



ARTICLE

Deep Learning-Based Glass Detection for Smart Glass Manufacturing Processes

Seungmin Lee¹, Beomseong Kim² and Heesung Lee^{3,*}

¹Department of AI Transportation Convergence, Korea National University of Transportation, Uiwang-si, 16106, Republic of Korea

²Department of Artificial Intelligence, Gyeonggi University of Science and Technology, Siheung-si, 15073, Republic of Korea

³Department of Railroad Electrical and Information Engineering, Korea National University of Transportation, Uiwang-si, 16106, Republic of Korea

*Corresponding Author: Heesung Lee. Email: hslee0717@ut.ac.kr

Received: 31 March 2025; Accepted: 16 May 2025; Published: 09 June 2025

ABSTRACT: This study proposes an advanced vision-based technology for detecting glass products and identifying defects in a smart glass factory production environment. Leveraging artificial intelligence (AI) and computer vision, the research aims to automate glass detection processes and maximize production efficiency. The primary focus is on developing a precise glass detection and quality management system tailored to smart manufacturing environments. The proposed system utilizes the various YOLO (You Only Look Once) models for glass detection, comparing their performance to identify the most effective architecture. Input images are preprocessed using a Gaussian Mixture Model (GMM) to remove background noise present in factory environments. This approach minimizes distractions caused by varying backgrounds and enables accurate glass identification and defect detection. Traditional manual inspection methods often require skilled labor, are time-intensive, and may lack consistency. In contrast, the proposed vision-based system ensures high accuracy and reliability through non-contact inspection. The performance of the system was evaluated using video data collected from an actual glass factory. This assessment verified the accuracy, reliability, and practicality of the system, demonstrating its effectiveness in real-world production scenarios. Beyond automating glass detection and defect identification, the proposed system integrates into manufacturing environments to support data-driven decision-making. This enables real-time monitoring, defect prediction, and improved production efficiency. Moreover, this research is expected to serve as a model for enhancing quality control and productivity across various manufacturing industries, driving innovation in smart manufacturing.

KEYWORDS: Object detection; glass detection; artificial intelligence (AI); smart manufacturing; quality management

1 Introduction

Smart manufacturing, a core concept in modern industries, maximizes the manufacturing process efficiency and realizes customer-tailored production by integrating information and communication technology with automation technology. In particular, the introduction of real-time data analysis and decision-support systems is essential for manufacturers to maintain competitiveness and achieve sustainable growth [1–4]. Recent advancements in artificial intelligence, particularly deep learning technology, have accelerated the automation of manufacturing processes. Moreover, quality inspection systems have evolved into more precise and efficient methods utilizing deep-learning-based computer vision technologies. However, despite these technological advancements, several manufacturers rely on traditional manual quality inspection methods, which depend on the subjective judgment of workers, leading to inconsistent inspection results, a higher likelihood of errors, and ultimately, reduced productivity [5,6].



Generally, tempered glass, which is an essential material across industries, produced in factories varies in size and specifications according to order and is placed arbitrarily on production lines, making uniform measurements difficult [7–10]. In traditional manual methods, workers must measure the size of the glass directly, limiting the production speed. To address this issue, herein we propose a vision recognition-based glass detection method capable of automatically identifying the position of glass. Automated glass detection is essential to improve the accuracy of quality control and production efficiency [11–14]. This study aimed to develop an automated quality inspection system to resolve these challenges.

Many deep learning models have been proposed for glass detection [15–18], such as GDNet [15] and GlassNet [16]. However, their performance declines in industrial environments with complex backgrounds and lighting, as they are mainly trained on general scenes. This limitation is compounded by the scarcity of glass-specific images in public datasets and domain differences, which further reduce detection accuracy in manufacturing settings.

To overcome these issues, this study proposes a video transformation module that utilizes various image processing techniques to remove complex backgrounds from glass manufacturing processes and emphasize the features of glass. This approach demonstrates the effective detection of transparent objects even in challenging environments in which existing deep-learning-based object detection models face difficulties. Images that emphasize only the glass regions help deep-learning-based detection models learn more robustly and achieve high-performance glass detection. This study applies a hybrid approach combining traditional computer vision techniques, such as the Affine transform [19–21] and Gaussian Mixture Model (GMM) [22–24], with deep learning technologies, such as YOLO [25–28], to propose an adaptive glass detection method suited for complex environments. This study contributes to enhancing the efficiency of quality inspection systems in smart manufacturing environments by improving traditional manual quality inspection methods and enabling non-contact quality inspections based on vision recognition.

The structure of this paper is as follows. Section 2 analyzes the current challenges faced by small- and medium-sized glass factories in Korea and examines the features and limitations of existing glass detection algorithms. Section 3 describes the proposed glass-detection algorithm for smart factories. Section 4 presents the experimental results and discussion. Finally, Section 5 concludes the study.

2 Related Work

2.1 Measurement Issues in Small-Scale Glass Manufacturing

Glass-manufacturing factories typically operate on a made-to-order basis, resulting in inconsistent sizes and specifications of the produced glass, as shown in Fig. 1a, and the arrangement of the glass is also variable. Currently, in small- and medium-sized glass-manufacturing factories in Korea, workers manually measure the number and size of glasses produced, as illustrated in Fig. 1b.

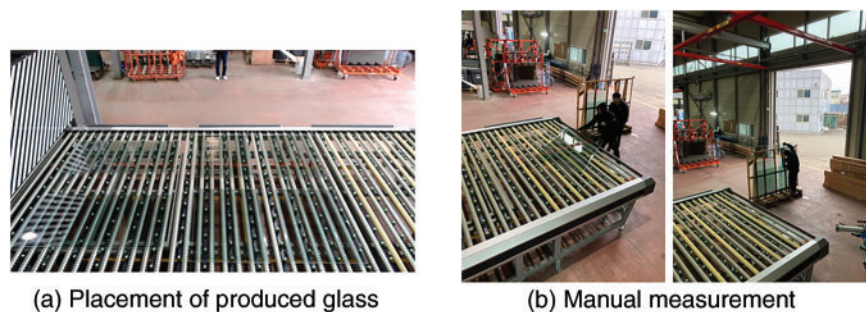


Figure 1: Measurements in small-scale glass manufacturing

This manual measurement method can compromise the consistency and accuracy of the measurements. Furthermore, because the size and arrangement of glass vary depending on the capacity and characteristics of the production equipment, the manufacturing process becomes more complex, making it challenging to perform consistent real-time measurements. Consequently, manual procedures for individually measuring and adjusting glass are commonly applied. These procedures not only slow down production processes, but also impose limitations on efficiently utilizing the maximum production capacity of equipment, ultimately leading to a decrease in overall productivity.

Glass manufactured on production equipment is placed at various positions on the conveyor belts based on its size and order specifications. The glass then moves along the conveyor belt, and workers manually measure its size for inspection purposes. This manual method relies heavily on the experience of skilled workers, which can reduce the accuracy and reproducibility of the measurement results. To address these issues, this study proposes a computer vision-based glass detection method. The proposed automated system can recognize and accurately detect glass in real time, thereby improving the production efficiency. Although the size of the glass is not measured directly, glass detection is utilized as a preprocessing step for size measurement to enhance the reliability of quality inspection systems.

2.2 Glass Detection

Current glass detection methodologies have evolved into two main approaches: detection- and segmentation-based methods. Detection-based methods estimate the locations of glass objects by generating bounding boxes, whereas segmentation-based methods detect glass objects using pixel-level classification. A representative study [29] on detection-based methods utilized a Regions with Convolutional Neural Network (R-CNN) architecture pre-trained on the beaker class of ImageNet and applied transfer learning to the PASCAL VOC dataset. This approach demonstrated the potential of deep-learning-based methods for detecting transparent objects. However, owing to limited performance, its practical application faces significant challenges. Moreover, research specifically focusing on detection-based approaches for glass detection remains insufficient.

