Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.12 n.58, p.1170-1178. March/April, 2026.
DOI: https://doi.org/10.5935/jetia.v12i58.3281

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

INTEGRATING BOTTLENECK ATTENTION MODULE (BAM) INTO YOLOV8 FOR AUTOMATED INDUSTRIAL LABEL QUALITY INSPECTION

Habib Ja'far Nuur[1], M. Munadi*[2] and Mochammad Ariyanto[3]

[1, 2, 3]Department of Mechanical Engineering, Diponegoro University, Semarang, Central Java, Indonesia.

[1]https://orcid.org/0009-0002-8948-7339, [2]http://orcid.org/0000-0002-0766-0955, [3]http://orcid.org/0000-0001-8125-6742

Email: habibjafar08@gmail.com, *munadi@ft.undip.ac.id, mochammad_ariyanto@ft.undip.ac.id

ARTICLE INFO

Article History
Received: January 10, 2026
Reviewed: February 12, 2026
Accepted: March 24, 2026
Published: April 30, 2026

Keywords:
Quality control,
YOLOv8,
Attention mechanism,
Expired date,
Labeling defects.

ABSTRACT

This study proposes an automated inspection solution to address the vulnerability of manual Quality Control (QC) systems to human error in detecting expired date labeling defects on plastic bottle products used for liquid packaging. The developed system employs the single-stage YOLOv8 architecture for object detection, which offers high inference speed, a crucial aspect for real-time applications. This study enhances model accuracy through the integration of a Bottleneck Attention Module (BAM) as an attention mechanism, strategically placed at the 9th layer of the network backbone. The selection of BAM is based on its capability to simultaneously capture channel dependencies and spatial relationships, which is essential for accurately recognizing subtle printing defect patterns in small-sized text. As a result, the enhanced model (YOLOv8-BAM) demonstrates a significant improvement in key performance metrics compared to the baseline YOLOv8 model, namely: an increase in Recall of 4.17%, Precision of 3.23%, and F1-score of 3.16%. These findings validate that YOLOv8-BAM is a more robust and reliable solution for automated industrial label quality inspection.

I. INTRODUCTION

Amid the rapid transformation of Industry 4.0, the manufacturing sector is required to adopt Quality Control (QC) systems that are not only efficient but also capable of ensuring a high level of accuracy. In the context of consumer goods production, particularly those utilizing bottle packaging for liquid products, the integrity and clarity of labeling are fundamental aspects. Labels, such as the expiration date, serve as a guarantee of consumer safety and security. However, ironically, the QC process for inspecting these critical labels is still largely performed manually, relying entirely on human visual perception. This conventional approach is inherently prone to human error, which often allows defective products to pass through the production line. This vulnerability is further exacerbated by operator fatigue and noisy working environment conditions, which are primary triggers for decreased inspection accuracy [1]. The consequences of such human error are highly significant for consumers, as unclear labeling can pose health risks due to products being consumed beyond their safe limits, while for companies, the impacts include substantial financial losses and reputational damage resulting from the erosion of consumer trust in product quality.

Therefore, innovations are required to replace manual visual inspection with methods that are more objective and reliable.The rapid advancement of deep learning now offers a promising solution for enhancing QC processes in labeling defect detection. The application of deep learning in visual inspection has been proven successful in various cases; for instance, by [2] utilized a CNN combined with a persistent spectrum to identify surface cracks in rotating systems. Since the emergence of [3], Convolutional Neural Network (CNN) architectures have become the backbone of defect detection tasks. Current object detection algorithms are generally classified into two main categories. Two-stage algorithms, such as R-CNN [4], Faster R-CNN [5], and Mask R-CNN [6], divide the detection process into two stages, achieving high accuracy and strong adaptability to object scale variations. In contrast, single-stage algorithms such as, according to [7], [8] simplify the problem into a single regression, simultaneously predicting object classes and locations.

The advantage of these single-stage methods lies in their high inference speed and compact architecture [9], making them well suited for applications requiring real-time processing. Overall, deep learning–based approaches have been proven to be more accurate and adaptive than conventional spectrum-based or statistical image analysis methods. To address the demands of fast-paced QC processes and minimize human error, this study proposes a YOLOv8 architecture as an automated system for detecting expiration date labeling defects. The selection of YOLOv8 is based on its characteristics as a single-shot detector that offers high detection speed, considering that the inspection process is conducted on a continuously moving conveyor. Therefore, the primary focus of this research is to enhance the robustness of the YOLOv8 model through the integration of an attention module. By incorporating the Bottleneck Attention Module (BAM), it is expected that the model's capability to extract relevant visual features can be sharpened, thereby resulting in an accurate, high-performing labeling defect detection system that is ready for industrial implementation.

II. RELATED WORKS

There are numerous approaches that can be employed to perform expired date labeling defect detection, ranging from rule-based models, Optical Character Recognition (OCR), machine learning, and other methods that form the foundation of computer vision. In [10] proposed a quality control method for zinc electrodes by extracting image data into numerical representations through image processing techniques, which were then processed using traditional machine learning methods. According to [11] employed various machine learning models (such as MLP, SVM, k-NN, RF, XGBoost, and LightGBM) with color-based inputs. Color extraction was performed using ImageJ, resulting in features across the R, G, B, L, a, b, H, S, and V color channels. These conventional machine learning–based methods are effective in certain cases; however, they face significant challenges when dealing with large-scale image data, variations in lighting conditions, or complex and diverse visual defects.

