NUEJ
Naresuan University
Engineering Journal

Benchmarking ROI Strategies for Real-Time Chicken Counting Using YOLOv8 and LLM-Assisted Development in Industrial Slaughterhouses

Panu Buranajarukorn[1], Somsakul Thongtab[1,*], Jiraporn Virunbut[1], Itsariyaporn Luanghan[3] and Phisut Apichayakul[2]

[1] Department of Industrial Engineering, Faculty of Engineering, Naresuan University, Phitsanulok, Thailand
[2] Department of Electrical and Computer Engineering, Naresuan University, Phitsanulok, Thailand
[3] Faculty of Industrial Technology, Uttaradit Rajabhat University, Uttaradit, Thailand

* Corresponding author e-mail: somsakult65@nu.ac.th

(Received: 10 November 2025, Revised: 15 December 2025, Accepted: 30 December 2025)

Abstract

Accurate chicken counting is essential for operational efficiency and compliance in poultry processing. Manual counting methods are prone to error and unsuitable for high-speed production. This study presents the validation of automated chicken counting in an industrial slaughterhouse using YOLOv8 detection with SORT tracking and ROI-based strategies. While the core pipeline follows established computer vision methods, the novelty lies in systematically benchmarking three ROI strategies under high-speed conveyor conditions where occlusion, motion blur, and unstable lighting are major challenges. Tested on real production line footage, the system was evaluated using precision, recall, and F1-score against ground truth counts. Video-based strategies centred on the conveyor line achieved the highest accuracy, with F1-scores up to 0.998 and a Mean Absolute Error (MAE) of 2.30, a Mean Absolute Percentage Error (MAPE) of 0.74%, and a Root Mean Square Error (RMSE) of 2.70, while image-based approaches undercounted by up to 13%. Confidence variability was markedly lower in video-based methods (CV < 9%), demonstrating robustness under dynamic production conditions. Beyond methodological integration, this work introduces LLM-driven code generation for rapid development of industrial vision systems. The findings provide practical guidance for camera positioning, threshold settings, and deployment in high-speed slaughterhouse environments, establishing a foundation for scalable, high-accuracy poultry processing automation.

Keywords: Automated chicken counting, Computer vision, Poultry industry, Video-based tracking, YOLOv8

1. Introduction

Thailand is currently positioned among the top five global exporters of chicken meat. In 2023, the nation's chicken exports were valued at approximately USD 1.54 billion (The Observatory of Economic Complexity, 2025), with more than 57% comprising cooked or processed products (Public Relations Department of Thailand, 2024). This upward trend is projected to continue at an annual growth rate of 3.5–4.5%, driven by increasing global demand for affordable protein sources, expanded access to halal markets in the Middle East, and rising demand from neighboring countries. Additionally, the avian influenza outbreak in Brazil in 2025 presents Thailand with a strategic opportunity to further expand its export market, with projected revenues reaching USD 1.7 billion (Reuters, 2025).

Concurrently, the poultry industry is undergoing a technological transformation through the adoption of Precision Livestock Farming (PLF) technologies. These systems utilize artificial intelligence (AI), the Internet of Things (IoT), and computer vision to enhance productivity and animal welfare monitoring (Jiang et al., 2023). PLF facilitates improved decision-making in areas such as animal health, feed efficiency, and traceability (Novus International, Inc., 2025). Within this framework, automated chicken counting has emerged as a critical component in slaughterhouse operations, contributing to consistency, traceability, and compliance with international export standards.

Despite its operational importance, traditional chicken counting methods such as manual tallying and contact-based sensors exhibit several limitations: human error (Wu et al., 2025) due to fatigue and subjective judgment, occlusion (Khanal et al., 2024; Wu et al., 2025) from overlapping carcasses, environmental instability (Feng et al., 2025) affecting detection accuracy, visual complexity (Khanal et al., 2024) from background elements, and scalability (Wu et al., 2025) issues in high-speed processing environments.

Recent advancements in deep learning-based object detection have demonstrated significant potential in addressing these challenges. Okinda et al. (2020) emphasized the potential of deep learning systems across multiple poultry welfare tasks. Among detection-based approaches, the YOLO (You Only Look Once) architecture has demonstrated robust real-time performance (Siriani et al., 2023; Qin et al., 2025).

NUEJ
Naresuan University
Engineering Journal

Specific adaptations for poultry have yielded high results; for instance, Zhu et al. (2022) achieved 95.87% accuracy in dense flocks, while Guo et al. (2023) improved YOLOv5 using attention mechanisms (CBAM) to reach 97.3% precision. More recently, Wu et al. (2025) proposed YOLO-CCA to enhance F1-scores, and comparative studies by Bumbálek et al. (2025) suggest that while YOLOv9c achieves the highest precision, YOLOv11n offers the fastest inference speed.

However, detection alone is insufficient for continuous counting. Tracking algorithms such as SORT (Bewley et al., 2016; Wojke et al., 2017) and DeepSORT are instrumental in maintaining object identity. Yang et al. (2024) demonstrated that combining YOLOv8 with DeepSORT achieved 94% MOTA in cage-free hen monitoring. To further enhance robustness against environmental noise, motion filtering techniques like MOG2 have been validated by Garcia-Garcia et al. (2020) and Iseki et al. (2025), while Stopassola et al. (2021) highlighted that pairing these with optimized thresholds improves precision. Alternatively, for extreme crowding, density-based models like DFCCNet (Lv et al., 2023) and CSRNet (Li et al., 2018) offer effective counting via density maps rather than bounding boxes. Table 1 summarizes these core AI components using real-world analogies.