Recently, prominent glass detection models have adopted segmentation-based approaches. The Glass Detection Network (GDNet) [15] employs Large-field Contextual Feature Integration to integrate contextual information to detect glass at various scales. In addition, it provides a Glass Detection Dataset that includes annotated data for various common glass objects and their locations. Similarly, GlassNet [16] uses a Rich Context Aggregation Module to aggregate rich contextual information at different scales and a Reflection-based Refinement Module to leverage reflection characteristics for the semantic segmentation of glass. In addition, a Glass Segmentation Dataset provides glass images, annotations, and reflection images that show reflections on the glass surface.

However, existing models trained on domain-specific datasets have shown significantly low performance in tempered glass-manufacturing environments. Moreover, GDNet and GlassNet have several limitations in their training and fine-tuning processes. To begin with, neither model provides a training code by default, requiring researchers to either reimplement the models themselves or adopt alternative approaches, which poses a significant challenge. In particular, GlassNet requires additional reflection data during the training process, which makes it difficult to achieve efficient training without constructing an appropriate dataset. These limitations hinder the fine-tuning of the model performance and further restrict the applicability of existing models in specialized conditions, such as tempered glass manufacturing environments. Fig. 2 illustrates the detection results of conventional glass detection models in such environments. GDNet mistakenly detects nonglass regions outside the conveyor as glass, whereas GlassNet fails to recognize the presence of glass. When the conveyor region is cropped and inputted, GDNet identifies most of the conveyor surface as glass, whereas GlassNet selectively detects the roller structure of the conveyor.

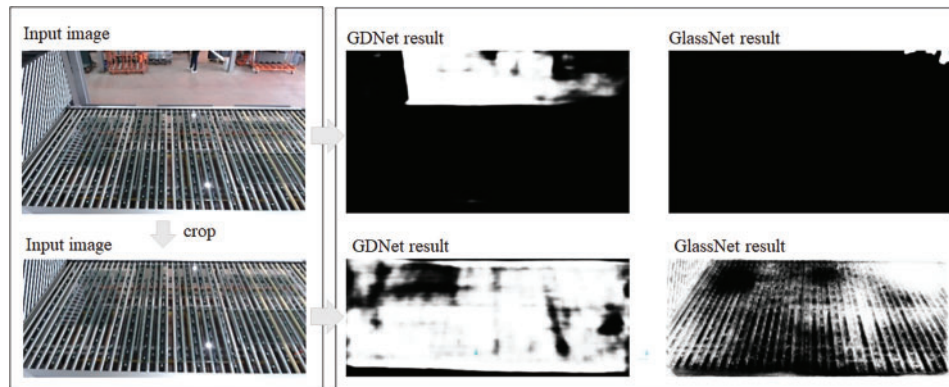


Figure 2: Detection results of GDNet and GlassNet using original and cropped input images, illustrating misclassification in a glass manufacturing environment

In addition, we conducted an in-depth analysis of various cases in which deep learning-based defect detection technologies have been applied in industrial settings. In particular, we examined how deep learning and computer vision techniques have been utilized across a range of industries, including defect detection in glass panels, textiles and packaging, waste classification, smartphone cover glass, and glass bottles. Although these studies do not address exactly the same problem as our work, they are closely related to our main keywords, such as “deep learning,” “defect detection,” “industrial application,” “glass,” and “embedded systems.” Therefore, by analyzing these previous studies, this paper provides a broad perspective on the industrial applicability and limitations of deep learning-based defect detection technologies, as well as important technical considerations for real-world implementation. The strengths and weaknesses of various algorithms under investigation in this paper are summarized in [Table 1](#). To address these limitations, this study makes the following key contributions:

1. Development of a glass detection algorithm tailored for smart glass factory environments.
2. A robust hybrid approach combining classical techniques with modern deep learning using the YOLO series.
3. Use of lightweight models optimized for real-time deployment in industrial settings, with comparative performance evaluation based on data collected from actual glass manufacturing environments.
4. Proposal of a Video Transformation Module that accelerates YOLO inference speed for practical video-based applications.

Table 1: Comparison of deep learning applications in industrial defect detection

Author	Summary	Pros	Cons
Pan Z et al., 2021 [30]	Proposed Mask R-CNN-based deep learning model and microscope camera scanning system for automatic scratch detection on architectural glass panels	• Pixel-level precision • Robust to background noise • Automated and efficient	• Requires custom equipment (coaxial lighting, Anyty 3R-CLSTMOIS microscope camera) • Speed improvement needed due to scanning with 2-axis sliding rail

(Continued)

Table 1 (continued)

Author	Summary	Pros	Cons
Ngo HQT, 2023 [31]	Designed and validated a CNN-based vision system for real-time surface defect inspection in textile and packaging industries	• Real-time automatic inspection • Low-cost open-source solution	• Requires custom equipment (servo motor, chain, lighting, etc.) • Sensitive to environment, complex setup increases initial burden
Majchrowska S, et al., 2022 [32]	Proposed new benchmark (detect-waste, classify-waste) integrating various public waste datasets and a two-stage deep learning framework	• Two-stage (detection/classification) structure addresses data imbalance/incompleteness • Semi-supervised learning improves classifier with unlabeled data	• Detection performance drops by environment, size, class • Moderate accuracy (70% detection, 75% classification)
Ghasemi Y, et al., 2022 [33]	Reviewed deep learning object detection in AR over the past decade, comparing server/local device computing	• Demonstrated AR object detection accuracy and real-time potential • Compared performance across compute environments	• Real-time limits on low-end AR devices • Lack of data diversity hinders real-world deployment
Park J, et al., 2020 [34]	Proposed ensemble deep learning model (4-DarkNet) trained on images under various lighting for automatic defect detection/classification on smartphone cover glass	• Uses diverse lighting data and ensemble weights • Lightweight for real-time industrial use	• Strongly affected by data quality • Images acquired under multiple lighting directions/methods (dark/bright field) using a line scan camera
Claypo N, et al., 2023 [35]	Proposed a system combining CNN-LSTM and instance-based classification for glass bottle defect classification	• High accuracy (98%) on rotating dataset • Efficient with small data via transfer learning	• Complex CNN-LSTM design and training • Requires dedicated rotary inspection device for dataset

3 Deep Learning-Based Glass Detection Approach

This study proposes a deep-learning-based algorithm for the effective detection of glass in smart glass manufacturing processes. Fig. 3 illustrates the overall structure of the proposed system.

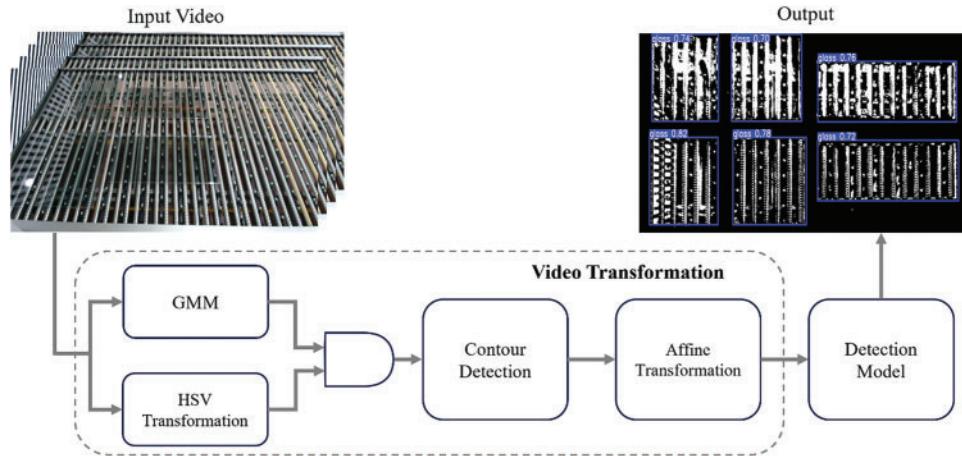


Figure 3: Model architecture

The proposed glass detection framework operates by first performing image transformation for glass detection in the Video Transformation module when the input video data are received, followed by the detection of glass using the Detection Model. Initially, the Video Transformation module performs image transformation to effectively isolate glass regions in the complex environment of the glass production processes. For this purpose, optimized image processing techniques such as Affine transformation [19–21] and the GMM [22–24] are applied to emphasize the glass regions. Subsequently, the transformed images are input into various YOLO-based detection models [25–28], in which glass is detected, and the performances are compared and analyzed. This step-by-step approach improves the accuracy of glass detection in smart glass manufacturing processes.

