EDUMATIC Jurnal Pendidikan Informatika

e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Volume 10, Issue 2, **430-439.**
https://e-journal.hamzanwadi.ac.id/index.php/edumatic
DOI: 10.29408/edumatic.v10i2.36243

Coordinate Attention-Based Neck Feature Fusion for Enhanced YOLOv11 Oil Palm FFB Ripeness Detection

Reza Hanif Kurniawan [1], Sri Winarno [1,*]

[1] Universitas Dian Nuswantoro, Indonesia
[*] Corresponding author: Reza Hanif Kurniawan, Universitas Dian Nuswantoro, Indonesia
sri.winarno@dinus.ac.id

Received: 19 July 2026 | Revised: 24 July 2026 | Accepted: 18 August 2026 | Published: 26 August 2026

Abstract
Detecting the ripeness of oil palm fresh fruit bunches (FFB) directly on trees is complicated by factors such as partial visibility, occlusion, low contrast, image blur, and scale variation in plantation images. This study evaluated how the choice and positioning of attention mechanisms affect neck feature fusion in the lightweight YOLOv11n model. The baseline YOLOv11n was compared with SE-YOLOv11n, CBAM-YOLOv11n, and CA-YOLOv11n, all trained under the same conditions, with the attention mechanisms applied at the first and second up-sampling fusion features. Additionally, Coordinate Attention was tested in three different neck placement setups. CA-YOLOv11n recorded the highest recall, F1-score, mAP@50, and mAP@50–95, with values of 0.772, 0.773, 0.813, and 0.518, respectively, while SE-YOLOv11n achieved the highest precision at 0.794. Compared with the baseline, CA-YOLOv11n added 4,640 parameters, increased GFLOPs from 6.4 to 6.5, and decreased inference throughput from 15.1 to 14.7 FPS. The placement ablation revealed that integrating attention at the first and second fusion features yielded the highest F1-score and mAP values among the three placements tested, whereas extending CA to all three fusion features diminished the detection performance. These findings suggest that within the YOLOv11n configuration assessed, both the selection and placement of attention mechanisms impact the detection performance, and adding more attention insertions does not necessarily enhance the results.

Keywords: coordinate attention; neck feature fusion; oil palm ffb ripeness detection; yolov11

To cite this article: Kurniawan, R. H., & Winarno, S. (2026). Coordinate Attention-Based Neck Feature Fusion for Enhanced YOLOv11 Oil Palm FFB Ripeness Detection. *Edumatic: Jurnal Pendidikan Informatika*, *10*(2), 430–439. https://doi.org/10.29408/edumatic.v10i2.36243

INTRODUCTION

Determining the ripeness of oil palm fresh fruit bunches (FFB) accurately is crucial for determining the harvest time. The maturity of the fruit influences both the rate of oil extraction and the quality of palm oil (Lai et al., 2022). In practical plantation settings, the evaluation of FFB ripeness still relies heavily on visual assessments and the judgment of workers (Junior & Suharjito, 2023). FFBs on trees may be only partially visible, hidden by leaves and fronds, surrounded by thick vegetation, or viewed at varying apparent scales. Images from plantations may also suffer from uneven lighting, motion blur, backlighting, and low contrast (Li et al., 2025; Naftali et al., 2024; Suharjito et al., 2025). These factors can make it challenging to locate objects and differentiate between ripeness categories.

YOLO-based detectors achieve a practical equilibrium between detection performance and computational efficiency, making them suitable for real-time object detection applications, such as in the oil palm industry for fresh fruit bunch (FFB) detection, ripeness recognition, and fruit counting (Ali & Zhang, 2024; Junior & Suharjito, 2023; Lai et al., 2022; Naftali et al., 2024). Recent advancements in this field have enhanced detection capabilities by focusing on various pipeline components, including parameter-efficient architectural design, improved feature extraction, and training objectives addressing class imbalance (Li et al., 2025; Naftali et al., 2024; Suparto & Pribadi, 2026). These collective efforts demonstrate that performance enhancements can be achieved through modifications in computational efficiency, feature representation, and model optimization. Nonetheless, there is limited evidence regarding the impact of the placement of attention modules within neck feature fusion on detection performance. This unresolved architectural issue necessitates an investigation into attention placement as a controlled variable within a lightweight YOLOv11n detector.

