engineering proceedings

MDPI

Proceeding Paper

Hybrid Decision Framework for Reliable Sashimi Detection and Size Estimation †

Ari Aharari [1,*] and Takeru Ishimoto [2]

[1] Faculty of Computer and Information Sciences, SOJO University, Kumamoto 860-0082, Japan
[2] Department of Computer and Information Sciences, SOJO University, Kumamoto 860-0082, Japan
* Correspondence: info@ahrary.org
† Presented at the 6th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB 2026), New Taipei, Taiwan, 24–26 April 2026.

Abstract

Vision-based inspection is increasingly applied in food processing to improve measurement accuracy and reduce manual workload. In frozen sashimi production, however, overlapping placement of products frequently degrades the reliability of conventional detection methods based on object centroids. We developed a compact visual inspection framework that combines You Only Look Once version 11 with oriented bounding boxes and a hybrid decision strategy based on center distance and intersection over union. Oriented detection enables rotation-invariant localization of trays and sashimi, whereas tray-based geometric rectification converts image measurements into real-world dimensions. The system was evaluated on continuous image sequences collected in a refrigerated processing environment, demonstrating that the hybrid logic mitigates misclassification caused by overlap more effectively than single-metric approaches. Experimental results on sequential image data collected in a refrigerated processing environment showed that the compact visual inspection framework mitigated misclassification caused by overlap more effectively than single-metric approaches under the tested conditions, supporting its applicability to real-time industrial inspection scenarios.

Keywords: AI-based inspection; sashimi size detection; oriented object detection; IoU–center hybrid decision logic; YOLO

1. Introduction

Automation has become a central driver of efficiency and reliability in modern food manufacturing, where production systems must simultaneously satisfy strict sanitation standards, dimensional consistency, and high processing throughput. Within seafood processing, accurate control of product size is particularly important, as irregular portioning directly affects downstream packaging, pricing, and regulatory compliance. Recent volatility in marine resources, driven by environmental change, temperature shifts, and biological stressors, has further increased pressure on processing facilities to operate with minimal waste and maximum reliability, thereby accelerating the adoption of automated inspection technologies.

Frozen sashimi processing exemplifies this challenge. After cutting, sashimi slices are rapidly arranged by hand onto trays under refrigerated conditions to maintain freshness and safety. Although skilled workers perform this task efficiently, manual placement inherently results in variation in object orientation, spacing, and relative position. Overlapping and stacked arrangements are common, especially as operators attempt to optimize tray

Check for updates

Academic Editors: Teen-Hang Meen, Shu-Han Liao and Cheng-Fu Yang

Published: 17 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

utilization. Under such conditions, purely manual visual inspection becomes vulnerable to inconsistency and fatigue, motivating the integration of vision-based systems capable of objective and repeatable assessment.

Advances in computer vision and deep learning have enabled practical noncontact inspection systems suitable for industrial environments. In particular, real-time object detection networks derived from the You Only Look Once (YOLO) architecture have demonstrated strong performance in production settings due to their balance of speed and accuracy. Prior studies have shown that deep-learning-based detectors can be combined with geometric calibration techniques to estimate physical dimensions of food products directly from image data. However, many existing systems rely on axis-aligned bounding boxes and simple positional heuristics, assumptions that break down when objects appear at arbitrary orientations or when placement patterns vary dynamically.

One of the most persistent technical difficulties in such inspection pipelines is determining whether a detected object represents a newly introduced item or one that has already been processed. A widely used approach compares the positions of detected object centers across successive images, classifying an item as new if its centroid shifts beyond a predefined distance threshold. While effective in scenes where objects remain well separated, this strategy is fragile in the presence of partial occlusion. Even slight overlap can alter the detected shape of an object and cause its estimated center to drift, resulting in erroneous identification of existing items as new and leading to compounding errors in counting and size estimation.

To improve robustness under rotational variance, oriented bounding box representations have emerged as an extension of conventional object detection. By explicitly modeling object orientation, oriented bounding boxes provide a tighter geometric fit for elongated or irregularly shaped objects, improving localization accuracy and reducing background inclusion. YOLOv11, a recent evolution of the YOLO family, incorporates oriented bounding box prediction while retaining real-time inference capability, making it well suited for industrial inspection scenarios such as sashimi processing.

Nevertheless, orientation-aware detection alone does not resolve the challenge of distinguishing newly placed items when overlap is involved. Objects that overlap spatially may exhibit significant changes in centroid position without corresponding physical novelty. Intersection over Union (IoU), which measures the spatial overlap between detected regions, offers complementary information by indicating whether two detections likely correspond to the same physical object. Although IoU is widely used in detection evaluation and tracking, its role as a decision cue for incremental object placement in food processing lines has received limited attention.

