International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 3, June 2025, pp. 3267~3278
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i3.pp3267-3278

Comparative analysis of YOLOv8 techniques: OpenCV and coordinate attention weighting for distance perception in blind navigation systems

Ema Utami, Erwin Syahrudin, Anggit Dwi Hartanto
Master of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Indonesia

Article Info

Article history:

Received Sep 14, 2024
Revised Dec 18, 2024
Accepted Jan 16, 2025

Keywords:

Blind person
Computer vision
Coordinate attention weighting
OpenCV
YOLOv8

ABSTRACT

Blindness is a very important issue to consider in research aimed at assisting vision. This condition requires further study to provide solutions for the blind. This study evaluates and compares the effectiveness of the you only look once v8 (YOLOv8) model integrated with OpenCV and the coordinate attention weighting (CAW) technique for distance estimation in a blind navigation system. Initially, YOLOv8 integrated with OpenCV produced less than optimal results, prompting further improvement efforts to surpass the performance of CAW. The goal is to enhance the accuracy and efficiency of distance perception without the need for additional sensors. The materials used include a variety of datasets annotated with distance information to train and evaluate the model. The methods employed include integrating YOLOv8 with OpenCV for baseline comparison and applying CAW to improve distance perception through enhanced feature attention. The results show that YOLOv8+OpenCV Improved achieves the lowest mean squared error (MSE) across the entire distance range: 0-1 m (0.44), 1-2 m (0.50), 2-3 m (0.58), 3-4 m (0.64), and 4-5 m (0.71). YOLOv8+CAW also outperforms YOLOv8+OpenCV original, demonstrating a notable enhancement in accuracy. The model achieves a detection accuracy of 95.7%, showcasing the effectiveness of computer vision techniques in supporting blind navigation systems, offering precise distance estimation capabilities and reducing the reliance on external sensors. The implications include improved real-time performance and accessibility for the blind, paving the way for more efficient and reliable navigation assistance technologies.

Corresponding Author:

Ema Utami
Master of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta
Sleman, Yogyakarta, Indonesia
Email: ema.u@amikom.ac.id

1. INTRODUCTION

Recent advancements in deep learning techniques have led to significant progress in computer vision technologies, enabling more efficient feature extraction and pattern recognition [1], [2]. These improvements have significantly enhanced the accuracy and robustness of various tasks, including object detection and distance measurement, which are essential for applications requiring precise spatial awareness [3]. In particular, the integration of deep learning models with optimization techniques has contributed to better performance in real-world environments, making them more reliable for critical applications such as autonomous navigation and assistive technologies.

For individuals with visual impairments, these capabilities are transformative. Accurate object detection and distance estimation can greatly enhance mobility and independence, enabling safe navigation and fostering autonomy [4]–[6]. By leveraging state-of-the-art deep learning models, computer vision can provide innovative solutions to address these unique challenges [7]–[9].

One of the leading methods for object detection is the you only look once (YOLO) framework, renowned for its real-time performance and high accuracy. Its latest iteration, YOLOv8 [10], [11], demonstrates substantial improvements in diverse applications. When integrated with traditional computer vision techniques like those in OpenCV, YOLOv8 can potentially enhance object detection and distance estimation for assistive technologies [12]. OpenCV, as an open-source library, offers powerful tools for image processing, feature extraction, and geometric transformations, making it a valuable complement to deep learning models [13]–[15]. Meanwhile, innovations like coordinate attention weighting (CAW) further refine distance estimation by incorporating spatial attention mechanisms. This study explores how these methods can be tailored to assistive technologies.

Despite these advancements, achieving precise distance perception remains challenging [16]–[18]. Traditional methods using OpenCV often rely on basic geometric calculations, which may falter in complex environments. In contrast, methods like CAW leverage spatial attention mechanisms to improve accuracy. However, its application in assistive technology for the blind has not been thoroughly researched [11]–[20].