The limitations of traditional machine learning and rule-based models that operate on predefined rules have driven a transition toward deep learning algorithms. These algorithms possess an intrinsic capability to automatically extract complex features and identify defect patterns from massive volumes of data [12]. The training process of deep learning models, ranging from data preprocessing to feature extraction and inference, demands high computational power. This has led to the necessity of utilizing Graphics Processing Units (GPUs) and high-performance computing systems to process data rapidly and efficiently. This study focuses on optimizing deep learning architectures, particularly the backbone component responsible for feature extraction, with the objective of improving the accuracy of complex target detection. A number of optimization studies have been conducted on existing object detection architectures. Jing [13] enhanced YOLOv3 by applying the K-means algorithm and adding a new detection layer on different feature maps.

This modification aimed to achieve better information fusion between the upper and lower layers of the feature maps, which was effective in reducing false detection cases. In another study, by [14] modified the YOLO11 architecture by integrating SPStem, ADown, SEDFF, and WFPIoU, enabling more effective edge extraction, stronger feature representation, and improved detection accuracy for small and occluded tomatoes without increasing computational cost. According to [15] improved YOLOv8n by integrating the GAM mechanism into the backbone, replacing the neck with CCFM, and adopting the DyHead head, which synergistically enhanced tomato fruit stem detection accuracy while maintaining a lightweight characteristic suitable for edge hardware deployment. According to [16] replaced traditional square convolutions with deformable convolutions to improve R-CNN, a method that allows the model to learn feature representations adaptively, thereby increasing robustness to small objects and complex edges.

Two-stage detection algorithms such as Faster R-CNN, although highly accurate, tend to suffer from slower processing speed and higher complexity [17]. The substantial computational resources and time required make these algorithms less optimal for real-time scenarios and speed-demanding applications. In the context of expired date labeling defect detection in industrial environments, the inspection process is performed on fast-moving conveyors, making detection speed a critical factor for implementation. Considering that manual inspection and rule-based/OCR models have been shown to be vulnerable to human error or incorrect predictions when dealing with complex defects, the need for a fast and accurate one-stage model becomes urgent. Based on these real-time processing considerations, the YOLOv8 model is proposed in this study, as it belongs to the category of one-stage models that excel in detection speed.

III. EXPERIMENT METHOD

III.1 YOLOv8 STRUCTURE

In this study, YOLOv8 is proposed as the model for detecting expired date labeling defects in the quality control process. Building upon the success of previous YOLO versions, YOLOv8 introduces improvements and additional new features. Compared to traditional two-stage methods, YOLOv8 offers higher detection speed. YOLOv8 detects and fuses feature maps at different levels, enabling the algorithm to simultaneously detect targets at multiple scales, as well as providing better robustness and accuracy in detecting complex targets [18], such as expired date labeling defects. Compared to previous object detection algorithms, YOLOv8 integrates the advantages of earlier approaches and designs a new backbone network, C2F, along with enhanced detection heads and loss functions to improve the performance and flexibility of the algorithm [19]. The standard YOLOv8 network structure shown in Figure 1 consists of three main components the backbone, neck, and head.

In the backbone, the network begins with several Convolution (Conv) and C2f layers that function to progressively extract features from the input image, where the C2f module is used to improve gradient flow and feature representation efficiency. At the end of the backbone, the SPPF (Spatial Pyramid Pooling Fast) module is employed to expand the receptive field, enabling the model to capture contextual information at multiple scales. Subsequently, the neck section utilizes a combination of upsampling and concatenation operations to fuse features from different resolution levels (feature fusion), allowing high-level semantic information and fine spatial details to be effectively integrated. This process is further reinforced with C2f and Conv blocks to refine the fused features. Finally, in the head section, these multi-scale features are fed into multiple detection layers, enabling YOLOv8 to perform object detection at various scales simultaneously, thereby improving detection accuracy for small, medium, and large objects.

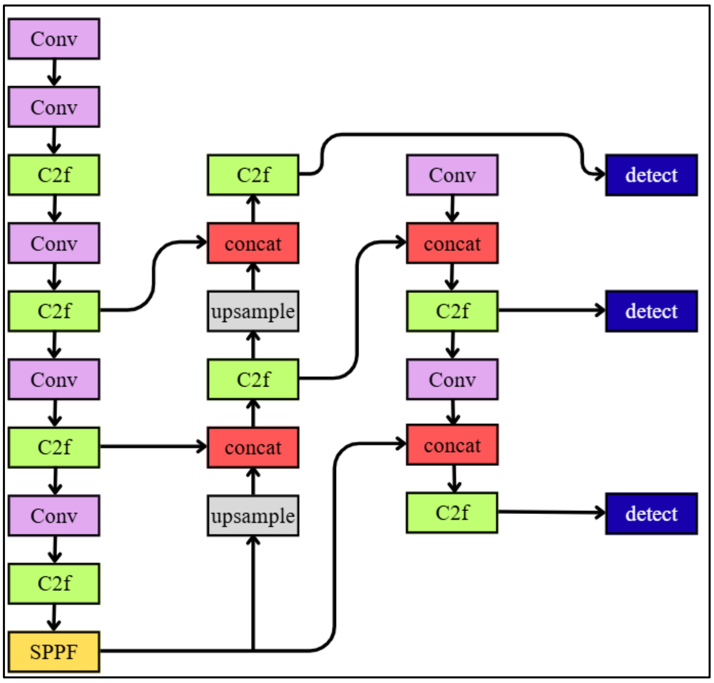

Figure 1: YOLOv8 network structure diagram.
Source: [20].