Attention mechanisms offer a way to selectively adjust intermediate feature responses. Coordinate Attention (CA) processes features independently in the horizontal and vertical directions, preserving positional information within channel attention (Hou et al., 2021). In YOLO-based detectors, the neck combines feature maps from various resolution levels prior to making predictions (Ali & Zhang, 2024; Bian et al., 2025). Additionally, attention mechanisms have been integrated into architectures that fuse features across multiple and cross scales for object detection (Liu et al., 2025; Liu et al., 2022; Wu et al., 2024; Zhao et al., 2024). In the context of on-tree FFB detection, pertinent visual information might appear at different feature levels owing to the varying apparent scale of bunches, which may be only partially visible. Thus, the neck serves as a crucial stage for assessing whether the placement of attention affects detection performance during the fusion of multi-scale features.

Numerous studies pertinent to this research have examined attention in conjunction with other architectural changes. Liu et al. (2022) integrated CA with cross-scale weighted feature fusion in a fruit-detection framework. Similarly, Chen et al. (2025) incorporated the CBAM into a modified YOLOv11 framework alongside other architectural adjustments. Although these studies illustrate the application of attention in altered object detectors, they lack a controlled comparison in which the same attention mechanism is applied across various neck fusion positions, keeping the rest of the detector configuration constant. In the literature reviewed for this study, there is also a scarcity of evidence for such a comparison in YOLOv11n-based on-tree FFB ripeness detection. The research gap identified here pertains to the selection of attention under a fixed neck configuration, placement of attention across neck fusion positions, and resulting variations in detection performance and computational cost within a lightweight detector.

This study evaluates YOLOv11n for on-tree FFB ripeness detection under consistent experimental settings. The baseline model was compared with SE-YOLOv11n, CBAM-YOLOv11n, and CA-YOLOv11n using identical training conditions and the same neck integration configuration. This study provides a controlled comparison of the three attention mechanisms within this configuration. Furthermore, CA was evaluated across three neck placement configurations while keeping the underlying detector and training settings unchanged. The detection metrics are reported together with the parameter count, GFLOPs, model size, and inference speed to characterize the performance and computational differences among the evaluated configurations. The conclusions are restricted to the dataset, YOLOv11n architecture, and experimental settings examined in this study and do not establish a universally optimal attention placement for other datasets or detector architectures.

METHOD

This study used a quantitative experimental design to evaluate attention integration in YOLOv11n for on-tree oil palm fresh fruit bunch (FFB) ripeness detection. The experimental evaluation consisted of two stages. In the first stage, the baseline YOLOv11n was compared with SE-YOLOv11n, CBAM-YOLOv11n, and CA-YOLOv11n under identical training and evaluation settings. In the second stage, three coordinate attention placement configurations within the neck were examined to assess the effect of placement under the same detector configuration. The overall experimental workflow is illustrated in Figure 1.

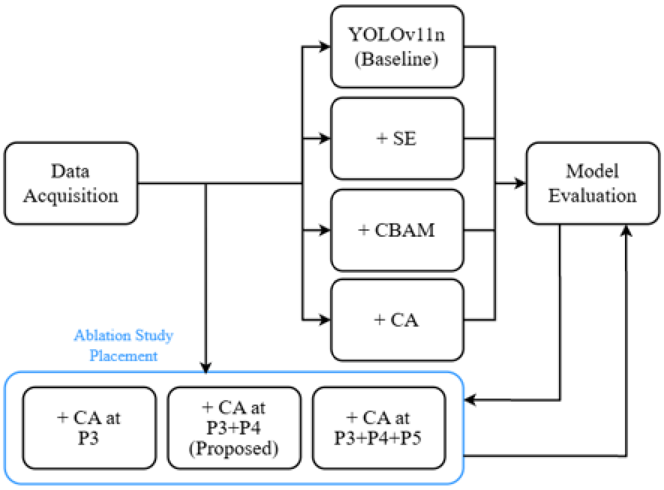


Figure 1. Research workflow

The Oil Palm Fruit Dataset on Plantations for Harvesting Estimation (version 3) was used in this study (Suharjito et al., 2024). The dataset was collected from commercial oil palm plantations in Central Kalimantan, Indonesia, and comprised five FFB categories: abnormal, flower, ripe, unripe, and unripe, with ripeness categories validated by palm oil experts. The experimental files contained 10,221 training, 1,583 validation, and 765 test images. The images present challenging conditions, including partial visibility, low contrast, occlusion, and image blur (Suharjito et al., 2025). As reported in Table 1, the class distribution is imbalanced, with unripe containing the largest number of instances, consistent with the dataset descriptor's explanation that ripe bunches are promptly harvested in plantation environments.