In response to these challenges, we developed a compact visual inspection framework for sashimi inspection that integrates oriented object detection with a hybrid decision strategy. The framework employs YOLOv11 with oriented bounding boxes to detect trays and sashimi slices under arbitrary orientations. Tray geometry is then leveraged to perform perspective rectification, enabling conversion of pixel-based measurements into real-world dimensions. To identify newly placed sashimi, the proposed method jointly evaluates centroid displacement and bounding-box overlap, classifying an item as new if either spatial separation or insufficient overlap indicates novelty. This formulation is designed to combine the strengths of center-based and overlap-based reasoning while mitigating their individual weaknesses.

Practical deployment considerations are embedded throughout the system design. A lightweight graphical user interface provides real-time visualization of detection results and allows interactive adjustment of detection confidence, center-distance thresholds, and

https://doi.org/10.3390/engproc2026159047

IoU thresholds. This flexibility is essential in operational environments, where camera placement, illumination, and product appearance may vary across facilities or production lines.

The developed framework is lightweight and robust, tailored to incremental product placement in industrial food processing environments. Whereas conventional multi-object tracking approaches preserve object identity under continuous motion, the developed framework addresses a distinct operational scenario in which objects are sequentially added to an otherwise static scene. A hybrid geometric decision mechanism integrates centroid displacement and bounding-box overlap into a unified novelty judgment, enabling reliable discrimination of newly introduced items under partial overlap and stacking. In addition, the framework combines orientation-aware detection with tray-based geometric normalization to achieve consistent real-world size estimation independent of object pose or camera perspective. The framework can be deployed in real-time inspection workflows with minimal computational overhead.

2. Related Works

Automated visual inspection has gained increasing momentum in industrial manufacturing as a means of improving consistency, throughput, and hygiene, reducing reliance on manual assessment. In food processing environments, where direct contact measurements are undesirable, vision-based systems offer a practical alternative for evaluating product presence, size, and conformity. Early inspection approaches were built on rule-based image processing and handcrafted descriptors. Although sufficient for controlled laboratory conditions, these methods exhibited limited robustness in environments subject to lighting variation, background changes, and irregular object appearance, characteristics that are common in real production lines.

The introduction of deep learning has fundamentally reshaped visual inspection research. Convolutional neural networks have enabled inspection systems to learn discriminative features directly from data, resulting in improved adaptability and generalization. In the food industry, deep-learning-based vision systems have been successfully applied to tasks such as product counting, size estimation, and inventory control. Recent research has demonstrated that AI-driven inspection can support both quality assurance and logistical decision-making under operational constraints, particularly in food-related applications where manual monitoring is costly and error-prone [1–3].

Among contemporary detection frameworks, single-stage architectures based on the YOLO paradigm have been widely adopted due to their fast inference and compact design. YOLO demonstrated that object localization and classification could be unified into a single forward pass, enabling real-time performance suitable for industrial deployment [4]. Subsequent generations of YOLO improved detection stability and accuracy, leading to successful application in food inspection scenarios that require rapid response and continuous monitoring. These properties have made YOLO-based detectors a common backbone for AI-driven inspection and counting systems in food processing and inventory management [2].

Despite their advantages, most YOLO-based applications in food inspection rely on axis-aligned bounding boxes, which impose geometric constraints that do not align well with real-world object variability. Food products such as fish fillets or sashimi slices are frequently placed at arbitrary angles, causing axis-aligned boxes to include unnecessary background regions and distort size measurements. To address orientation-related inaccuracies, oriented bounding box detection has emerged as an effective extension of conventional detection. By explicitly encoding object rotation, oriented representations offer tighter spatial coverage and improved geometric fidelity, particularly for elongated or irregular objects [5–7]. Recent studies, including Oriented R-CNN [7], have demonstrated that orientation-aware detection can improve object localization while maintaining computational efficiency. Recent research

has shown that such orientation-aware detection can be incorporated into single-stage detectors without sacrificing real-time performance.

Beyond detection accuracy, identifying newly introduced objects across sequential observations remains a persistent difficulty in automated inspection. A prevalent solution relies on comparing object centroids across frames and classifying an object as new if its center shifts beyond a predefined distance threshold. This approach is attractive for its simplicity and low computational cost; however, it implicitly assumes that object geometry remains stable between observations. In practice, partial occlusion or overlapping placement often alters detected shapes, causing centroid displacement even for unchanged objects. Similar limitations have been reported in the multi-object tracking literature, where geometry-based association can become unreliable under occlusion [8,9]. Recent tracking approaches, such as ByteTrack [9], improve object association by incorporating low-confidence detections, thereby reducing missed associations in challenging visual conditions.

Overlap-based reasoning offers an alternative perspective for object association. IoU is widely used to quantify spatial correspondence between detected regions and plays a central role in detection evaluation, suppression of duplicate detections, and tracking-by-detection pipelines [10]. IoU-based matching is robust when objects share substantial spatial overlap, even if their detected centers fluctuate. However, overlap metrics alone are insufficient in scenarios involving vertical stacking or incremental placement, where newly added items may exhibit limited planar overlap with existing objects despite being physically distinct.