3.1 Video Transformation for Glass Detection

The images obtained from the glass manufacturing process include scenes of glass moving along a conveyor. These images depict multiple rollers rotating to move glass. However, because of the transparent nature of glass, as shown in Fig. 1, the conveyor rollers beneath the glass are visible through it, and reflections from factory lighting make it difficult to clearly distinguish between the glass and the background. To effectively detect glass in such complex environments, we designed a Video Transformation module, which aims to improve the accuracy and reliability of glass detection by combining techniques, such as GMM-based background removal, HSV transformation, contour detection [36–38], and Affine transformation.

In this study, the GMM was used to learn and remove backgrounds where no glass was present. GMM is a probabilistic method that models data by combining multiple Gaussian distributions, and is used for clustering, density estimation, and data distribution analysis through unsupervised learning [39]. In essence, GMM enables the identification of underlying groupings and patterns within data without requiring labeled examples, making it a versatile tool for exploratory data analysis. After applying GMM to learn background images without glass, it was used to isolate the glass regions effectively. Fig. 4a shows the results of background removal using the GMM. The experimental results demonstrate that most of the background is effectively removed, except for some noise and roller areas.

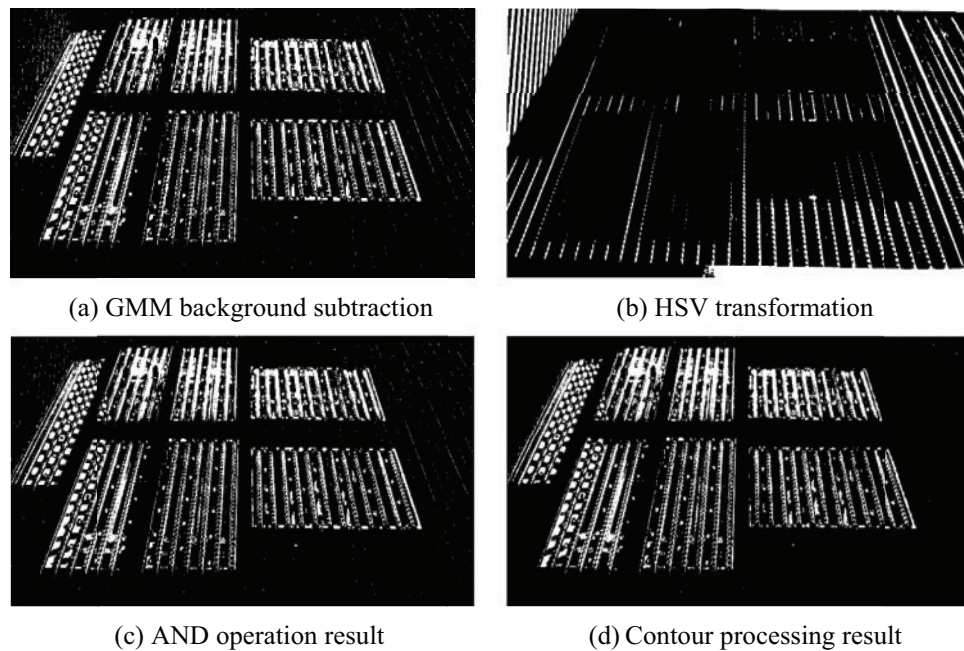


Figure 4: Image transformation and processing results

To enhance the accuracy of glass detection, HSV color-space transformation was applied. The HSV color space consists of the hue, saturation, and value (brightness). In this study, brightness thresholds were adjusted to detect glassy regions. However, reflections from lighting can cause excessive brightness on the glass surface or cause parts of the conveyor rollers to be misidentified as glass. To address this issue, the detected glass regions from the HSV transformation were combined with the results of the GMM-based background removal using an AND operation to achieve a more stable detection performance. Fig. 4b shows the inverted image after applying the HSV transformation. Fig. 4c illustrates the results of combining GMM-based background removal with HSV transformation techniques. The experimental results show that glass regions are accurately detected; however, there are instances in which non-glass regions are incorrectly identified as glass. This suggests that it is challenging to achieve perfect glass detection using a single technique.

Even after combining the GMM with the HSV transformation, some noise may remain; therefore, a contour detection algorithm [36–38] was applied to further refine the extraction of glass regions. The contour detection algorithm operates by analyzing the color differences between pixels in binary images or by grouping connected pixels to form contours. The application of this algorithm effectively removed background noise while clearly emphasizing the boundaries of the glass regions. Fig. 4d shows the results obtained after applying the contour detection algorithm, confirming that the glass regions are detected more distinctly.

Finally, Affine transformation [19–21] was applied to normalize the positions and shapes of the detected glass regions. Images obtained from the glass manufacturing process often appear tilted, depending on the camera angle. Applying Affine transformation ensures a consistent viewpoint regardless of the camera angle and reduces shape distortions caused by conveyor positioning, transforming all the detected glass into rectangular shapes. For example, without Affine transformation, glass near the edges of conveyors may appear stretched similar to parallelograms or trapezoids at central positions. However, Affine transformation preserves the true shape of the glass and allows deep-learning-based object detection models to predict the bounding boxes more accurately. Fig. 5 shows the results after applying Affine transformation.

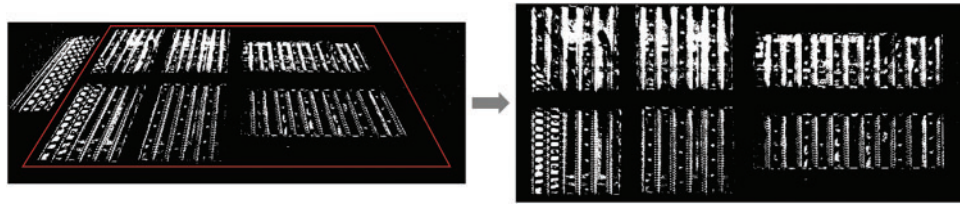


Figure 5: Affine transformation result

3.2 Detection Model

Tempered glass detection experiments were conducted using various YOLO series models (YOLOv8 [27], YOLOv9 [25], YOLOv10 [26], YOLO11 [27]) by applying bounding box annotations directly to images processed through a Video Transformation module. Among the tested models, YOLO11 demonstrated the best performance and was selected as the primary focus of the structural characteristics of YOLOv8 and YOLO11.

YOLOv8, released by Ultralytics in January 2023, improves upon YOLOv5 [40] by offering higher accuracy and faster speed. This is achieved through the integration of three key modules: Convolution with Batch Normalization and Swish (CBS), CSP Bottleneck with two convolutions-Fast (C2F), and Spatial Pyramid Pooling-Fast (SPPF). The CBS module performs basic feature extraction using standard convolution, batch normalization, and SiLU activation [41]. The C2F module enhances the C3 structure of YOLOv5 with a parallel-branch architecture. One branch preserves the original features, while the other extracts multilevel features through repeated bottleneck modules, thereby improving boundary detection accuracy. The SPPF module fuses information from different spatial scales, allowing the model to recognize both fine details and broader context within an image. The neck section of YOLOv8 combines Feature Pyramid Network and Path Aggregation Network to enable features to be transmitted in both directions-both from lower to higher layers and *vice versa*. This bidirectional feature transmission helps the model integrate detailed and contextual information, improving detection across various object sizes. The head section is divided into multiple heads based on functionality. The detection head uses a loss function combining Complete Intersection over Union and Distribution Focal Loss, whereas the classification head employs Varifocal Loss. In addition, YOLOv8 uses an anchor-free head, which predicts object locations directly without relying on predefined reference boxes. This approach simplifies the detection process and can improve accuracy, especially for small or irregularly shaped objects.