Table 1. Class-wise instance distribution

Class	Train	Validation	Test	Total
Abnormal	2,802	318	139	3,259
Flower	2,673	251	182	3,106
Ripe	2,669	443	195	3,307
Underripe	2,605	305	181	3,091
Unripe	6,439	1,789	834	9,062

YOLOv11n served as the foundational detector to assess the incorporation of attention mechanisms within a streamlined YOLO setup. Each model began with weights pre-trained on COCO and underwent training under the same experimental conditions. In line with the neck integration points identified by Chen et al. (2025), SE, CBAM, and CA were incorporated into the first and second upsampling fusion features of the YOLOv11n neck. This integration setup was uniformly applied across all three attention mechanisms, whereas the rest of the detector's architecture remained unchanged. Figure 2 depicts the resulting architecture and the positions of the attention integration.

The experimental assessment was divided into two phases. Initially, baseline YOLOv11n was compared with SE-YOLOv11n, CBAM-YOLOv11n, and CA-YOLOv11n under identical integration settings. Subsequently, CA placement ablation was conducted across the first up-sampling fusion feature (P3), the first and second up-sampling fusion features (P3 and P4), and all three fusion features (P3, P4, and P5). Each model was trained under the same conditions outlined in Table 2, utilizing a fixed random seed of 42 to minimize stochastic differences in comparisons. The experiments were executed on a Kaggle Notebook with an NVIDIA Tesla T4 GPU featuring 15 GB VRAM. The detection performance metrics included precision, recall, F1-score, mAP@50, and mAP@50–95, while computational efficiency was measured by parameter count, GFLOPs, model size, and inference throughput in frames per second (FPS).