III.2 BAM ATTENTION MECHANISM

Attention mechanisms are techniques that have been widely used in the field of deep learning to improve model performance by focusing computational resources on the critical parts of the input data. This study focuses on detecting expired date labeling defects, which fall into the category of small-sized objects with visual characteristics that are difficult to identify at a global scale. Although the standard YOLOv8 architecture is equipped with a multi-scale fusion mechanism capable of integrating feature maps from various resolution levels, this study integrates the Bottleneck Attention Module (BAM) to further refine the feature extraction capability of the convolutional neural network. The selection of BAM is based on its ability to enhance feature representation without significantly disrupting the core network structure [21]. In its implementation, the BAM module is strategically placed after a convolutional layer while preserving the original number of network channels to maintain computational efficiency. The architectural configuration of the BAM attention mechanism is visually illustrated in Figure 2.

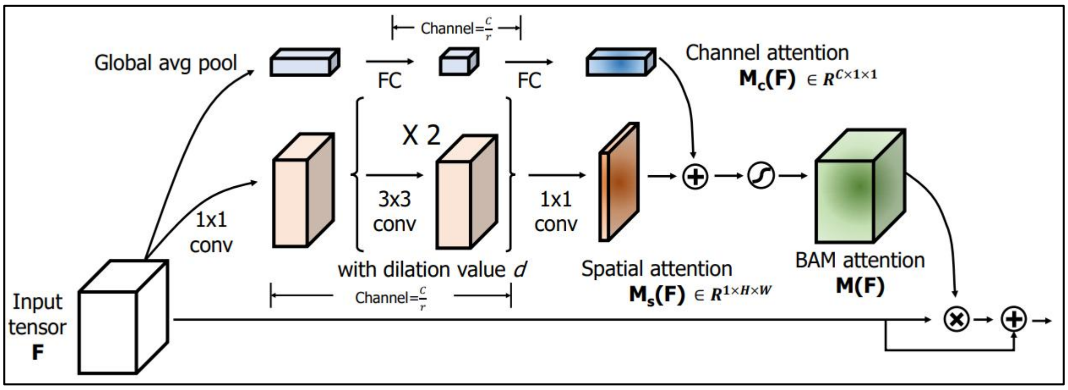

Figure 2: BAM attention mechanism diagram.
Source: [22].

BAM operates by generating an attention map that acts as an automatic information selector within a conventional neural network. The process begins by receiving an input feature map F , which is then processed through two parallel pathways to extract salient characteristics. The final output of this module is a refined feature map, where the attention values $M(F)$ are multiplied element-wise with the original features before being added back to the initial input. This residual learning technique ensures that gradients can flow effectively while simultaneously reinforcing features that are relevant for object detection or classification, as shown in Equation 1.

$$Output\ Feature\ Map \quad F' = F + F \otimes M(F) \tag{1}$$

Mathematically, the attention map $M(F)$ is not generated as a single entity, but rather through the combination of two different perspectives. The first component is channel attention $M_{C(F)}$ and the second component is spatial attention $M_{S(F)}$, which are summed before passing through an activation function. The use of the sigmoid function σ at the final stage aims to map the attention values into the range of 0 to 1. Within this range, the module can function as a “gate” that determines the intensity of information allowed to be propagated to the subsequent layers in the model architecture, as shown in Equation 2.

$$M(F) = \sigma\left(M_{C(F)} + M_{S(F)}\right) \tag{2}$$

The two attention components within BAM have specific and complementary roles in understanding the content of an image. Channel Attention $M_{C(F)}$ focuses on the channel dimension to determine "what" features are the most influential, such as specific textures or colors. On the other hand, Spatial Attention $M_{S(F)}$ employs dilated convolution to understand "where" these important locations are geographically positioned within the feature map. The combination of both allows the model to assign greater weight to defect regions or target objects while suppressing irrelevant background noise.

III.3 PROPOSED YOLOV8 FRAMEWORK ENHANCED WITH ATTENTION MECHANISM

The optimization of the YOLOv8 architecture was performed to improve detection accuracy for expired date labeling defects, which are generally categorized as small-sized objects with subtle and complex visual features. This approach was implemented through the selective integration of an attention mechanism into the backbone of the network to strengthen the model's capability to extract relevant features. In this study, several commonly used attention modules, namely CBAM (Convolutional Block Attention Module) and BAM (Bottleneck Attention Module), were selected as attention mechanism schemes to be experimentally evaluated. Both modules are designed to enhance feature representation by considering channel and spatial information, thereby potentially supporting the detection of small text and numerical characters on product labels. The implementation of the attention mechanism was focused on strategic placement at one of the YOLOv8 backbone layers, specifically the 9th layer, which was selected based on an analysis of feature map characteristics at this stage that were considered most effective for multi-scale feature processing and enhancement prior to the feature fusion stage [15]. This approach is relevant for handling variations in expiration date marking defects, such as printing defects, blurring effects, and incomplete characters, all of which demand a high level of precision during identification.