YOLO11, released by Ultralytics in September 2024, introduces architectural refinements for an enhanced speed-accuracy balance. The backbone replaces C2F modules of YOLOv8 with C3k2 blocks, thereby improving the low-to-high-level feature integration efficiency. The novel Convolutional Block with Parallel Spatial Attention (C2PSA) module enhances the object boundary recognition accuracy through parallel spatial attention mechanisms. The neck section incorporates these C2PSA modules to achieve improved spatial attention with fewer parameters, thereby accelerating inference speeds. The head is explicitly optimized for latency reduction in the final prediction layers, while maintaining detection precision.

Two datasets were constructed to train the object detection model. The first dataset was generated by applying the Video Transformation module, whereas the second dataset was created by applying Affine transformation only to the original images. Both datasets were annotated with identical bounding box labels using the VGG Image Annotator. Dataset splitting was performed in the same manner for both datasets. These two datasets were then utilized to evaluate the effectiveness of the Video Transformation module and comparatively analyze the performance of the object detection model.

4 Experiments

4.1 Details

The dataset used for glass detection experiments in this study consists of images captured at a factory that produces glass. During production, the glass moves from the left to the right along a conveyor system, as illustrated in Fig. 6.

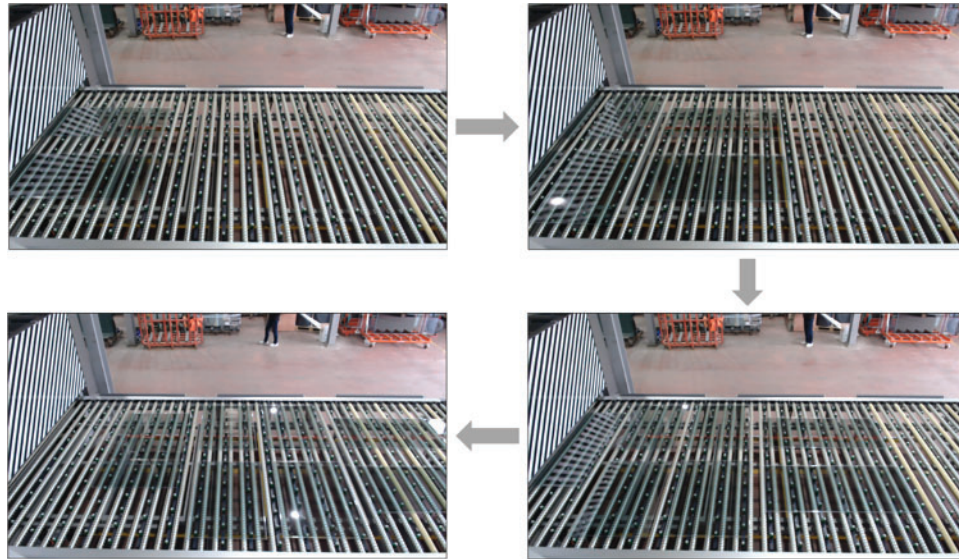


Figure 6: Glass manufacturing process

To construct the background using a GMM, only 40 images of factory facilities without any glass were utilized. During model training, the number of Gaussian distributions was set to 15 and the variance threshold was set to 50. The dataset for the YOLO model training consisted of 230 images, excluding those used for GMM training. This dataset was divided into training (40%), validation (10%), and testing (50%) sets comprising 92, 23, and 115 images, respectively. The training and validation data were used exclusively during the training process and were excluded from the testing set. Additionally, grayscale images processed with the Video Transformation module were employed to enhance the object detection performance, and the input channel of the YOLO model was modified to a single channel accordingly. To train and evaluate the glass detection model in this study, a system equipped with an Intel Core i9 CPU, an NVIDIA GeForce RTX 3060 GPU, and two 32 GB RAM modules was utilized.

4.2 Results

The results emphasizing the glassy regions in the input images using the Video Transformation module are shown in Fig. 7.

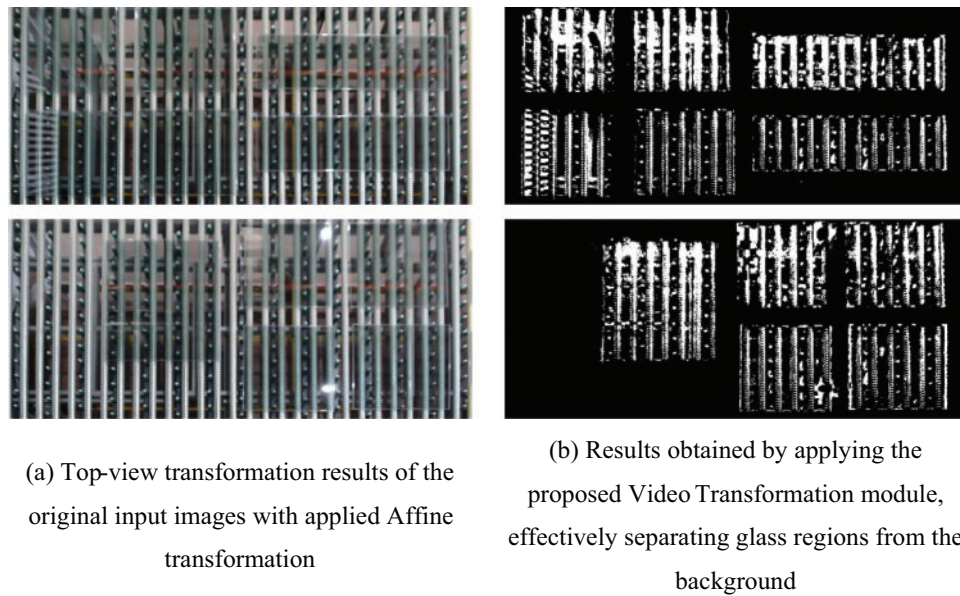


Figure 7: Results of applying the Video Transformation module, highlighting glass regions in the input images

To compare the outcomes of Video Transformation, Fig. 7a shows the top-view transformation results of the original images with the applied Affine transformation, and Fig. 7b illustrates the results obtained by applying the proposed Video Transformation module. The experimental results confirm that the proposed module effectively separates glass regions from complex glass images. In particular, the GMM-based background removal technique improves the robustness of the model by reducing the complex background noise caused by the conveyor rollers.

In glass-production environments, resistance to lighting variations is a critical factor in detection models. Owing to the refractive and reflective properties of glass, there is a high likelihood of false detections, and glass-detection models in real production environments require high robustness. Background learning using GMM is an effective approach for meeting these requirements. The adaptive background learning mechanism based on the GMM was introduced in this study to ensure consistent background removal performance against various external factors, such as time-dependent lighting changes, seasonal environmental variations, and equipment rearrangements in production lines. This mechanism standardized the quality of input images for object detection models.

A major limitation of previous studies on glass detection was that semantic segmentation and object detection models often misdetected objects behind the glass because of the transparent nature of the glass projecting onto the background. This study effectively addressed this issue by leveraging the static background characteristics of conveyor systems in glass production lines. The proposed method hierarchically combined the background removal and object detection stages to effectively separate moving glasses from stationary backgrounds. This approach offers a novel solution to the challenging problem of detecting transparent objects using deep-learning-based object detection. Notably, the integrated framework of dynamic background removal and static object detection provides a methodological foundation for designing moving-object monitoring systems in industrial environments.

In this experiment, a dataset emphasizing glass regions was constructed using a Video Transformation module to comprehensively evaluate the tempered glass detection performance of the object detection

models. Considering the computational resource constraints in embedded systems, comparative experiments were conducted targeting lightweight architectures (nano/tiny and small) of the YOLOv8, YOLOv9, YOLOv10, and YOLO11 models. The tempered glass detection results for each model are listed in Table 2.