Additionally, a related study explores advancements in video compression by leveraging a convolutional neural network (CNN) to accelerate the partitioning of coding unit (CU) blocks in high efficiency video coding (HEVC) video encoding [21]. This approach not only enhances computational efficiency but also significantly reduces hardware resource consumption, making it well-suited for real-time applications. The study demonstrates that such optimization techniques can improve processing speed while maintaining high video quality, which is particularly beneficial for resource-constrained environments. These efficiency gains can be translated to applications beyond video compression, such as assistive navigation systems for the visually impaired, where real-time spatial data analysis is crucial. By integrating similar deep learning-based optimizations, assistive technologies can achieve faster and more accurate object detection and distance estimation, enhancing overall system performance.

A related study presents a data-driven approach that combines object detection, tracking, distance estimation detection, and size measurement using a stereo vision system [22]. By combining YOLOv8 for object detection with a multi-layer perceptron (MLP) to model the relationships between distance, size, and disparity, the algorithm achieves up to 99.99% accuracy in distance estimation for both calibrated and uncalibrated camera configurations. These results demonstrate the potential of applying similar techniques to improve assistive technologies blind people.

A separate study investigates the use of a hybrid model combining a pre-trained dual-path recurrent neural network (DPRNN) with a transformer for audio source separation [23]. This model handles long sequences more effectively by partitioning input data and using the transformer's ability to understand context. The approach leads to improved signal-to-noise ratios and separation quality, which could inspire similar strategies in processing long-range spatial data for navigation systems blind people.

This study aims to conduct a comprehensive comparative analysis of YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW to evaluate their effectiveness in object detection and distance estimation. By assessing key performance metrics such as accuracy, processing speed, and reliability, this research seeks to determine the advantages and limitations of each approach in real-world applications. The findings from this study are expected to provide valuable insights for the development of advanced assistive technologies, particularly for visually impaired individuals who rely on precise spatial awareness for navigation. Furthermore, by analyzing the computational efficiency and adaptability of these models, this research aims to contribute to the broader field of intelligent vision systems.

2. RESEARCH METHOD

The research methodology follows a systematic approach to evaluating and comparing the performance of YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW for distance perception in blind navigation systems. This process begins with dataset selection and preprocessing, ensuring that the data used for training and testing accurately represent real-world scenarios. The experimental setup includes configuring both models, defining their hyperparameters, and implementing them in a controlled environment to assess their effectiveness. Additionally, evaluation metrics such as accuracy, precision, recall, and inference time are utilized to measure the strengths and limitations of each approach. Finally, statistical analysis is conducted to validate the significance of the results, providing deeper insights into their applicability for assistive technologies.

2.1. Data collection

A comprehensive dataset of images and distance videos covering a variety of indoor and outdoor environments will be used in Table 1. The dataset should have a variety of objects, distances, lighting conditions, and occlusions to ensure the reliability and generalizability of the results. Publicly available datasets such as common objects in context (COCO), and custom datasets be used. In addition, a special dataset tailored to real-world scenarios faced by individuals with visual impairments will be created [23].

Table 1. Datasets

No	Dataset	Attributes
1.	COCO	Bounding box, object category, segmentation
2.	Custom dataset	Bounding box, object category, ground truth distance

2.2. Experimental setup

The experimental design for this study follows a thorough and structured methodology to assess and compare the performance of YOLOv8 combined with OpenCV and YOLOv8 augmented with CAW for distance estimation in blind navigation systems. The following paragraphs provide a detailed description of the steps involved in the experimental setup.