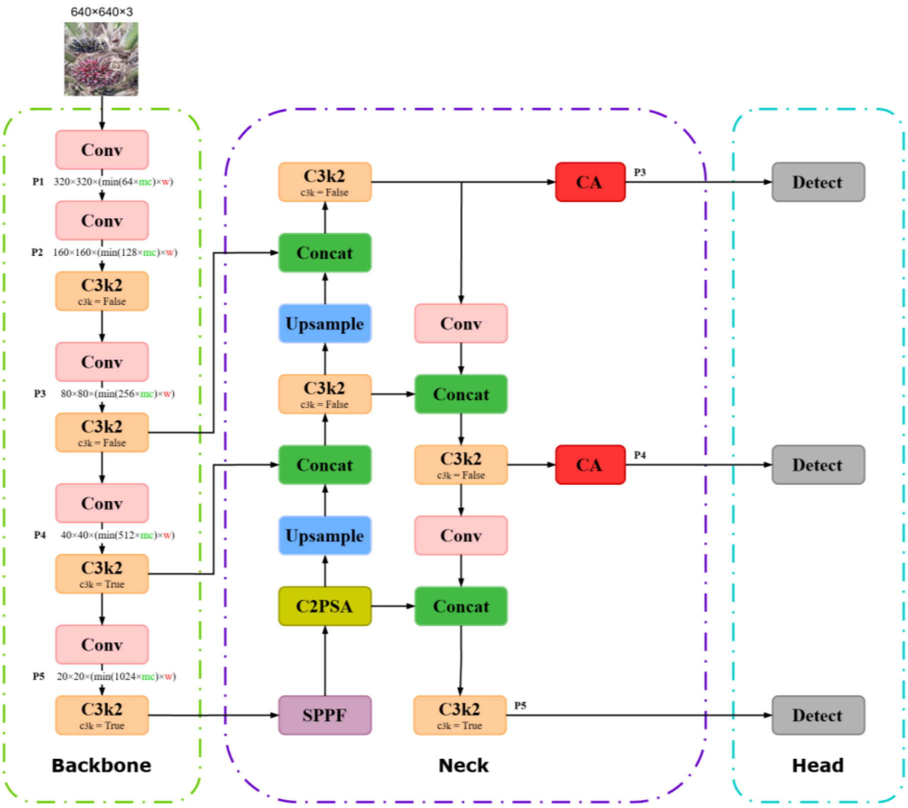


Figure 2. Coordinate Attention integration in the yolov11n neck

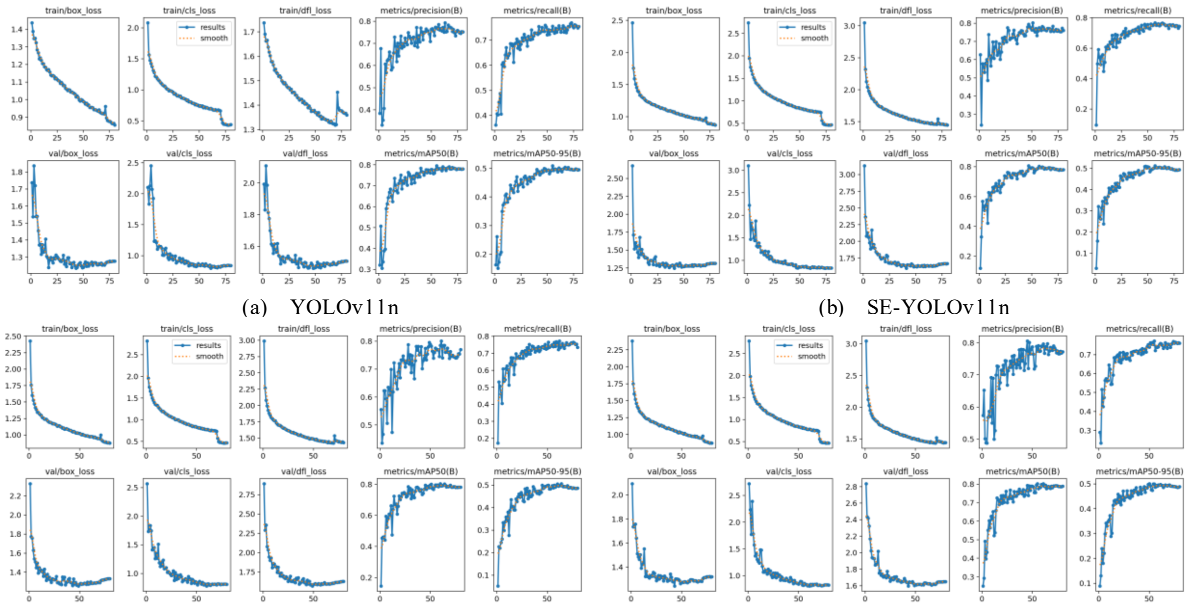
Table 2. Summary of training hyperparameters

Parameter	Configuration
Input image size	640×640
Batch size	32
Epochs	80
Warmup epochs	3
Optimizer	AdamW
Initial learning rate	0.001
Learning rate factor	0.01
Weight decay	0.01
Scheduler	Cosine learning rate schedule
Classification Positive Weight	0.26
Augmentation	Photometric (brightness, contrast, saturation, hue, rgb, noise) and geometric (mosaic, crop, rotation90, flipping)

RESULTS AND DISCUSSION

Result

Throughout the training phase, all four models demonstrated largely similar convergence trends. The training losses consistently diminished over time, whereas the validation losses dropped quickly in the initial epochs and then leveled off. No significant divergence was noted between the paths of training and validation losses. Slight alterations in the loss curves appeared towards the end, aligning with the deactivation of the Mosaic augmentation in the last 10 epochs. In summary, the models exhibited similar convergence characteristics by the conclusion of training, as depicted in Figure 3.



(a) YOLOv11n (b) SE-YOLOv11n

(c) CBAM-YOLOv11n (d) CA-YOLOv11n

Figure 3. Training and validation curves of the four models

Table 3. Overall performance of the four models

Model	Precision	Recall	F1	mAP@50	mAP@50–95	Params	GFLOPs	FPS	Size (MB)
YOLOv11n	0.763	0.753	0.757	0.791	0.513	2,590,815	6.4	15.1	5.26
+ SE	0.794	0.732	0.761	0.794	0.509	2,593,887	6.4	15.9	5.27
+ CBAM	0.776	0.759	0.767	0.810	0.517	2,594,083	6.5	14.4	5.27
+ CA	0.774	0.772	0.773	0.813	0.518	2,595,455	6.5	14.7	5.27

The detection efficiency profiles of the four models varied significantly across the assessed metrics (Table 3). SE-YOLOv11n recorded the highest precision of 0.794. In contrast, CA-YOLOv11n excelled in terms of recall (0.772), F1-score (0.773), mAP@50 (0.813), and mAP@50–95 (0.518). Compared to the baseline, CA-YOLOv11n incorporated an additional 4,640 parameters, increased the GFLOPs from 6.4 to 6.5, and expanded the model size from 5.26 to 5.27 MB, whereas the inference speed decreased from 15.1 to 14.7 FPS. Consequently, CA-YOLOv11n led to four out of five detection metrics, with only slight adjustments in computational demands.