Importantly, inspection tasks involving incremental product placement differ fundamentally from conventional multi-object tracking. In food processing workflows, objects are typically introduced sequentially into a largely static scene, exhibiting minimal independent motion and limited appearance diversity. Conventional multi-object tracking methods, such as BoT-SORT [11], employ motion and appearance information to maintain object identities across successive frames. However, incremental food placement differs from conventional tracking because objects remain largely stationary while new items are introduced sequentially. Under these conditions, complex appearance-based or motion-based association strategies may offer limited additional benefit while increasing computational overhead.

Prior work on AI-based size detection and visual counting in the food industry has emphasized the importance of lightweight, task-specific decision logic tailored to operational constraints rather than adopting generic tracking frameworks wholesale [1,2].

Consequently, recent research has shifted toward reframing new-object identification as a geometric decision problem rather than a full identity-preserving tracking task. By focusing on spatial cues that directly relate to placement behavior, such as displacement and overlap, inspection systems can achieve greater robustness while maintaining real-time performance. Nevertheless, the combined use of center-distance and overlap-based criteria for industrial inspection remains insufficiently explored, particularly in food processing environments where overlapping and stacking are frequent. This gap is addressed in the present study, which integrates complementary geometric cues into a hybrid decision framework optimized for incremental placement inspection.

3. Framework Architecture

The developed framework enables dependable sashimi detection, precise dimensional measurement, and consistent recognition of newly introduced items under practical factory conditions. These environments are characterized by uncontrolled object orientation, frequent partial overlap, and unavoidable visual disturbances. To cope with these challenges, the framework combines orientation-aware object detection, tray-based spatial normalization, and a lightweight decision strategy specifically designed for incremental placement scenarios.

The inspection workflow reflects the actual processing sequence used at the collaborating frozen seafood manufacturing facility and integrates conventional manual operations with automated AI-driven quality assessment (Figure 1). Fresh fish are delivered to the facility by refrigerated transport and undergo manual preprocessing, including cleaning, trimming, and slicing. The resulting sashimi portions are then arranged on standardized trays before handling and packaging.

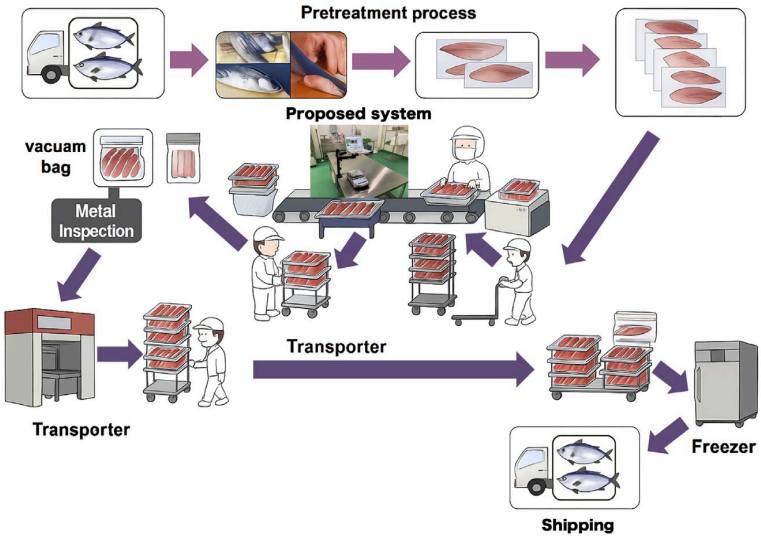


Figure 1. Workflow in frozen seafood processing facility.

A fixed overhead camera is installed above the tray placement area to capture image sequences during the loading process. Each image is analyzed in temporal order using a structured processing pipeline. The pipeline consists of tray localization, image rectification to correct perspective distortion, sashimi detection using oriented object representations, conversion of image measurements into real-world dimensions, and identification of newly placed items. Information from previously analyzed images is retained and reused throughout the sequence, enabling cumulative inspection rather than independent frame-based evaluation. All components are implemented using general-purpose computing resources and are designed to operate in real time without dependency on specialized hardware accelerators. This design choice ensures that the system can be deployed flexibly within existing processing lines while maintaining compatibility with industrial operating constraints.