Table 1 Analogy-based summary of core AI components

Component	Real-World Analogy	Method used
Object Detection	Face scanner or barcode reader	YOLOv8
Object Tracking	Bib number tracking in a race	YOLO + SORT
Motion Filtering	Audio noise cancellation	MOG2
Density Map	Estimating a crowd via drone	DFCCNet, CSRNet
Prompt Engineering	Giving instructions to a smart assistant	ChatGPT + LLMs

Beyond the vision pipeline, Prompt Engineering is emerging as a transformative method for system development. It involves structured input design to guide Large Language Models (LLMs) in executing complex tasks (Chen et al., 2023). Recent work by Xue et al. (2025) showed that LLMs can iteratively optimize system architectures to achieve near-human accuracy. Furthermore, Sahoo et al. (2024) emphasized its strategic value in both vision and language tasks. In this study, LLMs were leveraged to generate initial YOLO pipeline templates, suggest error-handling routines, and assist in debugging integration, effectively democratizing access to advanced AI solutions.

This study aims to develop a robust, real-time chicken counting system for slaughterhouses by benchmarking three ROI-based strategies (Central Box, Midline Crossing, and Left-edge Exit) against industrial challenges. Built on the YOLOv8 architecture and validated under real production conditions, the system addresses the gap in systematic benchmarking for high-speed conveyor lines. Performance is assessed using precision, recall, F1-score, and regression metrics (MAE, MAPE, RMSE) to identify the optimal configuration for scalable poultry operations.

2. EXPERIMENTAL SETUP AND METHODOLOGY

This study employed a comprehensive pipeline for real-time object detection and counting of poultry in an industrial processing environment. The dataset comprised 5-minute video clips recorded at a resolution of 1280 × 720 pixels and 30 frames per second, yielding approximately 9,000 frames per clip. On average, each video contained approximately 645 chickens , with an average density of 11 chickens visible per frame depending on the conveyor speed.

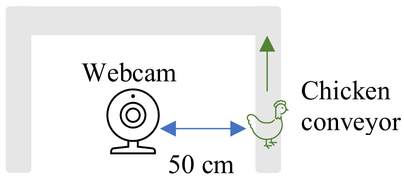

Figure 1 Setup USB webcam to chicken conveyor

Show the experimental data collection was performed using a USB webcam installed in the post-plucking area. The camera was mounted at a height of approximately 120 cm, maintaining a working distance of approximately 50 cm from the suspended chicken carcasses to ensure an optimal field of view (Figure 1). The frontal view was chosen to minimize obscuration between adjacent chicken carcasses, with lighting conditions typical of the processing process. All video processing and inference tasks were executed on a Windows 11 Pro system equipped with an Intel Core i5-12400 CPU (2.50 GHz), 8 GB RAM, and no dedicated GPU. The software stack included Python v3.10.11, OpenCV v4.8.1, and PyTorch v2.7.1 (CPU-only version). Object detection was performed using YOLOv8 (Ultralytics). Video input was captured live from a USB webcam (720p @ 30 FPS), and CUDA acceleration was not available.

The training, validation, and testing of YOLOv8 were conducted on a manually annotated dataset of 200 frames extracted from video recordings of chickens suspended on the processing line. These annotated images were collected across separate recording days to reduce temporal bias and ensure variation in carcass presentation and environmental conditions. The trained YOLOv8 model was then applied to an independent five-minute video clip, recorded on a different day, which was not part of the annotated dataset. This five-minute clip was used solely for benchmarking real-time counting performance and was not included in the training process.

YOLOv8 was selected for object detection, trained on a custom-labeled dataset of chickens with bounding boxes. A confidence threshold of 0.3 was applied, and Non-Maximum Suppression (NMS) was used to eliminate duplicate detections. Prior studies have shown that YOLOv8 can achieve over 95% detection accuracy in real-world environments (Zhu et al., 2022; Farjon et al., 2023).

To mitigate false positives arising from static background elements such as crate rails and conveyor edges, background subtraction was implemented using the Mixture of Gaussians version 2 (MOG2) algorithm from OpenCV. The algorithm was configured with a history of 100 frames and a variance threshold of 40, following the methodology proposed by Zivkovic and van der Heijden (2006). For multi-object tracking, the Simple Online and Realtime Tracking (SORT) algorithm was employed to maintain consistent object identities across frames. Key parameters included max_age = 30, min_hits = 3, and an Intersection-over-Union (IoU) threshold of 0.3. SORT has been validated for robust tracking in dynamic industrial environments (Bewley et al., 2016; Wojke et al., 2017).

To handle motion blur and temporary occlusions common in high-speed conveyors, the system relies on a multi-stage filtering mechanism rather than aggressive image pre-processing, which could induce latency. First, a confidence threshold of 0.3 was selected to prioritize Recall at the detection stage, ensuring no chicken is missed due to blur or lighting conditions. While this low threshold increases sensitivity, the risk of false positives is mitigated by the subsequent stages. The SORT algorithm acts as a temporal filter, validating detections based on trajectory consistency; transient noise or flickering detections that fail to establish a stable track over consecutive frames are discarded. Subsequently, the ROI counting logic serves as a spatial filter, ensuring that only objects exhibiting linear motion through the defined counting zone are registered. This synergy allows the system to maintain high sensitivity without compromising counting precision.

Three Region of Interest (ROI)-based strategies were developed for counting chickens: (1) the Central Box method, which counted chickens whose bounding box centers remained within a central region for at least two consecutive frames; (2) the Midline Crossing method, which triggered a count when an object's center crossed the vertical midline, with time-based suppression to prevent duplicate counts; and (3) the Left-edge Exit method, which counted objects exiting the left frame boundary, using track IDs to avoid repeated counts.

Detection results were logged in CSV format, with each entry containing the timestamp, object track ID, class ID, confidence score, and bounding box coordinates (x1, y1, x2, y2). This structured output facilitated subsequent auditing and quantitative analysis (Krizhevsky et al., 2012).