With this strategy, the core YOLOv8 architecture is preserved, while feature extraction capability is enhanced through the utilization of an attention mechanism that strengthens the representation of regions of interest (ROIs) containing expired date text and suppresses the influence of noise and complex packaging backgrounds. The structure of the YOLOv8 network modified with the integration of the attention mechanism is shown in Figure 3, with the aim of improving the network's ability to emphasize important features and reduce less relevant information. The addition of the attention mechanism (AM) enables the model to adaptively weight both spatial and channel features, resulting in more selective and informative feature representations being forwarded to the SPPF module. The subsequent effect of this process is an improvement in the quality of features entering the neck section, which involves upsampling and concatenation operations for multi-scale fusion. The fused features become semantically richer and more focused on object regions, allowing information integration across resolution levels to be performed more effectively. Finally, in the head section, the features enhanced by the attention mechanism provide higher-quality inputs to the detection layers, thereby contributing to improved detection accuracy, particularly for small or overlapping objects, without altering the main architectural flow or significantly increasing computational complexity.

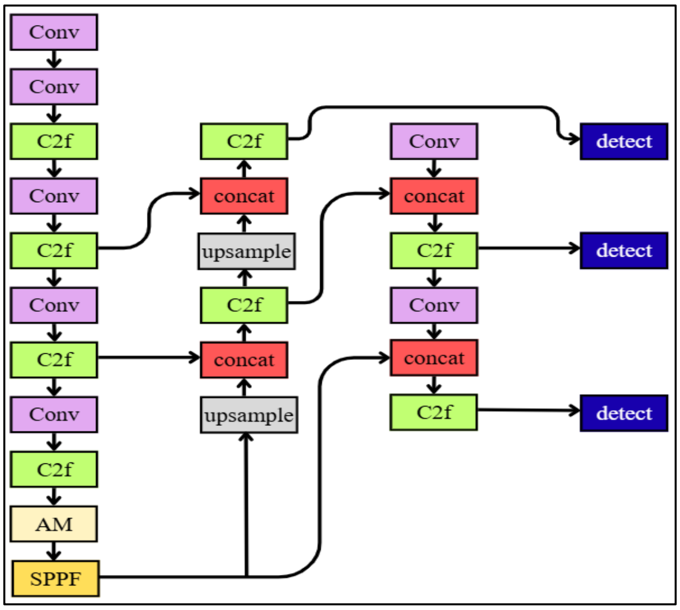

Figure 3: Proposed YOLOv8 network structure diagram.
Source: Authors, (2026).

III.4 CONSTRUCTION OF DATASET

Products that have entered the market should have passed the quality control stage, particularly in terms of expired date labeling. However, due to the influence of human error, products with defective and unreadable expired date labels may still pass through and reach the market. Therefore, the dataset samples were obtained from products already distributed in the market that exhibited expired date labeling defects. The selection of a specific brand aims to train the model to recognize consistent visual characteristics of a single product type, allowing the model to focus on detecting labeling anomalies. The total dataset used in this study consists of 312 images captured using an industrial camera, with 162 images representing products with labeling defects and 150 images representing products that passed quality inspection.

To enhance model generalization, image acquisition was conducted under varying lighting conditions and capture distances. The entire dataset was then randomly split with an 80:20 ratio, where 80% was allocated for training and 20% for testing. Prior to the training process, each image in the dataset underwent an annotation stage. This process involved assigning class labels (“defect” or “Pass”) and creating bounding boxes to identify the location and extent of defects on the products. Accurate annotation is crucial to effectively guide the model during the learning process. The annotation process is illustrated in Figure 4, where Figure 4a presents an example of annotation for products that pass quality inspection, while Figure 4b shows the annotation for products in defective (reject) condition.

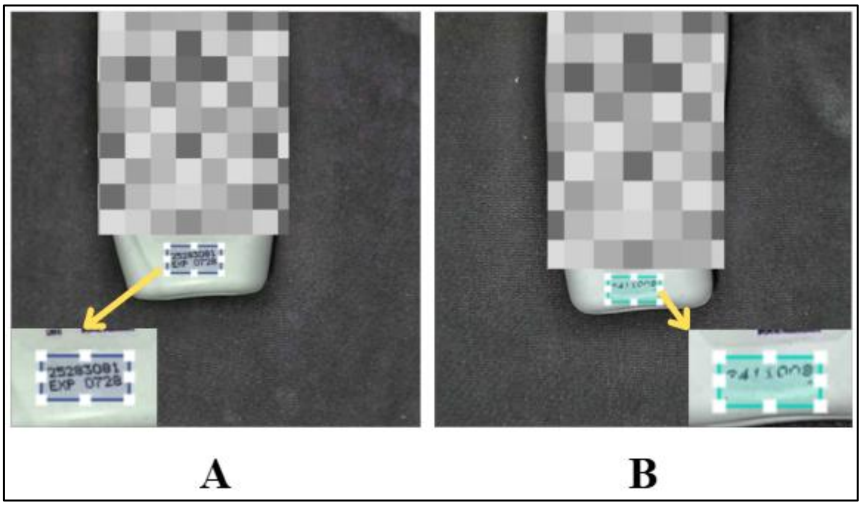

Figure 4: Annotation and labeling dataset.
Source: Authors, (2026).