Table 2: Performance comparison of YOLO models for glass detection

Model	Parameters	GFLOPs	Precision	Recall	F1-score	mAP50	mAP50-95	Inference (msec)
YOLOv8 Nano	2,690,115	6.9	1.0000	0.5154	0.6802	0.7577	0.7450	3.6387
YOLOv8 Small	9,838,771	23.5	1.0000	0.9015	0.9482	0.9507	0.9388	5.2110
YOLOv9 Tiny	1,764,835	6.6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
YOLOv9 Small	6,314,739	22.6	1.0000	0.4015	0.5730	0.7008	0.6920	7.4215
YOLOv10 Nano	2,707,142	8.3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
YOLOv10 Small	8,066,550	24.7	0.9902	0.9713	0.9806	0.9896	0.9645	5.2874
YOLO11 Nano	2,589,747	6.4	1.0000	0.9672	0.9833	0.9833	0.9681	3.8255
YOLO11 Small	9,427,603	21.4	0.9996	0.9826	0.9910	0.9949	0.9635	8.6456

As shown in Table 2, the YOLO11 small model achieves the highest performance, with an F1-score of 99.10% and mAP50 of 99.49%. The nano-version also demonstrates outstanding performance among the lightweight models with an F1-score of 98.33%. Although the YOLOv8 small model exhibits decent performance with an F1-score of 94.82%, its nano version exhibits a significant performance drop with an F1-score of 68.02%. Notably, YOLOv9 tiny and YOLOv10 nano-models fail to detect tempered glass entirely, with a 0% detection success rate.

All experimental models exhibit high precision, ranging from 99.02% to 100%; however, their recall values vary significantly. For example, YOLOv8 Nano achieves a recall rate of only 51.54%, missing nearly half of the actual glass regions, whereas YOLO11 Nano achieves a recall rate of 96.72%, demonstrating an excellent detection sensitivity despite its lightweight architecture. These results indicate that the improved feature extraction mechanism of YOLO11 effectively handles complex glass patterns. In terms of model complexity, YOLO11 Nano consumes only 6.4 GFLOPs and has 2.59M parameters, making it one of the most resource-efficient models in its class. Additionally, the inference speed of 3.83 ms is recorded, which is 2.26 times higher than that of small models. Fig. 8 compares the FLOPs, parameters, and F1-scores across different models.

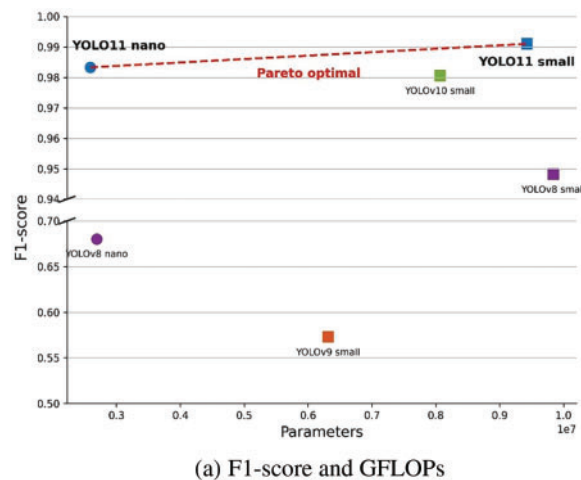
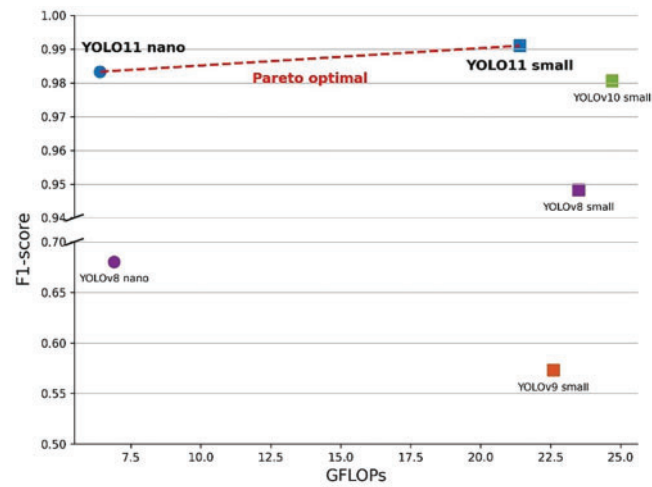


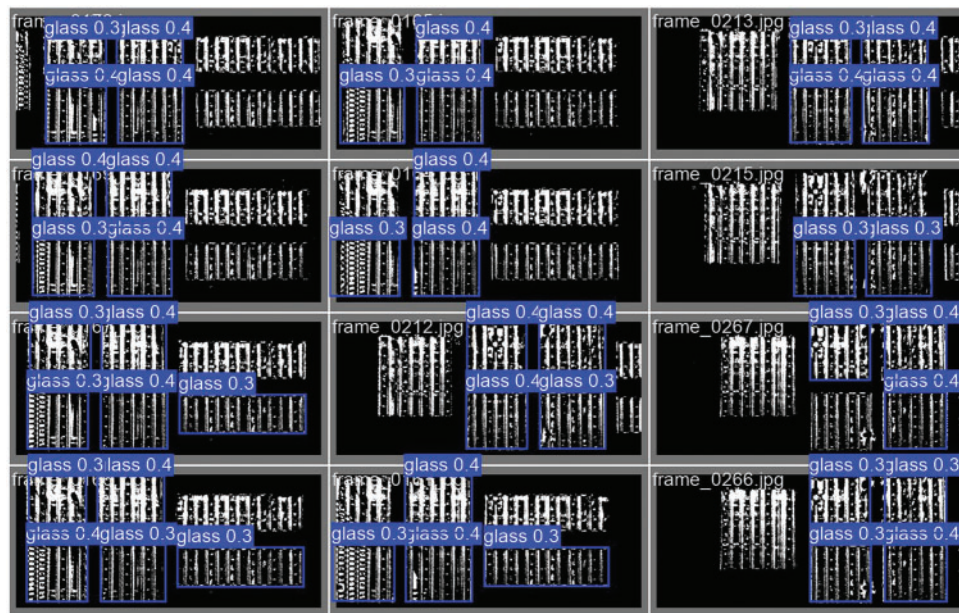
Figure 8: (Continued)



(b) F1-score and parameters

Figure 8: Performance-efficiency trade-off analysis

As shown in Fig. 8, based on the performance efficiency relative to computational resources and Pareto optimization analysis, YOLO11 Nano is determined to be the most suitable for embedded applications, achieving an F1-score of 98.33% and inference speed of 3.83 ms simultaneously. Accordingly, this study adopted the YOLO11 Nano as the standard model for experiments and ablation studies, confirming that it meets both real-time processing requirements and high detection reliability. Fig. 9 shows the tempered glass detection results for the YOLOv8 Nano and YOLO11 Nano models.