System performance was evaluated using three standard metrics, i.e., Precision, Recall, and F1-score, defined respectively as:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \tag{2}$$

$$\text{F1} - \text{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{3}$$

Where:

- TP: True Positives — Correctly detected chickens
- FP: False Positives — Incorrectly detected objects as chickens
- FN: False Negatives — Missed detections of actual chickens

These metrics are widely adopted in poultry detection research (Guo et al., 2023; Pangestu, 2025), with enhancements such as YOLOv5-CBAM shown to improve F1-score (Cheng et al., 2024).

In addition to classification metrics, the system's counting accuracy was rigorously evaluated using regression metrics to quantify performance over time. Specifically, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) were calculated based on cumulative counts over 30-second intervals.

Prompt engineering techniques were employed to assist in the development of the detection and tracking pipeline using GPT-4. Structured prompts incorporating role definitions, code constraints, and iterative feedback mechanisms were used to generate and refine Python scripts for each subsystem (Brown et al., 2020).

Statistical analysis of counting strategies was conducted using Python libraries statsmodels and scipy. Data normality was assessed via the Shapiro–Wilk test. Depending on the outcome, either one-way ANOVA (for normally distributed data) or the Kruskal–Wallis H-test (for non-normal data) was applied. Post hoc pairwise comparisons were performed using Tukey's HSD following ANOVA or the Mann–Whitney U test with Bonferroni correction following Kruskal–Wallis. Visualization tools included boxplots and kernel density plots of confidence scores, heatmaps of pairwise-adjusted p-values, and dendrograms for hierarchical clustering of counting strategies.

3. RESULTS AND DISCUSSION

The researcher applied the concept of Prompt Engineering with LLMs to build a real-time system composed of object detection, object tracking, counting, and result logging. The LLM helped modularize the code structure and automatically generated pseudocode or a skeleton pipeline from a single command. For example, changing the counting strategy from “middle-frame counting” to “left-edge counting” could be done by simply adjusting the prompt without altering low-level code or retraining the model. This made iterative system testing and development in a human in the loop format rapid and continuous. This concept aligns with recent studies in LLM-aided design, which show that LLMs can support all phases of system development including conceptualization, prototyping, verification, and optimization without requiring machine learning expertise (Gu et al., 2023; Sahoo et al., 2024). Research by Cruz et al. (2024) and Siriani et al. (2023) also confirms that LLMs can accelerate development, reduce errors, and streamline deployment in vision and object tracking systems.

The automated chicken counting system was evaluated not only for accuracy but also for computational efficiency, a critical factor for "Real-Time" applications. Running on the specified hardware, the system achieved an average inference time of 56.92 milliseconds per frame, translating to approximately 17.57 Frames Per Second (FPS). This performance confirms the system's capability to operate in real-time alongside the conveyor speed.

Object Detection Performance and Regression Metrics To evaluate each strategy, the system’s performance was statistically compared using Precision, Recall, F1-score, and Regression Metrics (MAE, MAPE, RMSE) based on a ground truth count of 645 chickens.

3.1 System and Input Data

The system operates as a sequential pipeline where YOLOv8 serves as the primary detection engine (Figure 2), extracting spatial coordinates of poultry from the actual processing line footage. These coordinates are then fed into three distinct ROI-based counting modules (Box Area, Middle Line, and Left Edge), which act as decision triggers to convert raw detections into cumulative counts based on specific spatial rules.

Each processed frame generated structured output consisting of timestamp, Object ID, Class ID, confidence score, and bounding box coordinates (x1, y1, x2, y2), as shown in Table 2.

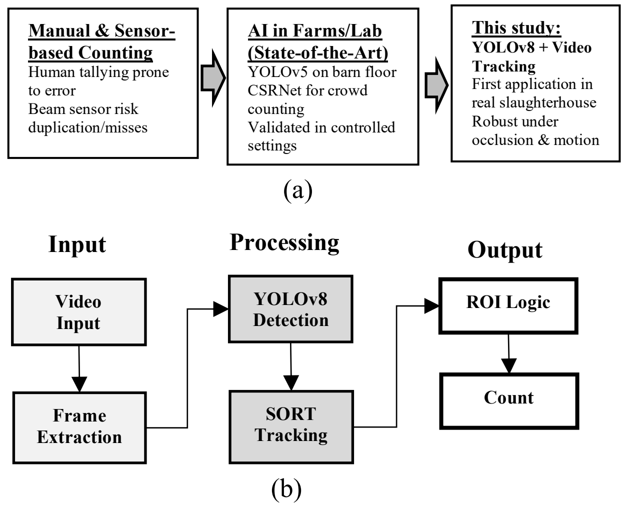

Figure 2 (a) State of-the-art and (b) chicken counting system developed using YOLOv8

Table 2 Sample structured output per frame

Timestamp	Object ID	Class ID	Confidence	x1, y1 (top-left)	x2, y2 (bottom-right)
12:00:01	01	Chicken	0.83	112, 305	202, 410
12:00:01	02	Chicken	0.76	215, 300	305, 395

3.2 Object Detection Performance

To evaluate each strategy, the system’s performance was statistically compared using Precision, Recall, and F1-score, based on a ground truth count of 645 chickens, as shown in Table 3.

Table 3 Accuracy Comparison of Six Strategies

Method	Count	Ground Truth	True Positive	False Positive	False Negative	Precision (%)	Recall (%)	F1 Score
Left edge-webcam	560	645	560	0	85	1.000	0.868	0.929
Middle line-webcam	579	645	579	0	66	1.000	0.898	0.946
Box area-webcam	696	645	645	51	0	0.927	1.000	0.962
Left edge-VDO	602	645	602	0	43	1.000	0.933	0.966
Middle line-VDO	643	645	643	0	2	1.000	0.997	0.998
Box area-VDO	649	645	645	4	0	0.994	1.000	0.997

Regression Analysis of Counting Accuracy to provide a more rigorous evaluation than simple total count comparison, regression metrics were calculated based on 30-second cumulative intervals. Table 4 presents the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) for all six strategies.