III.5 PREPROCESSING DATASET

In this study, object classification is divided into two main categories: pass and defect. The “defect” category includes various forms of nonconformity, such as blurred printing, incomplete print results, or partially erased text, which will subsequently be rejected during the quality control process. The dataset was augmented by applying Gaussian Blur and Random Noise methods, which simulate disturbances caused by varying or insufficient lighting conditions that result in high noise levels commonly encountered on production lines. Subsequently, all images were normalized by resizing the pixel dimensions to 300×300. This dimensional standardization ensures input uniformity required by the YOLOv8 model architecture while reducing pixel size to alleviate computational burden. In addition to manual augmentation, YOLOv8 also has the capability to perform internal augmentation during training through Mosaic and Mixup techniques. In Mosaic augmentation, four randomly selected training images are combined into a single frame, enabling the model to learn object recognition in more complex contexts and across diverse scales [23]. Meanwhile, Mixup linearly combines two images along with their labels, creating synthetic samples that can enhance model robustness [24]. The application of both techniques has been proven effective in increasing data diversity and reducing overfitting.

III.6 EXPERIMENT ENVIRONMENT AND PARAMETER SETTING

The experiments in this study were conducted locally by utilizing a computational environment designed to support efficient, stable, and well-controlled model training processes. The use of consistent hardware and software configurations constitutes an essential aspect in ensuring the reliability of the experimental procedures and minimizing result variability caused by differences in execution environments. All experimental environment settings employed in this study are presented in Table 1.

Table 1: Environment configuration.

Name		Name	
Video Cards	RTX 3090 Ti	Video memory	24 GB
PyTorch	2.2.1	Python	3.10
Processor	AMD Ryzen 9 7950X	Cuda	11.8

Source: Authors, (2026).

In addition to the computational environment, the selection of model training parameters was carried out systematically to support an optimal learning process. These parameter configurations play a critical role in controlling training dynamics, model convergence, and the stability of the optimization process. Detailed information regarding the parameters used during model training and evaluation is presented in Table 2.

Table 2: Training parameters.

Name		Name	
Learning rate	0.01	Weight decay	0.0005
Batch Size	16	Momentum	0.9
Image size	640 x 640	Number of iterations	208
Optimizer	Adam	Patience	100
Dropout	0.0	Nms	False

Source: Authors, (2025).

III.7 EXPERIMENT MODEL EVALUATION CRITERIA

The objective performance assessment of all trained models was conducted through quantitative evaluation using several complementary evaluation metrics to comprehensively describe the strengths and weaknesses of each model. The evaluation criteria employed include Precision, Recall, F1-score, as well as Average Precision (AP) and mean Average Precision (mAP). Precision represents the model's accuracy in classifying the positive class and is calculated based on the proportion of positive predictions that correspond to the actual condition, where a high precision value indicates a low false positive rate, as shown in Equation (3).

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Recall describes the model's ability to identify all instances that truly belong to the positive class, where a high recall value indicates that the model is capable of minimizing false negative errors and effectively capturing the majority of existing positive cases, as shown in Equation (4).

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

To balance these two metrics, the F1-score is employed, which represents the harmonic mean of precision and recall. A high F1-score value indicates that the model demonstrates stable and optimal performance in maintaining prediction accuracy while simultaneously preserving its ability to detect the positive class, as shown in Equation (5).

$$F1 = \frac{2\,x\,(Precision\ x\ Recall)}{(Precision+Recall)} \tag{5}$$

In addition to these classification metrics, object detection performance is also evaluated using Average Precision (AP), which is a metric obtained by integrating the Precision–Recall curve for a single object class across different confidence thresholds. Subsequently, mean Average Precision (mAP) is calculated as the average AP value across all classes used in this study, namely the defect and pass classes. The mAP@0.5 value indicates detection performance at an Intersection over Union (IoU) threshold of 0.5, while mAP@0.5:0.95 represents the average mAP over a range of IoU thresholds from 0.5 to 0.95, thereby providing a more stringent and comprehensive evaluation of detection quality.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

IV.1 COMPARE ANALYSIS OF TARGET DETECTION ALGORITHMS

In this study, a comparative analysis was conducted using three YOLO-based models, all trained and tested on the same dataset. This comparative approach is essential for objectively evaluating the benefits of the proposed enhancement (YOLOv8-BAM) relative to the baseline architecture. The performance comparison of each YOLO model version is presented in Table 3.

Table 3: Comparison of target detection algorithms.

Experiment Algorithms	Precision (%)	Recall (%)	F1 (%)	mAP:0.5 (%)	mAP:0.5-0.9 (%)
YOLOv5	82	93	87	99,3	67,4
YOLOv6	90	96	93	98,8	67,1
YOLOv8	93	96	95	99,5	69

Source: Authors, (2026).

The experimental results indicate that the YOLOv8 model achieves the best overall performance compared to the other YOLO architecture variants evaluated in this study. Specifically, the YOLOv8 baseline attains a precision of 93%, recall of 96%, an F1-score of 95%, mean Average Precision (mAP@0.5) of 99.5%, and mAP@0.5:0.95 of 70%. This robust performance demonstrates the suitability of YOLOv8 as a base architecture for defect detection tasks. The YOLOv6 model follows behind in performance, while YOLOv5 ranks lowest among all compared models. Based on these experimental results, YOLOv8 was selected as the baseline model for further development through architectural modifications with the integration of an attention mechanism, aiming to further enhance detection capability, particularly for small-sized objects and fine-grained defects.