Table 4 presents a summary of the detection performance per class for the four models. The findings indicate class-dependent variations in the attention mechanisms. CA-YOLOv11n attained the highest AP@50 for the Abnormal class, recording a value of 0.891, and shared the highest AP@50 for the Ripe class with SE-YOLOv11n, both achieving 0.869. CBAM-YOLOv11n recorded the highest AP@50 for the Flower and Underripe classes, with values of 0.681 and 0.762, respectively. In contrast, SE-YOLOv11n achieved the highest AP@50 for the Unripe class, with a value of 0.880.

Table 4. Per-class performance

Class	Metric	YOLOv11n	+ SE	+ CBAM	+ CA
Abnormal	Precision	0.932	0.897	0.881	0.905
	Recall	0.690	0.690	0.755	0.748
	AP@50	0.858	0.816	0.868	0.891
Flower	Precision	0.640	0.707	0.695	0.732
	Recall	0.637	0.681	0.681	0.706
	AP@50	0.608	0.667	0.681	0.677
Ripe	Precision	0.722	0.742	0.718	0.685
	Recall	0.813	0.805	0.835	0.851
	AP@50	0.861	0.869	0.860	0.869
Underripe	Precision	0.713	0.745	0.742	0.722
	Recall	0.795	0.710	0.729	0.747
	AP@50	0.743	0.736	0.762	0.752
Unripe	Precision	0.806	0.877	0.842	0.822
	Recall	0.828	0.771	0.792	0.805
	AP@50	0.879	0.880	0.876	0.874

Table 5. CA placement ablation

Placement	Precision	Recall	F1	mAP@50	mAP@50–95	Params	GFLOPs	FPS	Size (MB)
First (P3)	0.788	0.746	0.766	0.795	0.514	2,592,367	6.5	13.5	5.26
First & Second (P3, P4) (Proposed)	0.774	0.772	0.773	0.813	0.518	2,595,455	6.5	14.7	5.27
All Three (P3, P4, P5)	0.777	0.724	0.749	0.781	0.506	2,607,775	6.5	14.7	5.27

Table 5 presents the results of the CA placement ablation study. The highest precision of 0.788 was observed when CA was placed at the initial fusion feature. In contrast, positioning CA at both the first and second fusion features yielded the best recall, F1-score, mAP@50, and mAP@50–95, with values of 0.772, 0.773, 0.813, and 0.518, respectively. Including CA across all three fusion features led to an increase in the number of parameters to 2,607,775 but resulted in a decrease in recall, F1-score, mAP@50, and mAP@50–95 compared to the configuration with the first and second placements. The GFLOPs remained constant at 6.5 for all placements, while FPS varied between 13.5 and 14.7, and the model size ranged from 5.26 to 5.27 MB. Notably, the configuration with CA at the first and second fusion features achieved the highest scores in four out of the five detection metrics among the configurations tested.

Discussion

The key finding of this study is the need to jointly consider attention mechanism selection and placement when configuring neck feature fusion in YOLOv11n. Under consistent training and integration settings, SE, CBAM, and CA produced distinct detection profiles, while CA performance also varied according to its placement. Bian et al. (2025) and Liu et al. (2025) showed that attention can be incorporated into multi-scale feature-fusion strategies, while other studies have demonstrated that neck architecture influences how information is aggregated across feature levels (Wu et al., 2024; Zhao et al., 2024). These findings support the view that both attention type and integration location are important factors in YOLOv11n feature-fusion design. However, because placement was varied only for CA, the present study does not establish a formal interaction effect between attention type and placement. The conclusion therefore remains specific to the dataset, architecture, and configurations examined and should not be generalized to other detectors without further validation.