TABLE 4. Regression Metrics Comparison (MAE, RMSE, MAPE)

Strategy	MAE (Count)	RMSE (Count)	MAPE (%)
Line (Video)	**2.30**	**2.70**	**0.74**
Box (Video)	8.00	9.09	2.48
Edge (Video)	25.60	28.05	7.71
Box (Image)	23.10	27.76	5.85
Line (Image)	37.30	41.04	11.43
Edge (Image)	49.40	55.21	14.14

The results unequivocally demonstrate the superiority of video-based processing over static image-based methods. The Line (Video) strategy achieved the best overall performance with an exceptionally low MAPE of 0.74% and an MAE of 2.30, indicating that, on average, the system deviates by only approximately 2 chickens per 30-second interval.

In contrast, image-based strategies showed high volatility. The Edge (Image) strategy performed the poorest, with an MAE of 49.40 and MAPE of 14.14%. Even the best image-based method (Box-Image) had an error rate (MAE 23.10) nearly ten times higher than the best video method. This substantial difference highlights the critical role of temporal information in tracking algorithms (SORT) to resolve occlusions and maintain object identities on high-speed conveyors.

In the detection display, red boxes represent counted chickens, while green boxes are uncounted (Figure 3). Three ROI-based counting strategies are applied to slaughterhouse conveyor footage. Insets highlight typical error sources are (a) Central Box, where partial occlusion leads to missed counts; (b) Midline Crossing, where motion blur during conveyor movement causes inconsistent detection; and (c) Left-edge Exit, where overlapping carcasses at the belt margin increase undercounting.

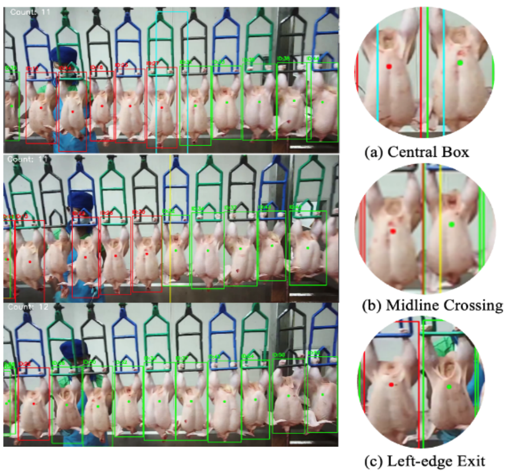

Figure 3 Comparison of three ROI-based counting strategies applied to slaughterhouse conveyor footage

Figures 3. Examples of counting errors observed during slaughterhouse testing, where (a) occlusion within the central ROI caused repeated detections and cumulative overcount, (b) rapid movement and partial occlusion led to missed midline-crossing events and undercount, and (c) object loss near the image boundary resulted in undercount due to incomplete trajectory tracking.

Figures 4 through 5 provide a comprehensive visual analysis of system performance across all six counting strategies. Figure 5 presents a bar chart comparing Precision, Recall, and F1-score, where video-based strategies consistently outperform still-image approaches, particularly in terms of Recall and F1-score. Figure 4 complements this with a line chart that captures the trends and variability of each metric across strategies, enabling clearer interpretation of the trade-offs between detection completeness and precision. Figure 6 further illustrates the absolute number of chickens detected by each method relative to the ground truth (645 chickens), using a horizontal reference line to highlight undercounting and overcounting behaviors.

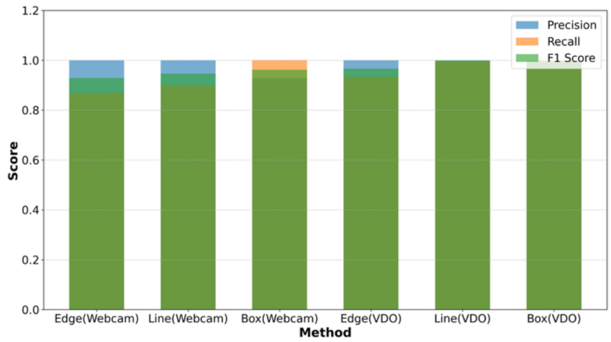

Figure 4 Comparison of Precision, Recall, and F1 Score across all six strategies

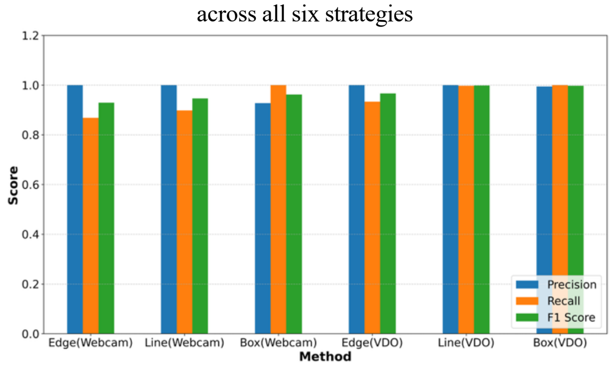

Figure 5 Line chart showing variation in metrics across strategies

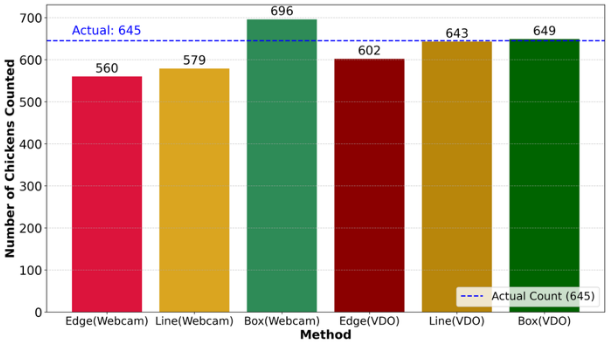

Figure 6 Detected chicken counts per method vs ground truth

Error Analysis and Robustness To address the robustness of the model, specific scenarios such as empty shackles (missing chickens) and foreign objects were analyzed. The model successfully distinguished between chickens and empty shackles, resulting in zero false positives from empty hooks. However, some errors persisted.