(a) YOLOv8 nano

Figure 9: (Continued)

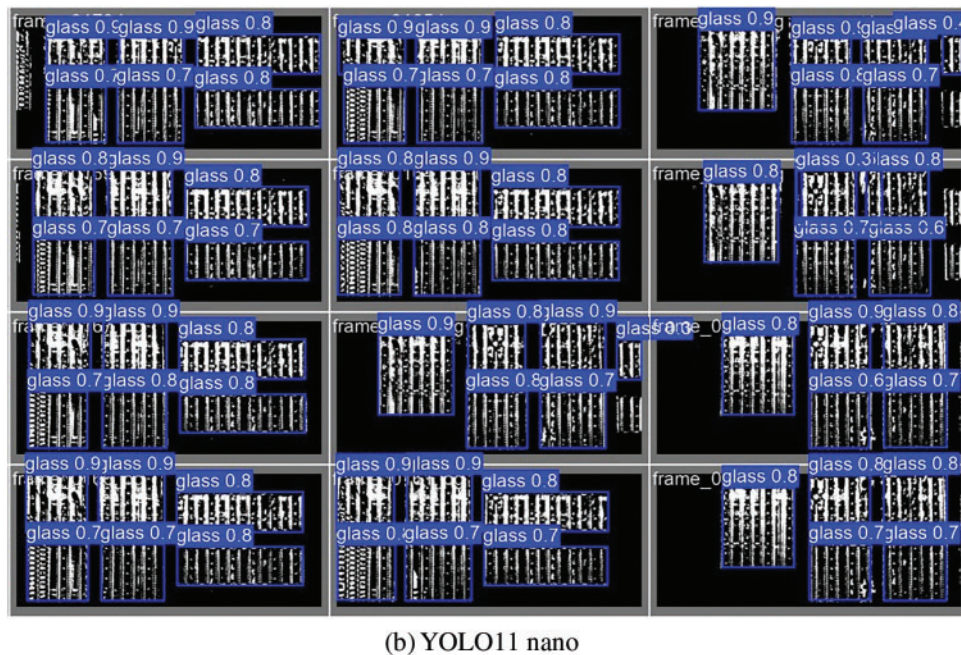


Figure 9: Glass detection results using YOLOv8 and YOLO11

The qualitative comparison shown in Fig. 9 demonstrates that while YOLOv8 Nano frequently fails to detect glass regions, YOLO11 Nano successfully detects most central glass regions despite missing some peripheral areas.

The performance evaluation of the developed object detection model showed that the precision reached 100%, indicating that no false positives occurred. In other words, the model did not mistakenly classify non-glass objects as glass. The recall was 96.72%, which is relatively high; however, some glass objects were not detected, resulting in false negatives. As analyzed in Fig. 10, these false negatives mainly occurred when glass objects were positioned at the left or right boundaries of the image, so that only a portion of the object was visible. This phenomenon is attributed to the model's inability to reliably recognize glass objects when their features are only partially captured. An effective approach to address this issue is to restrict the detection area to the central region of the image. Since all glass objects eventually pass through the center of the conveyor belt, this method can help prevent missed detections caused by partial occlusion at the image boundaries.

4.3 Ablations

Systematic ablation experiments were conducted to analyze the impact of Video Transformation module components (GMM, HSV transformation, and contour detection) on the tempered glass detection performance. YOLO11 Nano was adopted as the standard model; Table 3 presents the performance variations based on different combinations of these components.

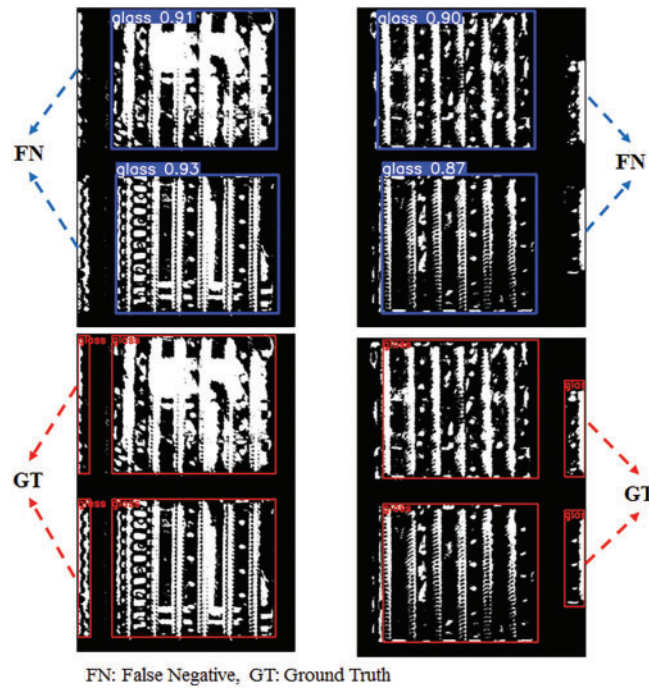


Figure 10: Analysis of model prediction for False Negative case

Table 3: Ablation study results of video transformation components

GMM	HSV	Contour	Affine	Precision	Recall	F1-score	mAP50	mAP50-95	Inference (msec)
			✓	1.0000	0.8822	0.9374	0.9410	0.9049	7.0286
✓			✓	1.0000	0.9421	0.9702	0.9709	0.9593	4.4155
✓	✓		✓	1.0000	0.9440	0.9712	0.9718	0.9570	3.8425
✓		✓	✓	1.0000	0.9691	0.9843	0.9844	0.9669	4.0481
	✓		✓	1.0000	0.8417	0.9140	0.9208	0.8720	3.8446
	✓	✓	✓	1.0000	0.9054	0.9504	0.9526	0.8848	4.0389
✓	✓	✓	✓	1.0000	0.9768	0.9883	0.9880	0.9727	3.8255

As shown in Table 3, under the condition (GMM $\times$, HSV $\times$, Contour $\times$, Affine $\checkmark$), the model achieves a recall of 88.22% and an F1-score of 93.74%, which represents the lowest performance compared to other combinations. When GMM-based background removal is added, recall improves by 5.99% to 94.21%, whereas inference speed is enhanced by 37.18%, achieving 4.42 ms. Under the condition excluding HSV transformation (GMM $\checkmark$, HSV $\times$, Contour $\checkmark$), recall reaches 96.91% and F1-score improves to 98.43%, indicating improved performance compared to the GMM-only condition. This suggests that the structural feature extraction components (GMM and contour detection) play a dominant role in driving performance improvements. When all components are applied (GMM $\checkmark$, HSV $\checkmark$, Contour $\checkmark$, Affine $\checkmark$), the model achieves the best performance with a recall of 97.68% and an F1-score of 98.83%. HSV transformation and contour detection contribute to an additional recall improvement of 3.47% compared to using the GMM alone; this can be interpreted as a synergistic effect between background removal and boundary enhancement. Table 4 summarizes the results of the ablation experiments.

Table 4: Summary of the ablation study results

Combinations	Recall	F1-score	Inference (msec)
Baseline (applying only Affine transformation)	0.8822	0.9374	7.0286
+GMM	0.9421 (+5.99%p)	0.9702 (+3.28%p)	4.4155 (−37.18%)
+GMM+HSV	0.9440 (+6.18%p)	0.9712 (+3.38%p)	3.8425 (−45.33%)
+GMM+Contour	0.9691 (+8.69%p)	0.9843 (+4.69%p)	4.0481 (−42.41%)
+GMM+HSV+Contour	0.9768 (+9.46%p)	0.9883 (+5.09%p)	3.8255 (−45.57%)

When all transformation techniques are applied, the model achieves a 45.57% higher inference speed and a 5.09% improvement in the F1-score compared to the baseline condition. These experiments demonstrate that GMM-based background removal contributes to both inference speed enhancement and detection sensitivity improvement, confirming that the hierarchical integration of multi-image processing techniques is an effective strategy for overcoming the limitations in object detection model performance. Furthermore, the system demonstrates strong adaptability to real-world manufacturing environments, where conditions such as lighting and conveyor layout can change over time. In particular, the GMM background model is constructed using the first 40 frames captured before the glass enters the field of view, during a stable idle phase. This enables automatic reconstruction of the background for each production cycle, ensuring that the model remains robust to environmental changes without requiring manual recalibration. The robustness of tempered glass detection is directly reflected in the confidence distribution of the reliability of the model. The proposed Video Transformation module further enhances system robustness by normalizing the input space—mitigating effects of camera position shifts, lighting variations, and background inconsistencies. This module ensures consistent input to the YOLO11 Nano model, making retraining unnecessary even as external conditions change. The proposed Video Transformation module can enhance the inference speed of YOLO11 Nano by utilizing classical computer vision algorithms, which can easily be implemented on edge devices. YOLO11 Nano, being a lightweight model, is well-suited for deployment on edge devices where computational resources are typically limited. As it is designed for high-speed inference while maintaining accuracy, it can be easily implemented on edge devices such as industrial cameras or embedded systems, even in environments with constrained processing power. Therefore, the proposed algorithm can be effectively deployed on real-world edge devices and provide practical benefits in factory environments, enabling real-time, non-contact quality inspection with minimal hardware requirements.

