] Higa, K., & Iwamoto, K. (2019). Robust shelf monitoring using supervised learning for improving on-shelf availability in retail stores. *Sensors*, 19, 2722.

[17] Kejriwal, N., Garg, S., & Kumar, S. (2015). Product counting using images with application to robot-based retail stock assessment. *In Proceedings of the 7th Annual IEEE International Conference on Technologies for Practical Robot Applications, Woburn, MA, USA*, 1–6.

[18] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110, 346–359.

[19] Rosado, L., Gonçalves, J., Costa, J., Ribeiro, D., & Soares, F. (2016). Supervised learning for out-of-stock detection in panoramas of retail shelves. *In Proceedings of the IEEE International Conference on Imaging Systems and Techniques, Chania, Greece*, 406–411.

[20] Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 105–119.

[21] Mohammadi-Balani, A., Nayeri, M. D., Azar, A., & Taghizadeh-Yazdi, M. (2021). Golden eagle optimizer: A nature-inspired metaheuristic algorithm. *Computers & Industrial Engineering*, 152, 107050.

[22] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.

[23] Wang, C. Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Vancouver, Canada*, (pp. 746–754).

[24] Jocher, G., Chaurasia, A., Qiu, J., & Ultralytics Team. (2023). YOLOv8: Ultralytics next-generation object detection and segmentation. *Ultralytics Technical Report*.

[25] Zhang, X., Liu, H., Wang, J., & Chen, Y. (2023). Dense object detection in retail shelves using improved YOLO architectures. *Expert Systems with Applications*, 213, 119091.

[26] Rao, S. N. (2024). YOLOv11 architecture explained: Next-level object detection with enhanced speed and accuracy. *Medium. Available online (accessed on 25 February 2025)*.

[27] Tian, Z., Shen, C., Chen, H., & He, T. (2019). Fully convolutional one-stage object detection (FCOS). *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9627–9636.

[28] Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. *arXiv preprint, arXiv:1904.07850*.

[29] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. *Proceedings of the European Conference on Computer Vision (ECCV)*, 213–229.

[30] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10012–10022.