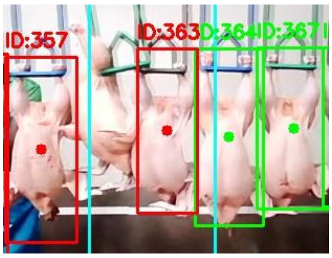

Figure 7 Examples of False Negatives

As shown in Figure 7, False Negatives (FN) primarily occurred due to extreme occlusion where two chickens overlapped significantly, the chickens are not hung properly, or motion blur caused by conveyor vibration. False Positives (FP) were rare but sometimes by double counting of the box area.

The integrated analysis of these visualizations and the corresponding quantitative data reveals that the Middle Line- VDO strategy delivered the highest overall performance across all metrics, followed closely by Box Area-VDO. In contrast, strategies based on still images exhibited the most deviation from the ground truth, particularly in Recall, which reflects missed detections (false negatives). These findings suggest that static image input is less suitable for accurately detecting fast-moving poultry on conveyor lines, where continuous motion and temporal context significantly enhance detection and tracking stability.

From the combined analysis of these figures and the supporting data tables, the Middle Line-VDO strategy yielded the best performance across all metrics, closely followed by the Box Area-VDO strategy. Conversely, still-image-based strategies showed the greatest deviation from the actual count, particularly in Recall, which reflects the rate of missed detections (false negatives). This suggests that static image input is less suited for fast-moving objects such as chickens on a production line, where motion continuity aids detection and tracking consistency.

3.3 Detection Confidence and Variability

Confidence scores were analyzed using mean, standard deviation (SD), and coefficient of variation (CV%) to assess system stability. Table 4 shows that video-based strategies (Middle Line-VDO and Box Area-VDO) yielded high mean confidence (0.727, 0.718) and low CV (<9%), indicating consistent detection.

In contrast, still image strategies (e.g., Box Area-Webcam and Left Edge-Webcam) had lower mean confidence (0.57–0.59) and higher variability (CV ~17–18%), suggesting less reliable performance.

Table 5 Confidence Statistics and CV% per Strategy

Strategy	Detection (Count)	Mean Confidence (Score)	Std. Dev. (Score)	Coefficient of Variation (%)
Left edge-webcam	560	0.590	0.109	18.46
Middle line-webcam	579	0.590	0.098	16.66
Box area-webcam	696	0.571	0.100	17.44
Left edge- VDO	602	0.574	0.097	16.95
Middle line- VDO	643	0.727	0.062	8.53
Box area-VDO	649	0.718	0.062	8.62

The results from Table 5 highlight the advantages of using video input for object detection on continuously moving production lines. Video-based strategies not only reduce occlusion-related errors, but also provide higher and more consistent confidence scores, contributing to a more stable system when deployed in real-world industrial environments.

The distribution of confidence scores for each strategy was further analysed using boxplots and density plots, as shown in Figures 8 and 9, respectively.

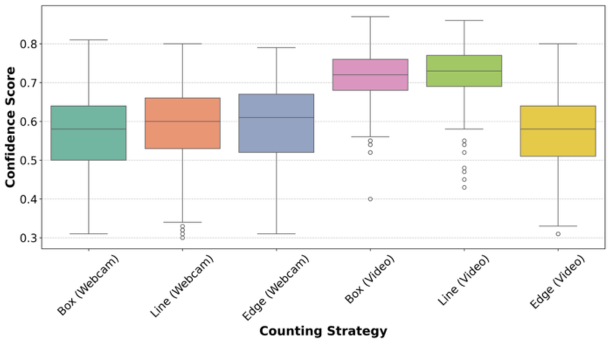

Figure 8 Boxplot showing interquartile range (IQR) and outliers in confidence scores

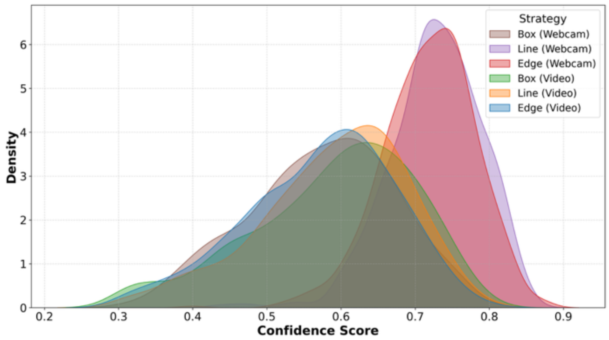

Figure 9 Density plot indicating distribution shape and skewness in confidence scores

The boxplot clearly illustrates the interquartile range (IQR) and identifies outliers in each strategy, providing a concise visualisation of confidence score variability. The density plot reveals the distribution shape, including skewness and the concentration of scores within high or low intervals.

From both figures, it is evident that video-based strategies, such as Line (Video) and Box (Video), exhibit narrower and steeper distributions. This reflects higher consistency in detection, which aligns with their low coefficient of variation (CV%) reported in Table 4. These results reaffirm that video input not only enhances accuracy but also improves the system's robustness and reliability under dynamic production conditions.

3.4 Statistical Significance Testing

The Kruskal–Wallis H-test revealed statistically significant differences in confidence scores among the six strategies ($H = 1576.54$, $p < 0.001$).

Pairwise comparisons using the Mann–Whitney U Test with Bonferroni-adjusted p-values showed that Edge (Video) differed significantly from almost all others, while no significant difference was found between Edge-Webcam and Line-Webcam, see Table 6.

A primary section heading is enumerated by a capital letter and is centered above the paragraph text.