The observed performance of CA-YOLOv11n may be associated with the characteristics of Coordinate Attention, although the present experiments did not establish a direct causal relationship. Hou et al. (2021) explained that Coordinate Attention encodes information separately along two spatial directions, allowing long-range dependencies to be captured while preserving positional information. Previous studies have demonstrated the compatibility of this mechanism with lightweight YOLO architectures, including its integration with cross-scale feature connections and feature-fusion strategies (Liu et al., 2022; Xie et al., 2022). More recent studies have also incorporated Coordinate Attention into YOLO11-based detectors (Ullah et al., 2025; An et al., 2026). However, these studies mainly demonstrate architectural compatibility and performance improvements rather than proving that positional information directly causes the performance variations observed in this study. This interpretation remains plausible for on-tree FFB detection because plantation images often involve partial visibility, low contrast, occlusion, and image blur, as also reported by Suharjito et al. (2025). Further analysis of intermediate feature maps or attention responses would therefore be required to verify this explanation.

The evaluation of SE, CBAM, and CA indicates that no single attention mechanism consistently outperformed the others across all metrics and categories. SE-YOLOv11n achieved the highest precision, whereas CA-YOLOv11n performed best in recall, F1-score, mAP@50, and mAP@50–95. At the class level, performance also varied across attention mechanisms, indicating distinct trade-offs rather than a clear overall hierarchy. Xu et al. (2024a) and Xu et al. (2024b) demonstrated the integration of attention modules within modified YOLO and feature-fusion architectures, while other studies have explored similar combinations using SE, CBAM, and CA (Zhou et al., 2026; Li et al., 2026). Furthermore, Hanafy et al. (2026) highlighted substantial diversity in attention mechanisms and architectural configurations across plant-health applications. Therefore, the observed differences are more appropriately interpreted as configuration-specific performance profiles rather than evidence that one attention mechanism is universally superior.

Evidence from the placement ablation shows that increasing the number of CA insertions did not consistently improve detection performance. The F1-score increased from 0.766 with CA at the first fusion feature to 0.773 when applied to the first and second fusion features, but decreased to 0.749 when all three were used. Similarly, mAP@50 changed from 0.795 to 0.813 and then declined to 0.781, while mAP@50–95 shifted from 0.514 to 0.518 before decreasing to 0.506. The three-placement configuration also increased the parameter count to 2,607,775 without improving performance. Bian et al. (2025) and Liu et al. (2025) showed that multiscale fusion effectiveness depends on how information is aggregated across feature levels, while cross-scale pathway design is also important (Wu et al., 2024; Zhao et al., 2024). Thus, the effect of CA may depend on the representation level and fusion context. However, the experiments do not establish whether the decline with three placements resulted from redundant processing, altered feature balance, optimization effects, or other factors. Therefore, the first-and-second placement should be regarded as the best-performing configuration among those tested, rather than a universally optimal CA placement.

The module is integrated within the feature-fusion pathway than by the number of insertions alone. Compared with the baseline YOLOv11n, the first-and-second CA configuration added 4,640 parameters, increased GFLOPs from 6.4 to 6.5 and model size from 5.26 to 5.27 MB, while reducing inference speed from 15.1 to 14.7 FPS. At the same time, the F1-score improved from 0.757 to 0.773, mAP@50 from 0.791 to 0.813, and mAP@50–95 from 0.513 to 0.518. Thus, the performance gain was accompanied by a small but measurable computational cost. Li et al. (2025) similarly evaluated orchard detection performance together with computational efficiency, while model complexity has also been emphasized as an important consideration in lightweight on-tree palm fruit detection (Sapkota et al., 2026). These results support further evaluation of CA-YOLOv11n in resource-constrained agricultural environments, although its suitability for specific harvesting, grading, mobile, or edge-computing systems remains to be validated.

There are several limitations that constrain the interpretation of these findings. This study relied on a single oil palm plantation dataset and one lightweight detector architecture, YOLOv11n; therefore, the observed differences may not generalize to other plantation environments, datasets, YOLO variants, or detection architectures. In addition, each configuration was evaluated using only one fixed random seed, which is important because some performance differences were relatively small, such as the increase in mAP@50–95 from 0.513 to 0.518. Gundersen et al. (2023) noted that neural-network performance can vary across training runs due to random initialization, potentially affecting model-comparison conclusions when performance margins are narrow. Repeated experiments using multiple random seeds and appropriate variability measures are therefore needed to assess the stability of the observed improvements.

Gong and Wu (2025) emphasized that factors such as occlusion, illumination variation, motion blur, and background complexity can substantially affect real-world fruit detection performance. Suharjito et al. (2025) similarly reported that these challenging field conditions are relevant in oil palm plantation imagery. However, robustness against such conditions was not evaluated separately in this study. Moreover, no feature-level analysis or attention-response visualization was conducted, so interpretations related to positional information and stage-dependent recalibration remain tentative. Future studies should therefore validate the evaluated CA placements on independent datasets and plantation environments, conduct repeated training runs, test additional detector architectures, and assess robustness under controlled field conditions. Feature-level methods such as Grad-CAM or direct attention visualization should also be employed to clarify how CA placement influences neck feature representations.