IV.2 ANALYSIS COMPARISON OF ATTENTION MECHANISM

A comparison between attention mechanisms was conducted to highlight the superiority of the Bottleneck Attention Module (BAM) over the Convolutional Block Attention Module (CBAM) within the YOLOv8 architecture. In this experiment, CBAM was integrated at the same location as BAM, namely the 9th layer of the backbone network. CBAM combines channel attention and spatial attention sequentially, adaptively refining feature maps by adjusting the importance of different channels and the significance of different spatial locations. However, CBAM requires separate computations for channel and spatial attention and necessitates feature map alignment across multiple scales. This sequential computation inherently results in higher computational complexity, which can negatively impact overall model performance and inference speed. The relative performance of these different attention mechanism integrations is detailed in Table 4.

Table 4: Performance comparison of attention mechanisms integrated into YOLOv8.

Attention mechanism module	Precision (%)	Recall (%)	F1 (%)	mAP:0.5 (%)	mAP:0.5-0.9 (%)	Latency (ms)
CBAM	100	93	96	99,5	71,6	14,6
BAM	96	100	98	99,5	72,2	13,3

Source: Authors, (2026).

Based on the experimental comparison between the CBAM and BAM attention mechanisms, BAM demonstrates superior performance in the label defect detection task. Specifically, the integration of BAM results in an increase in Recall of 7.5%, an improvement in F1-Score of 2.08%, and an enhancement in detection accuracy (mAP@0.5:0.95) of 0.84%. Although Precision experiences a slight decrease of 4%, the substantial gains in Recall and F1-Score indicate a more balanced and robust detector, particularly in minimizing missed detections, which is a critical requirement in industrial quality control.

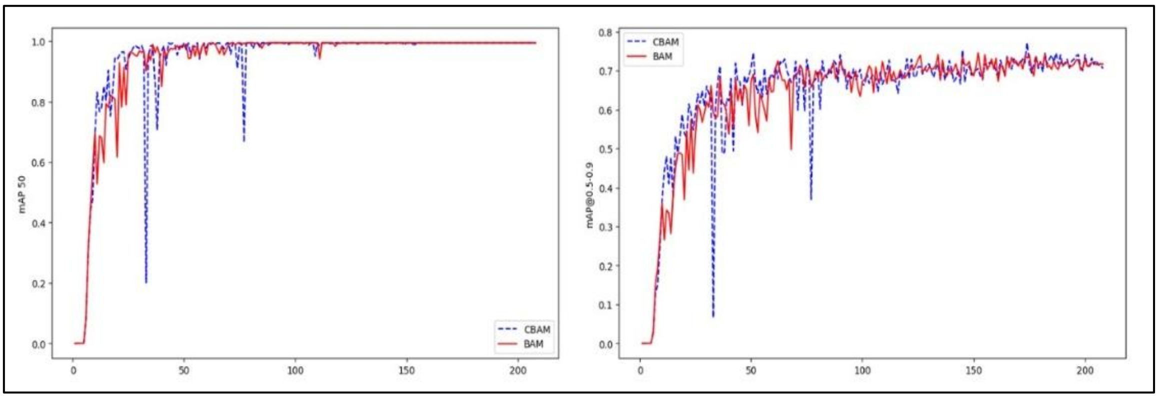

Figure 5: Curve mAP training experiment.

Source: Authors, (2026).

The implementation of the Bottleneck Attention Module (BAM) demonstrates lower inference latency compared to the implementation of the Convolutional Block Attention Module (CBAM). This characteristic is a crucial aspect, considering that defect detection on expired date labels requires high inference speed to enable the system to perform accurate detection and classification in real-time industrial environments. With more efficient latency, the YOLOv8-BAM model is considered more suitable for deployment on production lines that require high-speed visual inspection processes without compromising detection performance. In addition, Figure 5 illustrates a comparison of the model training results with the application of CBAM and BAM based on mAP performance curves over the number of epochs. The observations indicate that the model incorporating BAM achieves faster convergence and exhibits more stable curves compared to CBAM, particularly during the early to mid stages of the training process. The metric fluctuations in BAM tend to be smaller, indicating a more consistent learning process, while the CBAM-based model still experiences several relatively sharp performance drops at certain epochs. In the final stage of training, both models achieve high performance; however, BAM is able to maintain slightly better and more stable metric values. This demonstrates that the application of BAM is more effective in improving the quality of feature learning and contributes positively to the overall model performance. Therefore, based on the experimental results of the attention mechanisms, BAM was selected for integration into the YOLOv8 architecture. Subsequently, Figure 6 presents examples of defect detection results on expired date labels generated by the YOLOv8 model optimized with the Bottleneck Attention Module (BAM). Figure 6a illustrates the visualization of the model's prediction results for products that pass quality inspection, while Figure 6b shows the visualization of the model's prediction results for products in defective (reject) condition.