A secondary section (subsection) heading is enumerated by a capital letter followed by a period and is flush on the left of the column. All letters of each important word is capitalized. The text style is italic.

Table 6 Mann–Whitney U Test Results with Bonferroni Adjustment

Strategy 1	Strategy 2	Strategy 3
Edge (Webcam)	Edge (Video)	< 0.001
Line (Webcam)	Edge (Video)	< 0.001
Box (Video)	Line (Video)	< 0.001
Box (Webcam)	Edge (Video)	< 0.001
Line (Webcam)	Edge (Webcam)	0.992

3.5 Structural Analysis of Strategy Differences: Heatmap and Dendrogram

To understand the overall structure of differences between detection strategies, a heatmap was generated using Bonferroni-adjusted p-values from all pairwise comparisons, as shown in Figure 10.

- Dark tones in the heatmap indicate pairs with statistically significant differences ($p < 0.05$).
- Light tones represent pairs with no statistically significant difference.

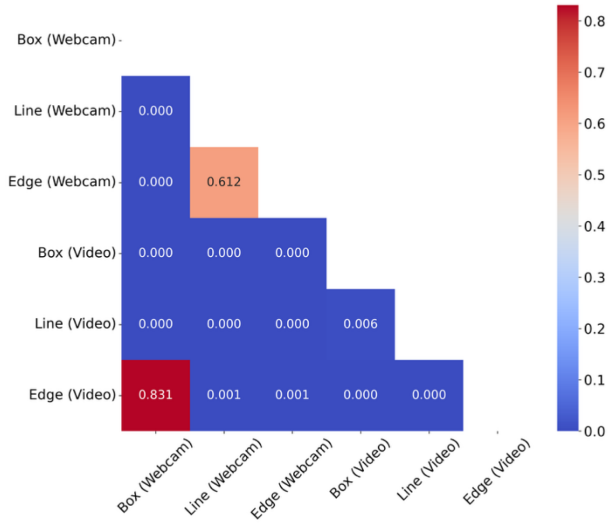

Figure 10 Heatmap of Bonferroni-adjusted p-values between strategies

The heatmap reveals that video-based strategies are clearly distinct from still-image-based strategies, with Line (Video) in particular showing significant differences from nearly all others.

To further illustrate statistical similarity between strategies, hierarchical clustering was performed using the dissimilarity metric ($1 - p_{adjusted}$). The results are visualized in Figure 11.

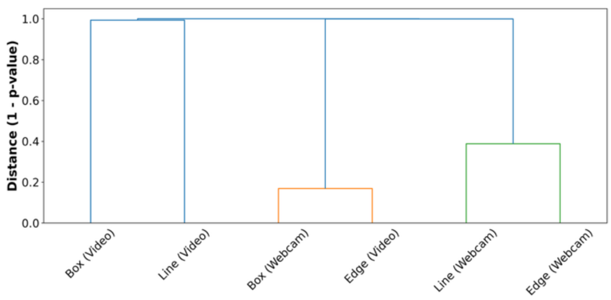

Figure 11 Dendrogram showing grouping of strategies based on statistical similarity

The dendrogram shows that Line (Video) and Box (Video) cluster closely together, reflecting similar detection behavior and high accuracy. In contrast, Edge (Webcam) is positioned far from the others, indicating a distinct and less effective detection pattern. This separation highlights its practical limitations, especially in complex or high-speed environments.

Overall, this structural analysis confirms that video-based strategies not only outperform in detection metrics but also exhibit consistent statistical characteristics, making them more reliable and suitable for industrial deployment compared to image-based approaches.

3.6 Accuracy Assessment vs Ground Truth

A direct comparison with human-labeled ground truth (645 chickens) quantified both absolute and relative errors for each strategy. Table 7 summarizes the results.

Table 7 Ground Truth vs System Count (GT = 645)

Strategy	Counted	Error (±)	Relative Error (%)
Box (Webcam)	696	+51	+7.91
Line (Webcam)	579	-66	-10.23
Edge (Webcam)	560	-85	-13.18
Line (Video)	643	-2	-0.31
Box (Video)	649	+4	+0.62
Edge (Video)	602	-43	-6.67

Video-based strategies showed minimal error, with Line (Video) deviating by only 0.31% (2 chickens), confirming high reliability. Still-image strategies, especially Edge (Image), showed the largest deviation (13.18%).

These findings align with earlier results in confidence scores and CV%, reinforcing that video strategies are more stable and practical in real-world conveyor scenarios.

3.7 Relation to Prior Studies

These findings support prior research by Wu et al. (2025), Zhang et al. (2019), and Cheng et al. (2024), which found that video inputs are superior for detecting fast-moving poultry or pigs in processing lines. Specifically, ROI strategies like Middle Line reduce duplication and occlusion errors.

Additionally, low CV values in video strategies indicate detection stability. Heatmap and dendrogram analysis confirmed statistically distinct behaviors between image- and video-based methods, offering insights for future system design tailored to real-world environments.

3.8 Limitations of the Study

While the proposed system demonstrates high accuracy in a real industrial setting, several limitations must be acknowledged. First, the validation was conducted using a single continuous video sequence from one specific slaughterhouse production line. Consequently, the system's robustness against significant environmental variations such as drastic changes in lighting, different conveyor speeds, or alternative poultry breeds was not extensively tested. Second, the dataset size used for validation is relatively small compared to large-scale public benchmarks, which may limit the generalization of the findings. These factors suggest that while the current ROI-based strategies are effective for the tested environment, further validation on larger, multi-source datasets is required to ensure broad applicability.