CONCLUSION

This study investigated the selection and positioning of attention mechanisms within the YOLOv11n neck for assessing the ripeness of oil palm fresh fruit bunches (FFBs) on trees. Using consistent training and integration parameters, the attention mechanisms SE, CBAM, and CA demonstrated different performance results. CA-YOLOv11n achieved the highest recall, F1-score, mAP@50, and mAP@50–95, while SE-YOLOv11n excelled in precision. Compared to the baseline, CA-YOLOv11n increased the F1-score from 0.757 to 0.773, mAP@50 from 0.791 to 0.813, and mAP@50–95 from 0.513 to 0.518, although it required an additional 4,640 parameters and reduced the inference speed from 15.1 to 14.7 FPS. The placement ablation study indicated that positioning CA at the first and second fusion features was more effective than using it solely at the first or at all three fusion points in four out of five detection metrics. Extending CA to all three fusion features did not improve the detection performance, despite increasing the parameter count. These results underscore the importance of attention placement as a key configuration factor, along with the choice of attention mechanism, within the YOLOv11n neck. However, these conclusions are constrained by the use of a single dataset, one detector architecture, and a fixed-seed evaluation. Future research should explore the consistency of these differences through repeated experiments, validate the placement strategy with independent datasets and detector architectures, assess robustness under controlled plantation conditions, and use feature or attention visualization to understand the impact of CA placement on neck representations.

REFERENCES

Ali, M. L., & Zhang, Z. (2024). The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection. *Computers*, *13*(12), 336. https://doi.org/10.3390/computers13120336

An, X., Sun, L., Zhao, Y., Wang, H., Zhang, M., Li, K., & Xie, R. (2026). CHCG-YOLO: a lightweight model for fruit recognition. *Journal of Food Measurement and Characterization*, *20*(2), 1971–1985. https://doi.org/10.1007/s11694-025-03775-9

Bian, Z., Yao, B., & Li, Q. (2025). SPAFPN: Wear a multi-scale feature fusion scarf around neck for real-time object detector. *Information Fusion*, *119*, 103034. https://doi.org/10.1016/j.inffus.2025.103034

Chen, S., Liu, Y., Zhang, H., & Cai, Y. (2025). A human location and action recognition method based on improved Yolov11 model. *Discover Artificial Intelligence*, *5*(1), 232. https://doi.org/10.1007/s44163-025-00492-6

Gong, X., & Wu, Q. (2025). Fruit Detection Methods Based on Deep Learning in Agricultural Planting: A Systematic Literature Review. *IEEE Access*, *13*, 96092–96110. https://doi.org/10.1109/ACCESS.2025.3573364

Gundersen, O. E., Shamsaliei, S., Kjærnli, H. S., & Langseth, H. (2023). On Reporting Robust and Trustworthy Conclusions from Model Comparison Studies Involving Neural Networks and Randomness. *Proceedings of the 2023 ACM Conference on Reproducibility and Replicability*, 37–61. https://doi.org/10.1145/3589806.3600044

Hanafy, A. El, Hessane, A., & Farhaoui, Y. (2026). Exploring attention mechanisms in deep learning for plant health: A scoping review on disease and pest detection. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2026.05.006

Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate Attention for Efficient Mobile Network Design. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350

Junior, F. A., & Suharjito. (2023). Video based oil palm ripeness detection model using deep learning. *Heliyon*, *9*(1), e13036. https://doi.org/10.1016/j.heliyon.2023.e13036

Lai, J. W., Ramli, H. R., Ismail, L. I., & Hasan, W. Z. W. (2022). Real-Time Detection of Ripe Oil Palm Fresh Fruit Bunch Based on YOLOv4. *IEEE Access*, *10*, 95763–95770. https://doi.org/10.1109/ACCESS.2022.3204762

Li, H., Hu, Z., Wu, F., Li, J., Meng, H., Yang, H., & Wang, Z. (2026). A Study on a Method for Detecting Leaf Diseases in Pumpkins Based on an Improved YOLO11 Model. *AgriEngineering*, *8*(7), 287. https://doi.org/10.3390/agriengineering8070287

Li, J., Zhang, T., Zeng, S., Gao, Q., Wang, L., & Lu, J. (2025). StarNet-Embedded Efficient Network for On-Tree Palm Fruit Ripeness Identification in Complex Environments.