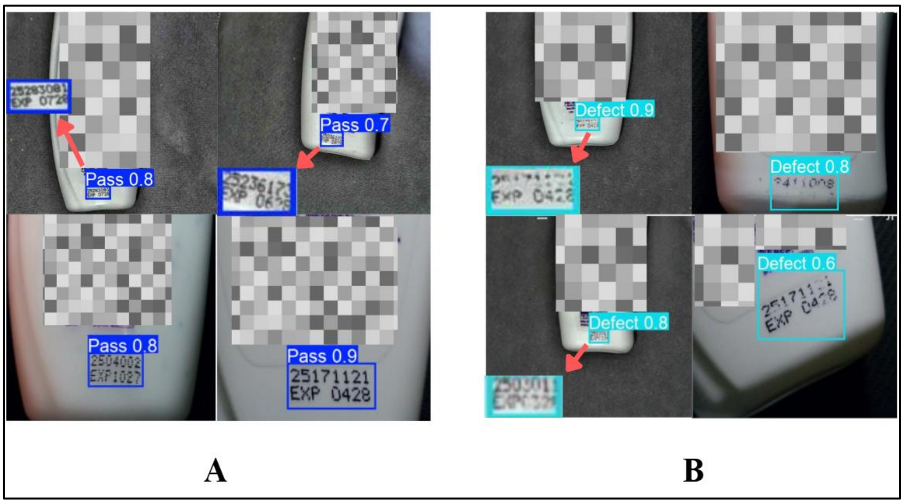

Figure 6: Visual sample result prediction model.

Source: Authors, (2026).

IV.3 CONTRIBUTION OF IMPROVED YOLOV8 ALGORITHM FOR DEFECT DETECTION LABELING EXPIRED DATE

The determination of whether an expiration date label is accepted or rejected depends on its readability for consumers. Traditional methods that rely on manual human inspection are highly susceptible to human error. In addition, the use of rule-based models is prone to detection errors when faced with complex defects and has difficulty adapting to new defect patterns, as classification depends entirely on explicit rules defined by developers. Furthermore, two-stage detection models are difficult to implement in industrial environments, where detection processes must be executed rapidly, which constitutes a major limitation of such architectures. To address these challenges, this study proposes the YOLOv8 model, a single-stage detector that offers high detection speed and the ability to adapt to evolving label defects without the need to explicitly define new rules within the model. For further optimization, the YOLOv8 model was enhanced through the integration of the Bottleneck Attention Module (BAM). This optimization significantly improves the model's performance in detecting and classifying defects in expiration date labels. YOLOv8-BAM achieves notable improvements in evaluation metrics, with Precision increasing by 3.23%, Recall by 4.17%, and F1-score by 3.16% compared to the standard YOLOv8 model. These enhanced results demonstrate that YOLOv8-BAM is a more robust, reliable, and industrially viable solution for automated expiration date label inspection.

IV.4 LIMITATION IN WORK

The construction of this dataset was carried out by collecting product samples that were already circulating in the market. This resulted in limited diversity and complexity of defects in the expired date labels. Consequently, for direct implementation in an active production environment, the model should be retrained with the inclusion of a more diverse dataset. In summary, this study focuses on the enhancement and optimization of the YOLOv8 model, which successfully improves overall model performance in detecting defects in expiration date labels.

V. CONCLUSIONS

This study successfully enhances the baseline YOLOv8 model through the integration of the Bottleneck Attention Module (BAM) attention mechanism, with the primary objective of improving the model's feature extraction capability. Based on experimental results evaluating the effectiveness of this architectural enhancement, the proposed YOLOv8-BAM model demonstrates consistent and significant performance advantages over the standard baseline YOLOv8 model. Testing results show improvements across all major evaluation criteria, with Precision increasing by 3.23%, Recall by 4.17%, and F1-score by 3.16%. These findings indicate that YOLOv8-BAM is not merely an incremental improvement, but rather a more robust, reliable, and industrially viable solution for automated visual inspection systems. Its enhanced ability to balance detection accuracy and prediction speed makes it particularly well suited for automated expired date labeling defect detection applications in industrial environments. The implementation of this model can reduce human error associated with manual inspection and provide an effective solution to the limitations of rule-based models.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Habib Ja'far Nuur.
Methodology: Habib Ja'far Nuur, Mochammad Ariyanto.
Investigation: Habib Ja'far Nuur, M. Munadi.
Discussion of results: Habib Ja'far Nuur, M. Munadi, Mochammad Ariyanto
Writing – Original Draft: Habib Ja'far Nuur.
Writing – Review and Editing: Habib Ja'far Nuur.
Resources: Habib Ja'far Nuur
Supervision: M. Munadi, Mochammad Ariyanto.
Approval of the final text: Habib Ja'far Nuur, M. Munadi, Mochammad Ariyanto.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and facilities provided by Diponegoro University. The laboratory access and technical support were indispensable to the successful execution of this research project. We deeply appreciate the University's enduring commitment to fostering research and innovation, which has been instrumental in making this work possible.

VIII. REFERENCES

[1] T. Badrick and A. S. A. Shane Brown, "Identifying human factors as a source of error in laboratory quality control," J. Lab. Precis. Med., vol. 8, pp. 6–11, 2023, doi: 10.21037/jlpm-23-7.

[2] N. Rezazadeh, M. Ashory, and S. Fallahy, "Identification of shallow cracks in rotating systems by utilizing convolutional neural networks and persistence spectrum under constant speed condition," pp. 135–147, 2021, doi:
10.21595/jmeacs.2021.22221.

[3] A. Krizhevsky and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," pp. 1–9.

[4] S. Kyung, H. Lee, and J. Kim, "Comparison of RNN-based deep learning models for tension force estimation for monitoring tension members using yoke-type elasto-magnetic sensors," Measurement, vol. 261, no. December 2025, p. 119939, 2026, doi: 10.1016/j.measurement.2025.119939.