4. CONCLUSION

This study provides the first empirical evaluation of ROI-based automated chicken counting strategies under real slaughterhouse conveyor conditions. Results show that video-based approaches, particularly the Midline and Central Box ROIs, achieved the highest performance. Instead of relying solely on count differences, regression metrics confirmed the precision of the system. The proposed Midline Crossing (Video) strategy achieved a Mean Absolute Error (MAE) of 2.30, a Mean Absolute Percentage Error (MAPE) of 0.74%, and a Root Mean Square Error (RMSE) of 2.70, proving its reliability for industrial application. In contrast, edge-based strategies tended to undercount due to occlusion and camera angle limitations, despite their effectiveness in avoiding double counting. Aligning ROIs with the conveyor center and relying on video input instead of static images significantly improved detection stability, as reflected by a low coefficient of variation (CV < 9%). Furthermore, the system demonstrated robust real-time capabilities with an average processing speed of 17.57 FPS (56.92 ms inference time) on standard CPU hardware, confirming its feasibility for continuous monitoring at industrial conveyor speeds without requiring high-end GPU acceleration.

For industrial deployment, cameras should be installed perpendicularly above the conveyor to minimize occlusion, with confidence thresholds set between 0.65 and 0.70 to balance sensitivity and false positives. The integration of advanced tracking algorithms, such as DeepSORT or BoT-SORT, can further enhance recall and identity consistency. Adding feedback mechanisms to flag miscounts and confidence anomalies would also enable iterative model refinement and active learning. While Midline and Central Box ROIs achieved the highest accuracy, qualitative inspection revealed distinct error sources. Undercounting in Edge ROIs often resulted from carcass occlusion at the belt margins, while occasional misclassifications were linked to motion blur during peak conveyor speed, specular reflections from metallic surfaces, and carcasses appearing in abnormal poses. Providing a taxonomy of these error types highlights the operational challenges that simple ROI placement cannot fully resolve.

This study did not perform a full robustness test; however, several industrial factors are likely to affect performance, including camera height and tilt, LED flicker and banding, water droplets on lenses, and vibration from processing equipment. Future research should systematically vary these conditions to quantify their impact on accuracy and ensure reliable deployment across diverse slaughterhouse environments. Limitations of this study include the relatively small dataset, consisting of only 200 annotated images used for training, validation, and testing. Although these frames were collected across different recording days, the overall sample size remains restricted and may lead to sampling bias or overestimation of performance. In addition, system-level validation relied on a single independent five-minute video clip recorded on a separate day, which, while not part of the training data, still provides limited coverage of operational variability. Other limitations include the absence of robustness testing under variable conveyor speeds, inconsistent lighting, and camera vibration, as well as the lack of ablation studies on LLM-assisted development. While this setup is sufficient for proof-of-concept evaluation, future research should expand data collection across multiple days, shifts, and facilities, supported by larger annotated datasets, to strengthen robustness and generalizability in real slaughterhouse deployment. Future research should therefore expand dataset diversity across multiple clips, production shifts, and slaughterhouse environments; report additional effect size metrics (e.g., Cliff's Delta, mAP); explore domain adaptation techniques for cross-site generalization; and conduct robustness testing across diverse industrial conditions.

In conclusion, the proposed YOLOv8-based chicken counting system demonstrates high accuracy and reliability in real slaughterhouse conditions, offering practical recommendations for deployment and establishing a foundation for future innovation in automated poultry processing. Finally, while this study utilized a fixed confidence threshold of 0.3 to maximize detection recall for the tracker (Wojke et al., 2017), future research will focus on conducting a comprehensive sensitivity analysis. This will involve systematically varying threshold values to quantify their impact on precision-recall trade-offs and F1-scores, thereby fine-tuning the system for varying lighting conditions in slaughterhouse environments.

5. Acknowledgment

This work received support and funding from the Naresuan University Faculty of Engineering Scholarships.

6. References

Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. (2016). Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/ICIP.2016.7533003

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., and Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. https://doi.org/10.48550/arXiv.2005.14165

Bumbálek, R., Umurungi, S. N., Ufitikirezi, J. D. D. M., Zoubek, T., Kuneš, R., Stehlík, R., Lin, H.-I., and Bartoš, P. (2025). Deep learning in poultry farming: comparative analysis of yolov8, yolov9, yolov10, and yolov11 for dead chickens detection. *Poultry Science*, 104(9):105440. https://doi.org/10.1016/j.psj.2025.105440

Chen, B., Zhang, Z., Langrené, N., and Zhu, S. (2023). Unleashing the potential of prompt engineering in large language models: a comprehensive review. *arXiv preprint arXiv:2310.14735*. https://doi.org/10.48550/arXiv.2310.14735

Cheng, T., Song, L., Ge, Y., Liu, W., and Wang, X. (2024). Yolo-world: Real-time open-vocabulary object detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR52733.2024.01599

Cruz, E., Hidalgo-Rodriguez, M., Acosta-Reyes, A. M., Rangel, J. C., and Boniche, K. (2024). Ai-based monitoring for enhanced poultry flock management. *Agriculture*, 14(12):2187. https://doi.org/10.3390/agriculture14122187

Farjon, G., Huijun, L., and Edan, Y. (2023). Deep-learning-based counting methods, datasets, and applications in agriculture: a review. *Precision Agriculture*, 24(5):1683–1711. https://doi.org/10.1007/s11119-023-10034-8

Feng, Y., Pallerla, C., Lin, X., Sohrabipour, S., Crandall, P., Shou, W., She, Y., and Wang, D. (2025). Synthetic data augmentation for enhanced chicken carcass instance segmentation. *arXiv preprint arXiv:2507.18558*. https://doi.org/10.48550/arXiv.2507.18558

Garcia-Garcia, B., Bouwmans, T., and Silva, A. J. R. (2020). Background subtraction in real applications: Challenges, current models and future directions. *Computer Science Review*, 35:100204. https://doi.org/10.48550/arXiv.1901.03577