Agriculture, *15*(17), 1823. https://doi.org/10.3390/agriculture15171823

Liu, S., Shao, F., Chu, W., Dai, J., & Zhang, H. (2025). An Improved YOLOv8-Based Lightweight Attention Mechanism for Cross-Scale Feature Fusion. *Remote Sensing*, *17*(6), 1044. https://doi.org/10.3390/rs17061044

Liu, X., Li, G., Chen, W., Liu, B., Chen, M., & Lu, S. (2022). Detection of Dense Citrus Fruits by Combining Coordinated Attention and Cross-Scale Connection with Weighted Feature Fusion. *Applied Sciences*, *12*(13), 6600. https://doi.org/10.3390/app12136600

Naftali, M. G., Hugo, G., & Suharjito. (2024). Palm Oil Counter: State-of-the-Art Deep Learning Models for Detection and Counting in Plantations. *IEEE Access*, *12*, 90395–90417. https://doi.org/10.1109/ACCESS.2024.3419835

Sapkota, R., Meng, Z., Churuvija, M., Du, X., Ma, Z., & Karkee, M. (2026). Comprehensive performance evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on detecting and counting fruitlet in complex orchard environments. *Agriculture Communications*, *4*(1), 100125. https://doi.org/10.1016/j.agrcom.2026.100125

Suharjito, Naftali, M. G., & Hugo, G. (2024). *Oil Palm Fruit Dataset on Plantations for Harvesting Estimation (Version 3) [Dataset]*. Science Data Bank. https://doi.org/10.57760/sciencedb.12647

Suharjito, Naftali, M. G., Hugo, G., Priyadi, M. R. A., Asrol, M., & Utama, D. N. (2025). Oil Palm Fruits Dataset in Plantations for Harvest Estimation Using Digital Census and Smartphone. *Scientific Data*, *12*(1), 972. https://doi.org/10.1038/s41597-025-05227-x

Suparto, A., & Pribadi, M. R. (2026). Improving Oil Palm Fruit Detection under Class Imbalance Using Class-Balanced Focal Loss on YOLOv11. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *15*(02), 165–176. https://doi.org/10.32736/sisfokom.v15i02.2568

Ullah, M. R., Chuang, H., Kayes, A., & Rouf, M. A. (2025). Real-Time Aerial Object Detection via YOLOv11 with Coordinate Attention Enhancement. *Proceedings of the 2025 3rd International Conference on Machine Learning and Pattern Recognition*, 22–27. https://doi.org/10.1145/3760678.3760682

Wu, J., Dai, G., Zhou, W., Zhu, X., & Wang, Z. (2024). Multi-scale feature fusion with attention mechanism for crowded road object detection. *Journal of Real-Time Image Processing*, *21*(2), 29. https://doi.org/10.1007/s11554-023-01409-1

Xie, F., Lin, B., & Liu, Y. (2022). Research on the Coordinate Attention Mechanism Fuse in a YOLOv5 Deep Learning Detector for the SAR Ship Detection Task. *Sensors*, *22*(9), 3370. https://doi.org/10.3390/s22093370

Xu, J., Zhu, X., Shi, L., Li, J., & Guo, Z. (2024). Squeeze-and-excitation attention and bi-directional feature pyramid network for filter screens surface detection. *Journal of Electronic Imaging*, *33*(04). https://doi.org/10.1117/1.JEI.33.4.043044

Xu, M., Zhang, X., Ma, N., & Li, Y. (2024). Detection method of corn leaf diseases based on CA-YOLOv8. *INMATEH Agricultural Engineering*, 129–138. https://doi.org/10.35633/inmateh-73-11

Zhao, Y., Liu, Q., Su, H., Zhang, J., Ma, H., Zou, W., & Liu, S. (2024). Attention-Based Multiscale Feature Fusion for Efficient Surface Defect Detection. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1–10. https://doi.org/10.1109/TIM.2024.3372229

Zhou, Y., Huang, W., Liu, B., Chen, T., Wang, J., Zhang, Q., & Yang, T. (2026). Lightweight GAFNet Model for Robust Rice Pest Detection in Complex Agricultural Environments. *AgriEngineering*, *8*(1), 26. https://doi.org/10.3390/agriengineering8010026