3.1. Oriented Detection and Geometric Rectification

Tray localization constitutes the basis of both coordinate normalization and physical scale estimation within the proposed inspection pipeline. To achieve robust tray identification under rotational variance, a YOLOv11 detector configured for oriented bounding box (OBB) output is applied. The tray is described as a four-vertex polygon defined by corner coordinates:

$$p_i = (x_i, y_i), i = 1, \ldots, 4 \quad (1)$$

Because the detection order of vertices may vary depending on orientation, the extracted points are reordered into a fixed canonical layout corresponding to the upper-left, upper-right, lower-right, and lower-left tray corners. This ordering step guarantees consistent geometric processing regardless of the tray's pose in the camera view. Using the reordered tray vertices, the system performs perspective correction to compensate for camera inclination and projection effects. Equations (2) and (3) represent the corresponding points in the rectified coordinate space. These points are related through a planar homography [Equation (4)]. Planar geometric transformations and camera calibration techniques have

been extensively investigated in computer vision, with Zhang [12] establishing a widely adopted calibration framework based on planar reference patterns. In the proposed system, the known geometry of the sashimi tray is utilized to perform perspective rectification and enable consistent dimensional measurements.

$$\mathrm{x} = (x, y, 1)^{\mathrm{T}} \tag{2}$$

$$\mathrm{x}' = (x', y', 1)^{\mathrm{T}} \tag{3}$$

$$\mathrm{x}' \sim \mathrm{Hx} \tag{4}$$

Here, $\mathrm{H} \in \mathbb{R}^{3\times3}$ is the homography matrix. The matrix H is estimated from point correspondences between the detected tray corners and predefined reference coordinates derived from the known physical geometry of the tray. Applying this transformation produces a geometrically normalized, top-down representation of the tray, allowing spatial consistency across successive frames.

Physical calibration is achieved by exploiting the known dimensions of the tray. Let L_w and L_h denote the actual tray width and height in millimeters, and let P_w and P_h represent the corresponding edge lengths measured in pixels from the rectified image. The pixel-to-millimeter scaling factors are then computed as:

$$s_w = \frac{L_w}{P_w}, s_h = \frac{L_h}{P_h} \tag{5}$$

These scale coefficients are subsequently applied to all object measurements, enabling conversion from image coordinates to real-world dimensions. Following tray normalization, sashimi detection is performed on the rectified image using a second YOLOv11 model trained specifically for sashimi classes and configured for OBB prediction. Each detected sashimi is represented as an oriented quadrilateral with vertices:

$$q_j = \left(x_j, y_j\right),\ j = 1, \ldots, 4 \tag{6}$$

The spatial center of each sashimi is calculated as the centroid of its bounding box:

$$\mathbf{c} = \frac{1}{4}\sum_{j=1}^{4} q_j \tag{7}$$

Geometric attributes are derived by computing the average lengths of opposing sides of the oriented bounding box. These pixel-based values are then converted into real-world width, height, and projected area using the previously determined scaling factors $\boldsymbol{S_w}$ and $\boldsymbol{S_h}$. This procedure ensures consistent and orientation-invariant size estimation across all detected sashimi pieces.

The oriented object detection models were trained using annotated image data collected from the target processing environment. Bounding-box annotations were provided in oriented quadrilateral format to support rotation-aware detection. Training was performed offline using standard stochastic optimization procedures, and inference was executed on general-purpose computing hardware without reliance on specialized accelerators. This design choice was intended to ensure compatibility with typical factory computing infrastructure while maintaining real-time operation. As the focus of this study lies in incremental decision logic rather than detector optimization, detailed benchmarking of detection accuracy is deferred to future work.

3.2. Hybrid Decision Strategy for Identifying Newly Placed Items

To reliably identify sashimi pieces introduced over time, the developed framework maintains a cumulative memory of previously detected objects. Each registered sashimi instance is stored as an oriented bounding box together with its associated center coordinate

in the rectified tray coordinate system. When processing a new image, each newly detected sashimi candidate is compared against the stored set to determine whether it represents a newly placed item or an already registered one.

Two complementary geometric similarities are employed for this comparison: spatial separation based on centroid distance and geometric correspondence based on region overlap. These measures are evaluated independently and then combined through a unified hybrid decision rule designed specifically for incremental placement scenarios. Spatial separation is quantified using a centroid-based distance metric. Let $\boldsymbol{c}$ denote the center coordinate of a current detection and $\boldsymbol{c_k}$ the center of a previously registered sashimi. The Euclidean distance between the two is computed as

$$d(\mathbf{c}, \mathbf{c}_k) = \| \mathbf{c} - \mathbf{c}_k \|_2 \tag{8}$$

This measure is effective when newly placed sashimi are positioned in distinct regions of the tray, such as lateral placement adjacent to existing items. To account for variations in camera resolution and perspective, centroid distances are evaluated in millimeter units after geometric rectification, ensuring spatial consistency across image sequences. However, centroid-based reasoning alone is insufficient under partial overlap or stacked placement, where deformation or occlusion may cause significant centroid displacement even for unchanged objects. To address this limitation, geometric overlap between detections is also evaluated using IoU. Let A and B_k denote the polygonal regions corresponding to the oriented bounding boxes of the current detection and a previously registered sashimi, respectively. Their overlap is defined as:

$$\mathrm{IoU}(A, B_k) = \frac{\mathrm{Area}(A \cap B_k)}{\mathrm{Area}(A \cup B_k)}. \tag{9}$$