[Fig. 11](#) compares the recall-confidence curves between the models applying only Affine transformation and those applying the full Video Transformation module.

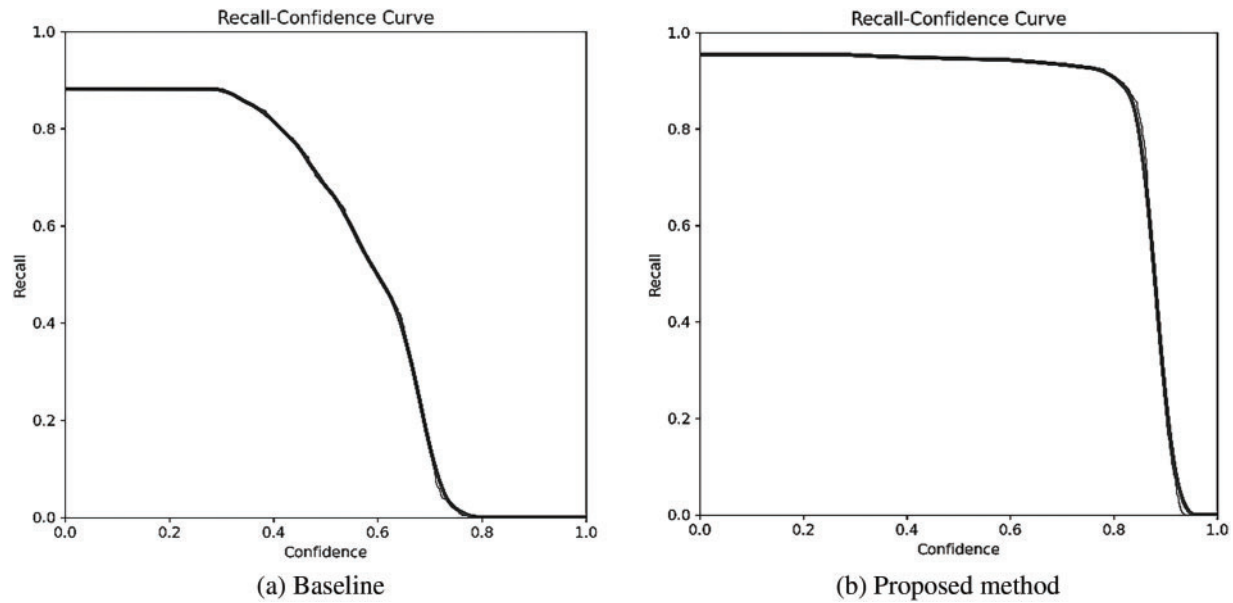


Figure 11: Comparison of Recall-Confidence curves

When only Affine transformation is applied, the recall drops sharply beyond a confidence score of 0.3, with most glass-object detections failing at confidence scores above 0.7. In contrast, the proposed Video Transformation module maintains a stable recall of more than 95%, up to a confidence score of 0.85. Notably, significant performance differences between the two approaches are observed in the high-confidence range of 0.8–0.9, where applying Video Transformation minimizes recall degradation until reaching a confidence threshold close to 0.9. This indicates that the proposed image transformation techniques enhance glass-feature extraction capabilities, thereby simultaneously improving both the reliability and stability of the detection results. In terms of confidence distribution, Video Transformation extends the maximum confidence range from approximately 0.7 to above 0.9 compared with simple Affine transformation, demonstrating its ability to significantly reduce false detections in practical applications. This improvement serves as a critical factor in ensuring robust tempered-glass detection performance, even under limited computational resource conditions in embedded environments.

5 Conclusions

This paper proposes a vision-based system for detecting and identifying defects in tempered glass in smart glass-manufacturing environments. The proposed system integrated a Video Transformation module that combined GMM-based background removal, HSV transformation, contour detection, and Affine transformation with an object detection model to effectively detect tempered glass, even in complex factory settings. The experimental results demonstrated that the approach combining the YOLO11 Nano model with the Video Transformation module achieved a recall of 97.68%, an F1-score of 98.83%, and a high inference speed of 3.83 ms. The proposed system offered higher accuracy, reliability, and faster processing than traditional manual inspection methods, enabling non-contact real-time quality inspection and data-driven decision-making support. Additionally, ablation experiments confirmed that GMM and contour detection played a critical role in improving the glass detection performance, highlighting the significance of effectively merging classical image processing techniques with state-of-the-art deep learning technologies. Future research should aim to enhance the background learning mechanism of GMM and improve its adaptability

to diverse lighting conditions. Furthermore, we plan to expand the usability of the system by integrating a high-resolution image analysis module for the precise identification of the detected glass regions. In addition, although the proposed algorithm was developed for use in tempered glass manufacturing, it can also be applied to other industrial environments that involve panel-shaped products conveyed via rollers, such as display or metal sheet production lines. We also plan to expand the dataset to include more diverse conditions through both additional data acquisition and the application of video augmentation techniques, such as random brightness adjustment, motion blur simulation, noise injection, geometric distortions (e.g., rotation and scaling), and background replacement. These techniques will help simulate various real-world factory environments and edge-case scenarios. In this context, generative AI models such as GANs may also be used to create synthetic defect samples or simulate rare environmental conditions. In parallel, we plan to adopt encoder-decoder-based architectures to enhance the segmentation and reconstruction of defect regions, allowing for more detailed analysis and interpretable results. Through these advancements, the system is expected to contribute further to automating quality control and improving productivity in the tempered glass manufacturing industry.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the Technology Development Program (RS-2024-00445393) funded by the Ministry of SMEs and Startups (MSS, Republic of Korea).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Seungmin Lee and Heesung Lee; software, Seungmin Lee; validation, Seungmin Lee and Heesung Lee; formal analysis, Seungmin Lee, Beomseong Kim and Heesung Lee; investigation, Beomseong Kim and Heesung Lee; resources, Beomseong Kim and Heesung Lee; data curation, Seungmin Lee and Heesung Lee; writing—original draft preparation, Seungmin Lee and Heesung Lee; writing—review and editing, Beomseong Kim and Heesung Lee; visualization, Seungmin Lee; supervision, Heesung Lee; project administration, Heesung Lee; funding acquisition, Heesung Lee. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Heesung Lee, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Shrouf F, Ordieres J, Miragliotta G. Smart factories in Industry 4.0: a review of the concept and of energy management approached in production based on the Internet of Things paradigm. In: 2014 IEEE International Conference on Industrial Engineering and Engineering Management; 2014 Dec 9–12; Selangor, Malaysia. Piscataway (NJ): IEEE; 2014. p. 697–701. doi:10.1109/IEEM.2014.7058728.
2. Soori M, Arezoo B, Dastres R. Digital twin for smart manufacturing, a review. *Sustain Manufac Service Econom.* 2023;2:100017. doi:10.1016/j.smse.2023.100017.
3. Kang HS, Lee JY, Choi SS, Kim H, Park JH, Son JY, et al. Smart manufacturing: past research, present findings, and future directions. *Int J Prec Eng Manuf-green Technol.* 2016;3(1):111–28. doi:10.1007/s40684-016-0015-5.
4. Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. *J Manuf Syst.* 2018;48(25):157–69. doi:10.1016/j.jmsy.2018.01.006.
5. Sundaram S, Zeid A. Artificial intelligence-based smart quality inspection for manufacturing. *Micromachines.* 2023;14(3):570. doi:10.3390/mi14030570.