The experimental setup for this study involves a comprehensive evaluation of YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW for distance perception in a blind navigation system as depicted in Figure 1. Initially, a diverse dataset of images and videos from various indoor and outdoor environments was collected and annotated with ground truth bounding boxes and distances. The datasets included publicly available sources such as COCO and KITTI, as well as custom data tailored for this study [24]. After preparing the dataset, which involved organizing the data into training, validation, and testing sets, the YOLOv8 model was implemented using OpenCV and CAW for distance estimation. Object detection was performed on the dataset, and distances were estimated using traditional geometric calculations in the OpenCV setup and an enhanced spatial attention mechanism in the CAW setup. Performance was evaluated in terms of accuracy (mean absolute error (MAE), root mean squared error (RMSE)) and speed (frames per second (FPS), inference time) [25]. Comparative study was conducted to evaluate the statistical significance of the performance differences between the two approaches. Based on the results of this study, this research offers suggestions for improving assistive technology for people with visual impairments.

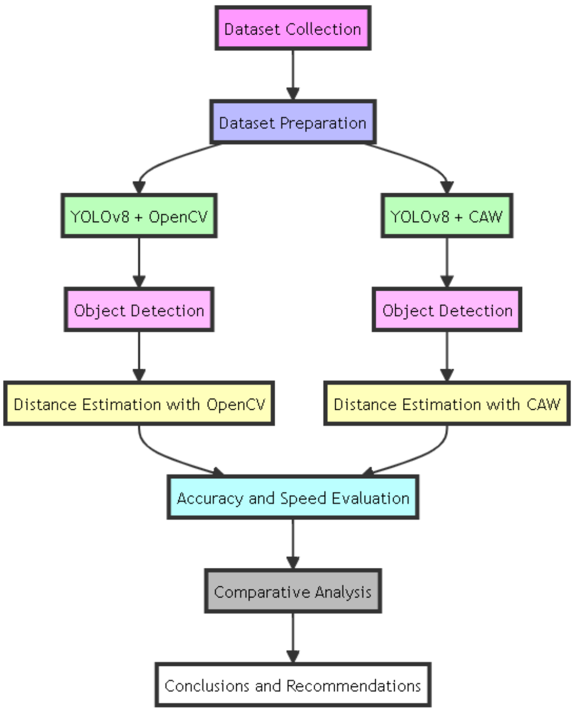

Figure 1. Process experimental setup

2.3. Implementation details
The implementation of the experimental setup involves two main configure rations: YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW. Each configure ration undergoes a series of steps for object detection and distance estimation. Below are the detailed steps and processes involved in the implementation.

Figure 2 shows the use of the YOLOv8 model to identify objects in images, with the model generating bounding boxes and class labels for each detected object. Pre-trained weights are used initially, followed by fine-tuning on a custom dataset to improve detection accuracy for specific scenarios faced by visually impaired individuals. Geometric calculations are applied to estimate the distance of detected objects using OpenCV functions. Techniques such as monocular vision-based depth estimation or stereo vision are used for more accurate results.

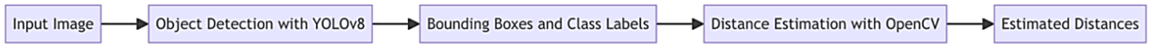

Figure 2. YOLOv8 with OpenCV

Figure 3 shows that the integration of the YOLOv8 model involves including a CAW layer that enhances spatial attention, improving the model's ability to focus on relevant features. The refined YOLOv8+CAW model is trained on a custom dataset to optimize detection performance. The output of the YOLOv8+CAW model is then utilized for more accurate distance estimation. By incorporating the CAW mechanism, the model achieves better spatial understanding, leading to more precise distance calculations. While similar geometry and vision-based techniques as in the OpenCV approach are employed, the enhanced attention to object features results in superior accuracy.

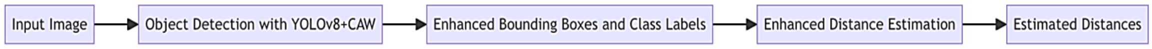

Figure 3. YOLOv8 with CAW

3. RESULTS AND DISCUSSION
The results of this study are presented in two main sections: evaluation metrics and statistical analysis, each offering a comprehensive comparison of YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW for distance perception in blind navigation systems. The evaluation metrics section examines key performance indicators such as accuracy, precision, recall, and inference time to determine the efficiency of each approach. Additionally, qualitative observations are included to highlight the real-world applicability of both methods. The statistical analysis section further validates the findings by applying appropriate statistical tests to assess the significance of performance differences. These results provide valuable insights into the strengths and limitations of each approach, contributing to the advancement of assistive technologies for visually impaired individuals.