IoU provides a robust indication of spatial correspondence when detections originate from the same physical object, even if their centroids shift due to shape changes or occlusion. Conversely, newly placed sashimi typically exhibit limited planar overlap with existing items, particularly in adjacent placement configurations. Final classification of a detection as newly placed is determined through a hybrid decision rule that jointly considers both geometric cues. Specifically, a sashimi is classified as new if

$$\min_k d(\mathbf{c}, \mathbf{c}_k) > \tau_d \mathrm{or} \max_k \mathrm{IoU}(A, B_k) < \tau_{\mathrm{IoU}} \tag{10}$$

Here, τ_d and τ_{IoU} denote predefined thresholds for center distance and overlap, respectively. The use of a logical OR condition reflects the complementary strengths of the two measures: center-distance evaluation is sensitive to spatial separation, while IoU-based reasoning provides robustness under overlap and stacked placement. Requiring both conditions to be satisfied simultaneously would risk suppressing valid novelty detection in common production scenarios.

Threshold values were selected empirically using a small validation subset independent of the final evaluation sequence. The center-distance threshold τ_d was chosen to be larger than the centroid fluctuations observed for unchanged sashimi under deformation and occlusion, yet smaller than the typical separation between distinct placements. The IoU threshold τ_{IoU} was set conservatively to preserve overlap-based matching for the same physical object while allowing newly introduced sashimi with limited overlap to be correctly classified as new.

Importantly, the hybrid formulation reduces sensitivity to exact threshold tuning. During experimental evaluation, stable classification behavior was observed across a broad operating range for both parameters. This robustness is essential for industrial deployment, where camera alignment, lighting conditions, and placement habits may vary. To further support practical use, both thresholds can be adjusted at runtime through the graphical

interface, allowing operators to perform rapid on-site calibration without modifying the underlying detection or decision logic.

By integrating complementary geometric cues within a lightweight decision framework, the proposed hybrid strategy enables reliable discrimination of newly placed sashimi under diverse placement patterns while preserving computational simplicity suitable for real-time inspection. Figure 2 shows a diagram of the developed framework.

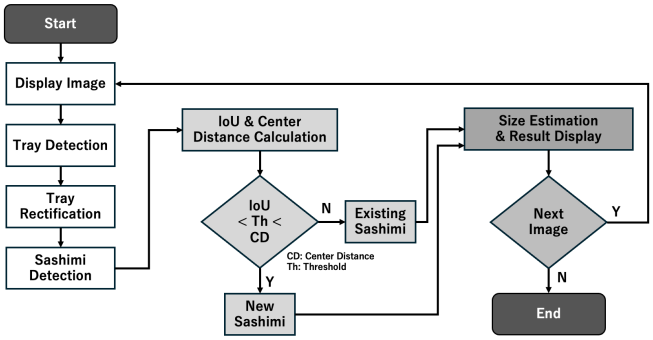


Figure 2. A diagram of the proposed system.

3.3. Implementation and Graphical User Interaction

To enable seamless integration into real production environments, the developed framework is realized through a compact and operator-oriented graphical user interface. The interface presents live visualization of analysis results, displaying oriented bounding boxes, object labels, and estimated dimensional information directly over the normalized tray image. This real-time feedback allows users to immediately verify detection outcomes and system behavior during operation. The interface also provides direct user control over critical processing parameters, including the detection confidence level, spatial separation threshold, and overlap threshold based on IoU. Through slider-based and interactive controls, operators can adapt system behavior to site-specific conditions such as camera alignment, illumination changes, or variation in product appearance. This capability is particularly important in industrial settings, where inspection conditions can differ across facilities or production lines.

In addition to routine operation, the user interface supports exploratory parameter adjustment for performance evaluation and tuning. By allowing rapid modification of thresholds and immediate visual feedback, the interface facilitates sensitivity analysis and aids in identifying stable operating regions that balance precision and robustness. Such functionality reduces the need for repeated code-level adjustments and supports efficient on-site calibration. By decoupling the core inspection algorithms from the interaction layer, the system preserves methodological transparency while delivering a practical and accessible tool for industrial users. This separation ensures that the underlying decision logic remains extensible and maintainable, while the interface focuses on usability and adaptability in real-world deployment scenarios.

4. Experimental Result

4.1. Dataset and Experiment

Performance evaluation was conducted using image data collected from an operational frozen sashimi processing facility to assess the practical behavior of the proposed inspection framework under realistic industrial conditions. The objective of the evaluation was to verify the feasibility of reliable new-item identification during incremental tray loading, particularly in the presence of partial overlap and stacked placement.