[5] A. Amir, S. Giri, S. Giri, and M. Butt, "Smart Agricultural Technology An advanced approach to tomato apex head thickness measurement using lightweight YOLO variants , faster RCNN , and RGB-depth sensor," Smart Agric. Technol., vol. 12, no. July, p. 101214, 2025, doi: 10.1016/j.atech.2025.101214.

[6] R. Liu et al., “Journal of Radiation Research and Applied Sciences A comparative study of automatic hippocampal segmentation in whole-brain radiotherapy applying DPNU-Net , Mask-RCNN , and nnU-Net models,” J. Radiat. Res. Appl. Sci., vol. 18, no. 1, p. 101274, 2025, doi: 10.1016/j.jrras.2024.101274.

[7] A. Ahmed, A. Shariq, A. Manaf, Z. Kastrati, and S. Muhammad, “Biomedical Signal Processing and Control Enhancing wrist abnormality detection with YOLO : Analysis of state-of-the-art single-stage detection models,” Biomed. Signal Process. Control, vol. 93, no. March 2023, p. 106144, 2024, doi: 10.1016/j.bspc.2024.106144.

[8] V. P. Raj, K. Radha, S. H. Gopalan, C. Venkataramanan, R. Dhanapal, and A. Manikandan, “COA-SSD : A novel chaotic optimization algorithm for enhancement of real-time video surveillance images using single shot detector with MobileNetV2,” Syst. Soft Comput., vol. 7, no. October, p. 200417, 2025, doi: 10.1016/j.sasc.2025.200417.

[9] A. Vijayakumar et al., “Real - time visual intelligence for defect detection in pharmaceutical packaging,” Sci. Rep., pp. 1–23, 2024, doi: 10.1038/s41598-024-69701-z.

[10] L. Azoulay-Younes, A. Nyabadza, M. Vazquez, and D. Brabazon, “Zinc electrode manufacturing quality control with machine learning: Using SMOTE & image augmentation to prevent overfitting,” J. Eng. Res., no. January, 2025, doi: 10.1016/j.jer.2025.01.002.

[11] B. Çiftci, N. Çetin, S. Günaydın, and M. Kaplan, “Machine learning approaches for binary classification of sorghum (Sorghum bicolor L.) seeds from image color features,” J. Food Compos. Anal., vol. 140, no. September 2024, p. 107208, 2025, doi: 10.1016/j.jfca.2025.107208.

[12] M. Prunella, R. M. Scardigno, G. S. Member, D. Buongiorno, A. Brunetti, and N. Longo, “Deep Learning for Automatic Vision-Based Recognition of Industrial Surface Defects : A Survey,” IEEE Access, vol. 11, no. May, pp. 43370–43423, 2023, doi: 10.1109/ACCESS.2023.3271748.

[13] J. Jing, D. Zhuo, and H. Zhang, “Fabric defect detection using the improved YOLOv3 model,” 2020, doi: 10.1177/1558925020908268.

[14] X. Deng, T. Huang, W. Wang, and W. Feng, “SE-YOLO : A sobel-enhanced framework for high-accuracy , lightweight real-time tomato detection with edge deployment capability,” Comput. Electron. Agric., vol. 239, no. PB, p. 110973, 2025, doi: 10.1016/j.compag.2025.110973.

[15] W. Weng et al., “Smart Agricultural Technology GCD-YOLO : A deep learning network for accurate tomato fruit stalks identification in unstructured environments,” Smart Agric. Technol., vol. 12, no. July, p. 101465, 2025, doi: 10.1016/j.atech.2025.101465.

[16] L. Xue, Q. Li, Y. Lu, D. Zhang, Q. He, and H. Wang, “Fabric Defect Detection Based on the Improved Cascade R-CNN,” vol. 4, no. 7, pp. 81–87, 2021, doi: 10.25236/AJCIS.2021.040712.

[17] S. Y. Mohammed, “Franklin Open Architecture review : Two-stage and one-stage object detection,” Franklin Open, vol. 12, no. July, p. 100322, 2025, doi: 10.1016/j.fraope.2025.100322.

[18] H. Zhang et al., “GAT-Enhanced YOLOv8 _ L with Dilated Encoder for Multi-Scale Space Object Detection,” Remote Sens., pp. 1–24, 2025.

[19] Z. Xu, Y. Bao, and B. Tian, “Improved YOLOv8 garment sewing defect detection method based on attention mechanism,” pp. 706–721, 2024, doi: 10.21595/jme.2024.24283.

[20] G. Yao, “HP-YOLOv8 : High-Precision Small Object Detection Algorithm,” Sensors, 2024.

[21] J. Park, S. Woo, J. L. In, S. Kweon, and J. Lee, “A Simple and Light-Weight Attention Module for Convolutional Neural Networks,” Int. J. Comput. Vis., 2020, doi: 10.1007/s11263-019-01283-0.

[22] J. Park, “arXiv : 1807 . 06514v2 [cs . CV] 18 Jul 2018 BAM : Bottleneck Attention Module”.

[23] B. Zhao, Y. Wu, X. Guan, and L. Gao, “An Improved Aggregated-Mosaic Method for the Sparse Object Detection of Remote Sensing Imagery,” pp. 1–19, 2021.

[24] X. Liu, F. Shen, J. Zhao, and C. Nie, “RandoMix: a mixed sample data augmentation method with multiple mixed modes,” Multimed. Tools Appl., vol. 84, no. 8, pp. 4343–4359, 2025, doi: 10.1007/s11042-024-18868-8.