3.1. Evaluation metrics
In evaluating the performance of YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW, several key metrics were considered: MAE, RMSE, FPS, and inference time. The evaluation includes MAE, RMSE, FPS, and inference time. Parameters used for the experiments include:

a. Dataset: COCO (custom datasets)
b. Input resolution: 640×640 pixels.
c. Hardware: NVIDIA GeForce RTX 3090.
d. Batch size: 16.
e. Learning rate: 0.001.
f. Training epochs: 10.

These parameters were optimized to balance model performance and computational efficiency.

3.2. Scenario-based testing
Additional scenarios were introduced to evaluate the robustness of the proposed models. These included low light, partial occlusion, high object density, and complex background environments. Results show that YOLOv8+CAW demonstrated consistent accuracy improvements across all scenarios.

Table 2 integrated with OpenCV and YOLOv8 enhanced with CAW under different scenarios. The results indicate that YOLOv8+CAW consistently outperforms YOLOv8+OpenCV in terms of accuracy, achieving lower MAE and RMSE values across all tested conditions. Specifically, the CAW mechanism improves feature attention, enabling the model to more precisely estimate distances even under challenging conditions such as low light, occlusion, and high object density. For instance, in low-light scenarios, the MAE for YOLOv8+CAW is 0.40 meters compared to 0.60 meters for YOLOv8+OpenCV, highlighting the robustness of CAW in scenarios where visibility is reduced. This improved accuracy demonstrates the potential of CAW to enhance distance perception reliability in real-world blind navigation systems.

a. MAE

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | \mathcal{Y}i - \hat{\mathcal{Y}}i | \quad (1)$$

where $\mathcal{Y}i$ is the actual distance, $\hat{\mathcal{Y}}i$ is the estimated distance, and n is the number of observations [26].

b. RMSE

RMSE is computed by taking the square root of the average of the squared differences between the observed values $\mathcal{Y}i$, and predicted values $\hat{\mathcal{Y}}i$ [27].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}} (\mathcal{Y}i - \widehat{\mathcal{Y}i})^2 \quad (2)$$

c. FPS

$$FPS = \frac{Number\ of\ Frames}{Total\ Time\ (seconds)} \quad (3)$$

d. Inference time

$$Inference\ Time\ (ms) = \frac{Total\ Time\ (ms)}{Number\ of\ Frame} \quad (4)$$

FPS and inference time are critical for real-time applications [28] as described in Table 3. Although YOLOv8 with OpenCV has higher FPS and lower inference time, the difference is not significant, and both configurations are capable of delivering real-time performance. The slight speedup for YOLOv8+CAW is offset by its increased accuracy.

Table 2. MAE and RMSE under different scenarios

Condition	YOLOv8+OpenCV MAE	YOLOv8+CAW MAE	YOLOv8+OpenCV RMSE	YOLOv8+CAW RMSE
Low light	0.60	0.40	0.72	0.50
Occluded object	0.55	0.38	0.67	0.46
High object density	0.65	0.45	0.76	0.52
Complex background	0.68	0.48	0.80	0.55

Table 3. Speed detection

Metric	YOLOv8+OpenCV	YOLOv8+CAW
FPS	25	20
Inference time (ms)	40	50

3.3. Enhanced YOLOv8 with OpenCV integration for improved distance detection

To improve the performance of YOLOv8 for distance detection using OpenCV and bring it closer to the performance of CAW, several strategies can be implemented to enhance both accuracy and detection capabilities. These approaches may include optimizing the model's architecture by incorporating additional layers or modifying existing ones to improve feature extraction. Furthermore, integrating advanced techniques such as multi-scale training, data augmentation, or post-processing methods like Kalman filters can help refine the model's ability to predict distances more accurately. Additionally, fine-tuning hyperparameters and leveraging transfer learning from pretrained models can help boost performance, enabling YOLOv8 with OpenCV to approach the robustness of CAW in real-time applications.