The dataset comprised a single continuous sequence of 17 frames, captured by a fixed overhead camera during routine production. Across the sequence, 17 sashimi placement events occurred, with exactly one new sashimi introduced between consecutive frames. As the placement sequence advanced, the number of sashimi pieces arranged on the tray increased step by step, mirroring the workflow typical of industrial freeze-processing lines. The images were analyzed in chronological order, with previously detected sashimi instances retained in memory to support cumulative inspection.

Oriented bounding box annotations were created for both tray boundaries and individual sashimi pieces. Each sashimi instance was annotated as a four-vertex quadrilateral representing its oriented footprint on the tray surface. Annotations were produced using a custom annotation tool supporting oriented polygon labeling. To ensure annotation reliability, all labels were independently reviewed by a second researcher familiar with the inspection task and processing environment. Discrepancies, particularly in frames involving partial overlap or stacked placement, were resolved through consensus review before training and evaluation. The evaluation is intended as a proof-of-concept validation, emphasizing the comparative behavior of decision strategies rather than exhaustive statistical benchmarking.

4.2. Performance Evaluation

Framework performance was evaluated based on frame-level identification of newly placed sashimi pieces. For each frame, a result was considered correct only when exactly one sashimi, the newly introduced item, was classified as new, and all previously registered sashimi were correctly identified as existing. Any false positive or false negative within a frame was counted as a misclassification. In this study, three decision strategies were evaluated: (1) center-distance-only logic, (2) IoU-only logic, and (3) the proposed hybrid decision logic integrating both metrics. Precision, recall, and F1-score were computed at the frame level for direct comparison.

Table 1 shows the evaluation results across the 17-frame sequence. The center-distance-only approach correctly identified newly placed sashimi in 11 frames, yielding precision and recall values of 0.65. Most errors occurred under partial overlap, where deformation of existing sashimi caused centroid displacement beyond the predefined threshold. The IoU-only strategy improved robustness in layered placement scenarios, correctly identifying new items in 13 frames with precision and recall of 0.76; however, it failed in stacked configurations where newly introduced sashimi exhibited limited planar overlap.

Table 1. Comparison of new-item identification performance.

Decision Strategy	Precision	Recall	F1-Score
Center-distance-only	0.65	0.65	0.65
IoU-only	0.76	0.76	0.76
Proposed hybrid logic	1.0	1.0	1.0

The proposed hybrid decision logic correctly identified newly placed sashimi in all 17 frames, resulting in precision, recall, and F1-score values of 1.0 for the evaluated sequence. By classifying a sashimi as new when either sufficient spatial separation or insufficient overlap was observed, the hybrid method effectively mitigated the complementary failure modes of the single-metric approaches.

In the reported experiments, the center-distance threshold τ_d was set to 25 mm, selected to exceed the centroid fluctuations observed for unchanged sashimi under deformation and partial occlusion while remaining smaller than typical separations between distinct placements. The IoU threshold τ_{IoU} was set to 0.30, a conservative value chosen to preserve

overlap-based correspondence for detections originating from the same physical object while allowing newly introduced sashimi with limited planar overlap to be correctly classified as new. Threshold values were determined empirically using a small validation subset independent of the final evaluation sequence.

Sensitivity analysis was performed by varying τ_d between 20–35 mm and τ_{IoU} between 0.25–0.40. Stable classification behavior was observed across these ranges for the evaluated sequence, with no change in new-item identification outcomes. This indicates that the proposed hybrid decision logic is not critically dependent on precise threshold tuning, an important property for practical deployment in industrial environments.

In addition to new-item identification accuracy, qualitative assessment of the estimated dimensional measurements demonstrated stable behavior across varying orientations and placement configurations. The combination of oriented bounding box detection with tray-based geometric rectification prevented measurement drift caused by camera perspective or object rotation. An example of correct new-item identification using the proposed hybrid logic is shown in Figure 3. Although the results are limited in scope, they confirm the feasibility of the proposed approach and justify further investigation using larger datasets and diverse production conditions.