6. Chouchene A, Carvalho A, Lima TM, Charrua-Santos F, Osório GJ, Barhoumi W. Artificial intelligence for product quality inspection toward smart industries: quality control of vehicle non-conformities. In: 2020 9th International Conference on Industrial Technology and Management (ICITM); 2020; Oxford, UK. Piscataway (NJ): IEEE. p. 127–31. doi:10.1109/ICITM48982.2020.9080396.
7. Yatsuk V, Kurylyak N. Quality monitoring during the installation of large tempered glass structures. *Meas Equip Metro.* 2023;84(4):39–43. doi:10.23939/istcmtm2023.04.039.
8. To QD, He QC, Cossavella M, Morcant K, Panait A, Yvonnet J. The tempering of glass and the failure of tempered glass plates with pin-loaded joints: modelling and simulation. *Mater Des.* 2008;29(5):943–51. doi:10.1016/j.matdes.2007.03.022.
9. Barsom JM. Fracture of tempered glass. *J Am Ceram Soc.* 1968;51(2):75–8. doi:10.1111/j.1151-2916.1968.tb11840.x.
10. McMaster RA. Fundamentals of tempered glass. In: *Proceedings of the 49th Conference on Glass Problems: Ceramic Engineering and Science Proceedings*; 1989; Hoboken, NJ, USA: John Wiley & Sons, Inc. p. 193–206. doi:10.1002/9780470310533.ch9.
11. Zaccaria M, Overend M. Nondestructive safety evaluation of thermally tempered glass. *J Mater Civ Eng.* 2020;32(4):04020043. doi:10.1061/(ASCE)MT.1943-5533.0003086.
12. Khoudi A, Barka N, Masrour T, El-Hassani I, Mazgualdi CE. Online prediction of automotive tempered glass quality using machine learning. *Int J Adv Manuf Technol.* 2023;125(3):1577–1602. doi:10.1007/s00170-022-10649-7.
13. Weiqi Z, Dejun C, Huixiong R, Hao Y, Ying L, Jieheng W. A method for surface defect detection of tempered glass based on polarization characteristics and unsupervised learning. In: *China Automation Congress (CAC)*; 2024; Qingdao, China: China Automation Congress (CAC). p. 3715–9.
14. Yue G, Wu H, Liu H, Li Y, Dun X. Online and real-time accurate prediction of tempered glass surface stress during quenching process. *J Therm Stress.* 2019;42(7):914–27. doi:10.1080/01495739.2019.1590169.
15. Mei H, Yang X, Wang Y, Liu Y, He S, Zhang Q, et al. Don't Hit Me! glass detection in real-world scenes. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020; Piscataway (NJ): IEEE; 2020. p. 3687–96. doi:10.1109/CVPR42600.2020.00374.
16. Zheng C, Shi D, Yan X, Liang D, Wei M, Yang X, et al. GlassNet: label decoupling-based three-stream neural network for robust image glass detection. *Comput Graph Forum.* 2022;41:377–88. doi:10.1111/cgf.14441.
17. Yan T, Gao J, Xu K, Zhu X, Huang H, Li H, et al. GhostingNet: a novel approach for glass surface detection with ghosting cues. *IEEE Trans Pattern Anal Mach Intell.* 2025;47(1):323–37. doi:10.1109/TPAMI.2024.3463490.
18. Han D, Lee S, Zhang C, Yoon H, Kwon H, Kim HC, et al. Internal-external boundary attention fusion for glass surface segmentation. *arXiv:2307.00212.* 2023. doi:10.48550/arXiv.2307.00212.
19. Bebis G, Georgiopoulos M, da Vitoria Lobo N, Shah M. Learning affine transformations. *Patt Recogn.* 1999;32(10):1783–99. doi:10.1016/S0031-3203(98)00178-2.
20. Xue X, Zhang K, Tan KC, Feng L, Wang J, Chen G, et al. Affine transformation-enhanced multifactorial optimization for heterogeneous problems. *IEEE Trans Cyber.* 2020;52(7):6217–31. doi:10.1109/tcyb.2020.3036393.
21. Choi H, Bajić IV. Affine transformation-based deep frame prediction. *IEEE Trans Image Process.* 2021;30:3321–34. doi:10.1109/tip.2021.3060803.
22. Reynolds D. Gaussian mixture models. *Encyclo Biomet.* 2009;741(1):659–63. doi:10.1007/978-0-387-73003-5_196.
23. Stock JH, Wright JH. GMM with weak identification. *Econometrica.* 2000;68(5):1055–96. doi:10.1111/1468-0262.00151.
24. Rasmussen C. The infinite Gaussian mixture model. In: *Advances in neural information processing systems*. Vol. 12; 1999.
25. Wang CY, Yeh IH, Liao HY. YOLOv9: learning what you want to learn using programmable gradient information. In: *European Conference on Computer Vision*; 2025; Cham: Springer. p. 1–15. doi:10.1007/978-3-031-72751-1_1.
26. Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. Yolov10: real-time end-to-end object detection. *Adv Neural Inform Process Syst.* 2024;37:107984–8011.
27. Jocher G, Qiu J. Ultralytics YOLO11. Version 11.0.0. 2024. [cited 2025 May 15]. Available from: <https://github.com/ultralytics/ultralytics>.

28. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023; Vancouver, BC, Canada: IEEE; 2023. p. 7464–75. doi:10.1109/cvpr52729.2023.00721.
29. Lai P, Fuh C. Transparent object detection using regions with convolutional neural network. Vol. 2. In: IPPR Conference on Computer Vision, Graphics, and Image Processing; 2015; Yilan, Taiwan.
30. Pan Z, Yang J, Wang XE, Wang F, Azim I, Wang C. Image-based surface scratch detection on architectural glass panels using deep learning approach. Constr Build Mater. 2021;282(8):122717. doi:10.1016/j.conbuildmat.2021.122717.
31. Ngo HQT. Design of automated system for online inspection using the convolutional neural network (CNN) technique in the image processing approach. Results Eng. 2023;19(16):101346. doi:10.1016/j.rineng.2023.101346.
32. Majchrowska S, Mikołajczyk A, Ferlin M, Klawikowska Z, Plantykowski MA, Kwasigroch A, et al. Deep learning-based waste detection in natural and urban environments. Waste Manag. 2022;138(6):274–84. doi:10.1016/j.wasman.2021.12.001.
33. Ghasemi Y, Jeong H, Choi SH, Park KB, Lee JY. Deep learning-based object detection in augmented reality: a systematic review. Comput Ind. 2022;139(12):103661. doi:10.1016/j.compind.2022.103661.
34. Park J, Riaz H, Kim H, Kim J. Advanced cover glass defect detection and classification based on multi-DNN model. Manuf Lett. 2020;23(2):53–61. doi:10.1016/j.mfglet.2019.12.006.
35. Claypo N, Jaiyen S, Hanskunatai A. Inspection system for glass bottle defect classification based on deep neural network. Int J Adv Comput Sci Appl. 2023;14(7): doi:10.14569/IJACSA.2023.0140738.
36. Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. IEEE Trans Pattern Anal Mach Intell. 2010;33(5):898–916. doi:10.1109/TPAMI.2010.161.
37. Zhao B, Wen X, Han K. Learning semi-supervised gaussian mixture models for generalized category discovery. In: IEEE/CVF International Conference on Computer Vision; 2023 Oct 26; Paris, France. Piscataway: IEEE; 2023. p. 16624–34. doi:10.48550/arXiv.2305.06144.
38. Nwobi FN, Kalu SI. Application of EM algorithm in classification problem & parameter estimation of gaussian mixture model. Int J Stat Appl. 2023;13(1):20–6. doi:10.5923/j.statistics.20231301.03.
39. Bouzouad M, Benhamadi Y, Slimani C, Boukhobza J. Adapting gaussian mixture model training to Embedded/Edge devices: a Low I/O, deadline-aware and energy efficient design. ACM SIGAPP Appl Comput Review. 2024;24(2):5–18. doi:10.1145/3687251.3687252.
40. Jani M, Fayyad J, Al-Younes Y, Najjaran H. Model compression methods for YOLOv5: a review. arXiv:2307.11904. 2023. doi:10.48550/arxiv.2307.11904.
41. Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: comparison of trends in practice and research for deep learning. arXiv:1811.03378. 2018. doi:10.48550/arxiv.1811.03378.