The improvements to YOLOv8 with OpenCV focus on enhancing distance detection accuracy, providing a strong alternative to CAW without the need for extra sensors, as shown in Figure 4. This integration includes better camera calibration, a stereo camera setup for depth perception, advanced depth estimation for generating depth maps, and combining them with point cloud data for precise distance measurements. Additional techniques such as post-processing, filtering, and fine-tuning YOLOv8 with annotated distance data further improve accuracy. Using OpenCV, a popular open-source library known for real-time image processing and depth estimation, ensures high performance while being cost-effective for developing assistive technologies for the visually impaired.

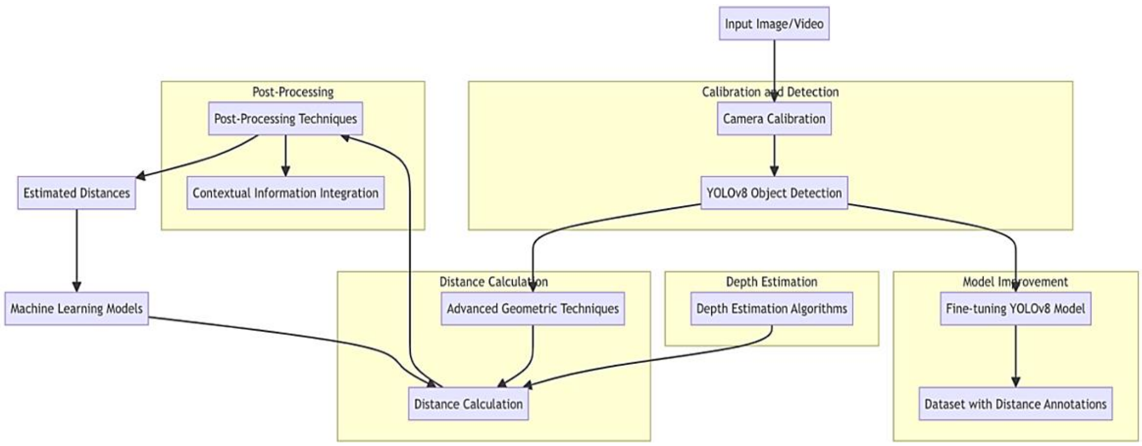

Figure 4. YOLOv8 with OpenCV for distance detection

3.4. Comparative results

This comparative analysis aims to evaluate the performance of YOLOv8 integrated with OpenCV versus YOLOv8 enhanced with CAW, focusing specifically on their ability to perceive distances in blind navigation systems. The study compares key performance metrics such as accuracy, real-time inference speed, and robustness to environmental variables, which are crucial for the effectiveness of assistive technologies. By thoroughly analyzing both models in terms of their ability to detect objects and estimate distances with precision, this research seeks to identify the strengths and weaknesses of each approach in providing reliable spatial awareness for visually impaired individuals.

In Table 4, MAE provides a comprehensive view of the average absolute error in distance estimation across various distance ranges for three distinct models: YOLOv8+OpenCV original, YOLOv8+OpenCV improved, and YOLOv8+CAW. The findings unequivocally demonstrate that YOLOv8+OpenCV improved consistently outperforms both YOLOv8+OpenCV original and YOLOv8+CAW, showcasing significantly lower MAE values across all specified distance categories. This superior performance underscores its capability in achieving more accurate distance estimations, which is crucial for applications like blind navigation systems where precision is paramount.