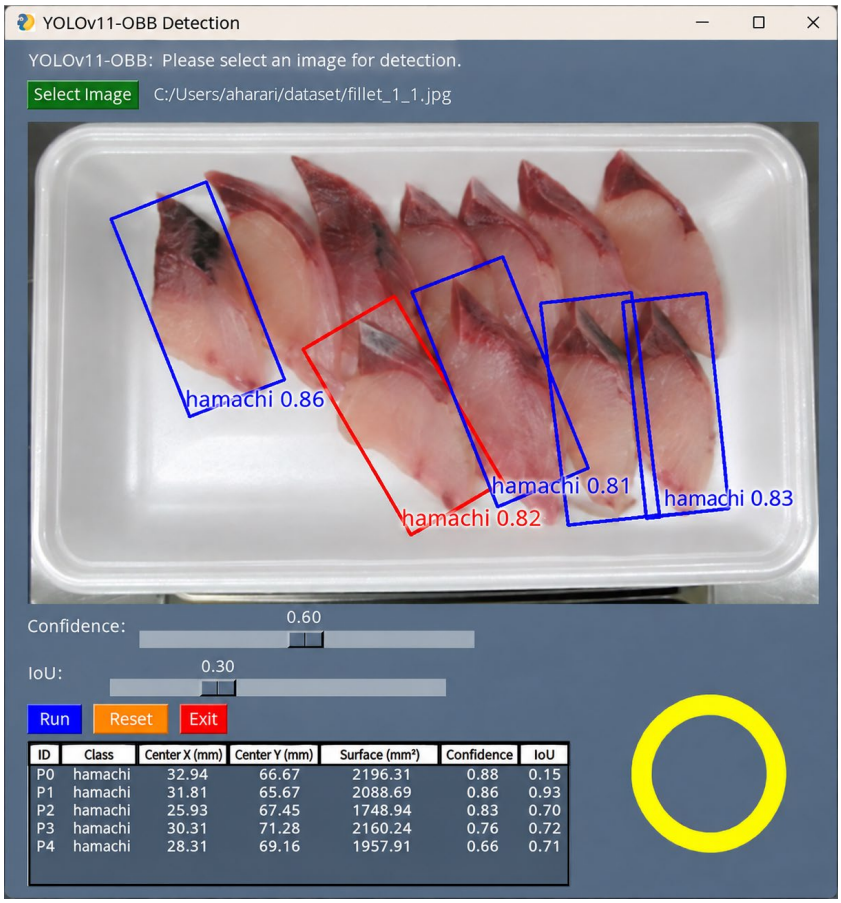


Figure 3. A sample of correctly identified newly placed sashimi with proposed hybrid decision logic. The blue bounding boxes indicate previously detected sashimi pieces, while the red bounding box highlights the newly placed sashimi identified by the proposed method.

5. Discussion

The experimental results show that the effectiveness of new-item identification in incremental inspection tasks depends strongly on the geometric cues employed. Whereas the proposed hybrid decision framework achieved perfect frame-level classification for

the evaluated image sequence, the significance of this outcome lies not in the absolute performance values but in the comparative behavior observed across decision strategies under realistic placement conditions. The findings confirm that combining complementary geometric measures can substantially improve robustness in scenarios involving overlap and stacking, which frequently undermine conventional inspection approaches.

The center-distance-only strategy, despite its simplicity and low computational cost, exhibited inherent limitations when sashimi pieces overlapped. Partial occlusion or deformation altered detected object shapes, leading to centroid displacement even when the physical object remained unchanged. As observed in the experimental sequence, this behavior resulted in erroneous classification of existing items as newly placed, highlighting a fundamental weakness of centroid-based reasoning in environments where overlap is common rather than exceptional. These observations are consistent with prior findings in the geometry-based association and tracking literature, which reports reduced reliability of centroid metrics under occlusion.

The IoU-only strategy addressed some of these shortcomings by focusing on spatial overlap between detections. This approach increased robustness in layered placement cases by correctly associating detections with substantial overlap. However, its reliance on planar overlap proved insufficient in stacked placement scenarios, where newly introduced sashimi pieces often exhibited limited overlap with existing items despite being physically distinct. Consequently, IoU-only logic failed to consistently identify novelty across all placement configurations, illustrating that overlap information alone does not fully capture the spatial relationships relevant to incremental placement inspection.

The hybrid decision framework developed in this study resolves these complementary failure modes. By integrating spatial separation and overlap analysis through a logical OR formulation, the framework preserves sensitivity to both lateral placement and overlap-induced deformation. The observed elimination of misclassification in the evaluated sequence indicates that reliable new-item discrimination in incremental inspection tasks requires the joint interpretation of multiple geometric cues rather than reliance on a single metric. Importantly, this robustness was ensured without introducing complex appearance modeling or temporal tracking mechanisms, maintaining computational simplicity suitable for real-time industrial deployment.

Beyond classification accuracy, the stability of physical size estimation is a critical consideration in industrial inspection applications. Qualitative assessment of the proposed framework demonstrated consistent dimensional measurements across varying orientations and placement configurations. The use of oriented bounding box detection combined with tray-based geometric rectification effectively decoupled size estimation from camera perspective and object pose. This geometric consistency is particularly important in food processing environments, where accurate size control directly influences packaging decisions, pricing, and regulatory compliance.

From an operational perspective, the robustness of the hybrid framework carries practical significance. Incremental placement processes are vulnerable to error accumulation, as a single misclassification can propagate through subsequent frames and compromise cumulative inspection results. The resistance of the proposed decision logic to overlap-induced errors represents a meaningful advantage for continuous production environments, where reliability and interpretability are often prioritized over algorithmic complexity.