Gu, J., Han, Z., Chen, S., Beirami, A., He, B., Zhang, G., Liao, R., Qin, Y., Tresp, V., and Torr, P. (2023). A systematic survey of prompt engineering on vision-language foundation models. *arXiv preprint arXiv:2307.12980*. https://doi.org/10.48550/arXiv.2307.12980

Guo, Y., Aggrey, S. E., Yang, X., Oladeinde, A., Qiao, Y., and Chai, L. (2023). Detecting broiler chickens on litter floor with the yolov5-cbam deep learning model. *Artificial Intelligence in Agriculture*, 9:36–45. https://doi.org/10.1016/j.aiia.2023.08.002

Iseki, H., Furukawa, E., Shimasaki, T., and Higaki, S. (2025). Automatic monitoring of activity intensity in a chicken flock using a computer vision-based background image subtraction technique: An experimental infection study with fowl adenovirus. *Smart Agricultural Technology*, 10:100821. https://doi.org/10.1016/j.atech.2025.100821

Jiang, B., Tang, W., Cui, L., and Deng, X. (2023). Precision livestock farming research: A global scientometric review. *Animals*, 13(13):2096. https://www.mdpi.com/2076-2615/13/13/2096

Khanal, R., Choi, Y., and Lee, J. (2024). Transforming poultry farming: A pyramid vision transformer approach for accurate chicken counting in smart farm environments. *Sensors*, 24(10):2977. https://www.mdpi.com/1424-8220/24/10/2977

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25. https://doi.org/10.1145/3065386

Li, Y., Zhang, X., and Chen, D. (2018). Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2018.00120

Lv, J., Wang, J., Peng, C., and Huang, Q. (2023). Dfccnet: A dense flock of chickens counting network based on density map regression. *Animals*, 13(23):3729. https://www.mdpi.com/2076-2615/13/23/3729

Novus International, Inc. (2025). Perspectives on Thailand's poultry industry. Retrieved July 27, 2025 from https://www.novusint.com/resources/perspectives-on-thailands-poultry-industry/

Okinda, C., Nyalala, I., Korohou, T., Okinda, C., Wang, J., Achieng, T., and Shen, M. (2020). A review on computer vision systems in monitoring of poultry: A welfare perspective. *Artificial Intelligence in Agriculture*, 4:184–208. https://doi.org/10.1016/j.aiia.2020.09.002

Pangestu, G. (2025). Real-time chicken counting system using yolo for FCR optimization in small and medium poultry farms. *Computational Agriculture Review*, 9(1):1–9. https://doi.org/10.28926/jdr.v9i1.436

Public Relations Department of Thailand (2024). Thai chicken exports rise by almost 5 percent in the first 10 months of 2024. Retrieved July 27, 2025, from https://thailand.prd.go.th/en/content/category/detail/id/48/iid/346892

Qin, W., Yang, X., Wang, Y., Wei, Y., Zhou, Y., and Zheng, W. (2025). Yolopoul: Performance evaluation of a novel yolo object detectors benchmark for multi-class manure identification to warn about poultry digestive diseases. *Smart Agricultural Technology*, 12:101145. https://doi.org/10.1016/j.atech.2025.101145

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788. https://doi.org/10.48550/arXiv.1506.02640

Reuters (2025). Thai poultry industry poised for growth on Brazil bird flu, lower feed costs. Retrieved July 27, 2025, from https://www.reuters.com/business/healthcare-pharmaceuticals/thai-poultry-industry-poised-growth-brazil-bird-flu-lower-feed-costs-2025-05-30/

Sahoo, P., Singh, A., Saha, S., Jain, V., Mondal, S., and Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. https://doi.org/10.13140/RG.2.2.13032.65286

Siriani, A. L. R., Miranda, I. B. d. C., Mehdizadeh, S. A., and Pereira, D. F. (2023). Chicken tracking and individual bird activity monitoring using the bot-sort algorithm. *AgriEngineering*, 5(4):1677–1693. https://www.mdpi.com/2624-7402/5/4/104

Stopassola, G., de Santana Weizenmann, G., and Costa, M. (2021). Machine vision algorithm for continuous counting of poultry in slaughterhouses. In *Proceedings of the 2021 ABCM International Congress of Mechanical Engineering (COBEM)*. https://doi.org/10.26678/ABCM.COBEM2021.COB2021-1707

The Observatory of Economic Complexity (2025). Thailand poultry meat trade profile. Retrieved July 27, 2025, from https://oec.world/en/profile/bilateral-product/poultry-meat/reporter/tha

Wojke, N., Bewley, A., and Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3645–3649. https://doi.org/10.48550/arXiv.1703.07402

Wu, Z., Yang, J., Zhang, H., and Fang, C. (2025). Enhanced methodology and experimental research for caged chicken counting based on yolov8. *Animals (Basel)*, 15(6). https://doi.org/10.3390/ani15060853

Xue, E., Huang, Z., Ji, Y., and Wang, H. (2025). Improve: Iterative model pipeline refinement and optimization leveraging LLM experts. https://doi.org/10.48550/arXiv.2502.18530

Yang, X., Bist, R., Paneru, B., and Chai, L. (2024). Monitoring activity index and behaviors of cage-free hens with advanced deep learning technologies. *Poultry Science*, 103(11):104193. https://doi.org/10.1016/j.psj.2024.104193

Zhang, L., Gray, H., Ye, X., Collins, L., and Allinson, N. (2019). Automatic individual pig detection and tracking in pig farms. *Sensors*, 19:1188. https://doi.org/10.3390/s19051188

Zhu, X., Wu, C., Yang, Y., Yao, Y., and Wu, Y. (2022). Automated chicken counting using yolo-v5x algorithm. In *2022 IEEE International Conference on Smart Agriculture and Intelligent Systems (ICSAI)*. https://doi.org/10.1109/ICSAI57119.2022.10005522

Zivkovic, Z. and van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780. https://doi.org/10.1016/j.patrec.2005.11.005