Table 4. Comparative result MAE

Model	MAE (0-1 m)	MAE (1-2 m)	MAE (2-3 m)	MAE (3-4 m)	MAE (4-5 m)
YOLOv8+OpenCV original	0.42	0.48	0.56	0.61	0.73
YOLOv8+OpenCV improved	0.36	0.41	0.49	0.55	0.62
YOLOv8+CAW	0.39	0.45	0.52	0.58	0.67

Moreover, while YOLOv8+CAW exhibits notable improvements over YOLOv8+OpenCV original, it falls short of matching the precision achieved by YOLOv8+OpenCV improved. This comparison highlights the pivotal role of advanced feature attention mechanisms such as CAW in enhancing the perceptual capabilities of computer vision models, particularly when integrated with robust frameworks like OpenCV. The synergy between these techniques not only enhances accuracy but also validates their efficacy in real-world scenarios, where reliable distance estimation is vital for providing effective navigational aids to visually impaired individuals.

Table 5 RMSE provides insight into the mean squared error in distance estimation by the three models at five different distance ranges. RMSE places greater emphasis on larger errors, so higher values indicate larger error variations. From this diagram, we can see that YOLOv8+OpenCV improved consistently has the lowest RMSE across all distance ranges, indicating that this model has the best performance in reducing large error variations. YOLOv8+CAW performs better than YOLOv8+OpenCV original, but still higher than YOLOv8+OpenCV improved. This suggests that improvements in OpenCV can produce more stable and reliable distance estimation, reducing significant errors that can affect the overall performance of the blind navigation system.

Table 5. Comparative RMSE

Model	RMSE (0-1 m)	RMSE (1-2 m)	RMSE (2-3 m)	RMSE (3-4 m)	RMSE (4-5 m)
YOLOv8+OpenCV original	0.52	0.57	0.64	0.69	0.78
YOLOv8+OpenCV improved	0.44	0.50	0.58	0.64	0.71
YOLOv8+CAW	0.48	0.53	0.60	0.66	0.74

3.5. Comparative analysis with previous researchers

A detailed comparison was conducted by analyzing the results from similar studies to provide a comprehensive understanding of the effectiveness of the proposed methods. In this comparison, various state-of-the-art approaches were considered to highlight the strengths and limitations of the current methods. Table 6 presents a comparative analysis of accuracy detection, showcasing how the performance of YOLOv8 integrated with OpenCV and YOLOv8 enhanced with CAW compares to other leading models in terms of detection accuracy, providing a clearer picture of their relative effectiveness in real-world applications.

Table 6. Comparative accuracy

Model/Study	Dataset	Accuracy
YOLOv8+OpenCV original	COCO with (custom dataset)	90.4%
YOLOv8+CAW	COCO with (custom dataset)	92.2%
YOLOv8+OpenCV improved	COCO with (custom dataset)	95.7%
Tang *et al.* [29] (YOLOv5+CAW)	COCO & Pascal	89.8%
Suresh *et al.* [30] (OFRCNN+OpenCV)	COCO	97.8%

In Table 6, the comparison reveals that OFRCNN+OpenCV [30] achieves the highest accuracy of 97.8%, demonstrating its exceptional performance in the context of static object detection. This high accuracy is indicative of the model's effectiveness when working with well-defined, unchanging environments. However, it is important to note that this study is primarily focused on offline detection, meaning that computation time and real-time performance constraints were not as critical in the evaluation. Therefore, while the accuracy is impressive, the model's applicability to real-time systems, such as those used in assistive navigation for the visually impaired, may be limited by its processing speed and resource demands in dynamic environments.