Several limitations of the framework must be acknowledged. The evaluation experiment was conducted using a single image sequence captured under a fixed camera configuration in one processing facility. Although the dataset reflects realistic industrial conditions, broader validation is necessary to assess generalization across different lighting setups, tray geometries, camera viewpoints, product types, and operator behaviors. In addition,

the current implementation relies on manually defined thresholds. Although sensitivity analysis demonstrated stable behavior across a practical operating range, adaptive or data-driven threshold selection must be used to enhance robustness under changing environmental conditions.

Finally, the framework processes images sequentially without exploiting temporal smoothing or motion constraints. Incorporating lightweight temporal consistency mechanisms may further reduce sensitivity to transient detection noise or brief occlusions while preserving real-time performance. Despite these limitations, the present study results provide a proof of concept and demonstrate that task-specific, geometry-driven decision logic can effectively address incremental placement challenges that are insufficiently handled by conventional tracking approaches.

6. Conclusions

A vision-based inspection framework was developed to support reliable sashimi detection and physical size estimation in frozen seafood processing environments. The objective was the accurate identification of newly placed sashimi during incremental tray loading, a scenario in which partial overlap and stacked placement frequently undermine conventional inspection and tracking approaches. To address this challenge, the framework integrates orientation-aware object detection with tray-based geometric normalization and a lightweight hybrid decision strategy combining spatial separation and region overlap analysis.

Evaluation results show the inherent limitations of single-metric decision strategies. Approaches based exclusively on center displacement exhibited pronounced sensitivity to geometric distortions induced by overlap, whereas IoU-based strategies failed to consistently detect novelty in stacked placement scenarios with restricted planar overlap. By contrast, the framework effectively mitigated these complementary weaknesses through the joint consideration of multiple geometric cues, thereby enabling stable identification of newly placed items across the evaluated sequence. These results can be used for proof-of-concept validation rather than conclusive evidence of general robustness. Nonetheless, the results underscore that reliable new-item discrimination in incremental inspection cannot be achieved through reliance on a single geometric criterion. Instead, coordinated interpretation of multiple spatial relationships is essential, particularly in industrial contexts characterized by irregular placement and partial occlusion. Notably, this robustness was attained without recourse to complex appearance modeling or full multi-object tracking, thereby preserving computational efficiency compatible with real-time deployment on standard factory hardware.

Beyond classification reliability, qualitative assessment results revealed that the integration of oriented bounding box detection with tray-based geometric rectification enabled consistent physical size measurements across diverse orientations and placement configurations. Such geometric consistency is critical for practical inspection tasks in food processing, where dimensional accuracy directly informs packaging decisions and quality control.

Author Contributions: Conceptualization, A.A.; methodology, T.I. and A.A.; software, A.A.; validation, T.I. and A.A.; formal analysis, T.I. and A.A.; data curation, T.I. and A.A.; writing—original draft preparation, T.I. and A.A.; writing—review and editing, A.A.; visualization, T.I. and A.A.; supervision, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The secondary data that were used for analysis in this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aharari, A.; Tanaka, S. Development of artificial intelligence-based product size detection system. Engineering Proceedings. In Proceedings of the IEEE International Conference on Engineering Innovation and Business (ICEIB 2025), New Taipei, Taiwan, 25–27 April 2025; Volume 108.
2. Aharari, A.; Kuwaduru, K.; Mehdipour, F. Development of an artificial intelligence (AI)-based visual counting system for the food industry. In *Proceedings of the 13th IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE 2023)*; IEEE: Piscataway, NJ, USA, 2023; pp. 136–139.
3. Ghonimy, M.; Barakat, H. Artificial intelligence and computer vision for automated quality inspection of processed foods: A comprehensive review. *Foods* **2026**, *15*, 4064. [CrossRef]
4. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*; IEEE: Piscataway, NJ, USA, 2016; pp. 779–788.
5. Yang, X.; Yan, J.; Zhang, Z.; He, L. R3Det: Refined single-stage detector with feature refinement for rotating object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 3163–3171.
6. Ding, Y.; Lu, Y.; Huang, J.; Ding, Z. Learning rotated bounding boxes for multi-oriented object detection. *IEEE Trans. Image Process.* **2021**, *30*, 4333–4346.
7. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021), Montreal, QC, Canada, 10–17 October 2021; pp. 3520–3529.
8. Bewley, S.; Ge, A.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2016)*; IEEE: Piscataway, NJ, USA, 2016; pp. 3464–3468.
9. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-object tracking by associating every detection box. In *Computer Vision—ECCV 2022*; Springer: Cham, Switzerland, 2022; pp. 1–21. [CrossRef]
10. Wojke, W.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2017)*; IEEE: Piscataway, NJ, USA, 2017; pp. 3645–3649.
11. Aharon, N.; Orfaig, R.; Bobrovsky, B.-Z. BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv* **2022**, arXiv:2206.14651. [CrossRef]
12. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.