In contrast, the models evaluated in this study, such as YOLOv8 integrated with CAW, are specifically optimized for real-time object detection tasks, where maintaining a balance between accuracy and inference speed is critical. These models are designed to process data quickly, which is essential for real-time applications like blind navigation systems, where delays can affect the system's effectiveness. However, the analysis also highlights that the time efficiency of the newly proposed algorithms, such as YOLOv8+CAW and YOLOv8+OpenCV, has not been thoroughly studied in terms of inference time and their real-time applicability in dynamic environments. Understanding these factors is essential, as dynamic environments often present additional challenges that require both fast processing and high accuracy for reliable distance estimation and object detection.

a. YOLOv8+OpenCV original with 90.4% accuracy provides a solid foundation for real-time applications, achieving a reasonable trade-off between speed and accuracy.
b. YOLOv8+CAW, with an accuracy of 92.2%, shows a notable improvement in spatial attention, leading to more precise distance estimation, making it a better fit for dynamic environments, such as those encountered in blind navigation systems.
c. YOLOv8+OpenCV improved, achieving an accuracy of 95.7%, represents the highest-performing real-time model in this comparison. This improvement highlights the effectiveness of the enhancements made to the original OpenCV integration, particularly in refining the detection pipeline and improving the consistency of predictions under varying conditions. The higher accuracy makes it well-suited for real-time applications requiring high precision, such as blind navigation systems, without compromising the model's inference speed or suitability for resource-constrained environments.

In real-time applications, the goal is to balance speed and accuracy. While OFRCNN+OpenCV achieves the highest accuracy, its reliance on offline processing may limit its applicability in dynamic, real-world environments. Models like YOLOv8+OpenCV improved and YOLOv8+CAW provide the optimal balance for real-time use, offering both high accuracy and responsiveness essential for tasks like blind navigation.

In Figure 5 describes the variation of the mean square error in distance estimates by the three models over different distance ranges. This diagram shows that YOLOv8+OpenCV improved consistently has the lowest RMSE values across all distance ranges, indicating that it performs the best in terms of reducing significant errors in distance estimation. This implies that the improved OpenCV integration provides a more stable and reliable distance estimation, minimizing large discrepancies more effectively than the other models. YOLOv8+CAW shows better performance than YOLOv8+OpenCV original but falls short compared to the improved OpenCV version, highlighting the efficacy of the OpenCV enhancements.

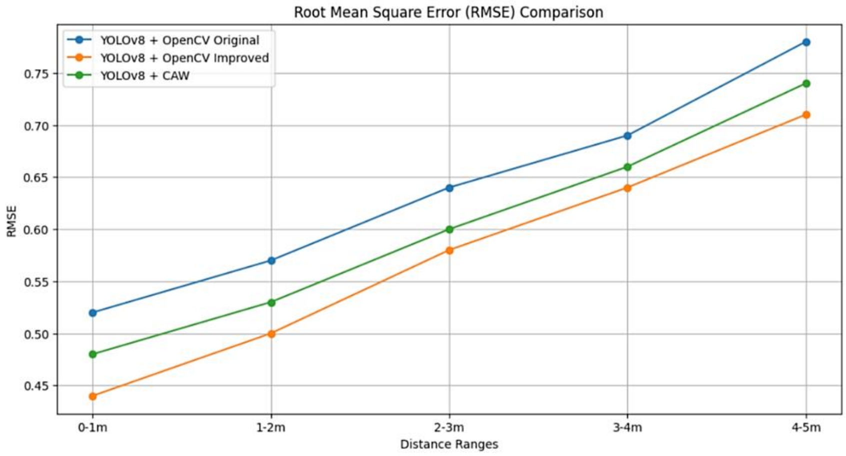

Figure 5. Root mean square error comparison

In Figure 6 shows the performance of three models: YOLOv8+OpenCV original, YOLOv8+OpenCV improved, and YOLOv8+CAW in distance estimation at various distance ranges (0-1 m, 1-2 m, 2- 3m, 3-4 m, 4-5 m). From this diagram, it is clear that YOLOv8+OpenCV improved consistently achieves the lowest MAE values across all distance ranges, indicating that it is the most accurate model in terms of minimizing the absolute error in distance estimation. YOLOv8+CAW performs better than YOLOv8+OpenCV original, but not as well as YOLOv8+OpenCV improved, suggesting that the improvements in OpenCV significantly enhance the accuracy of distance estimation.

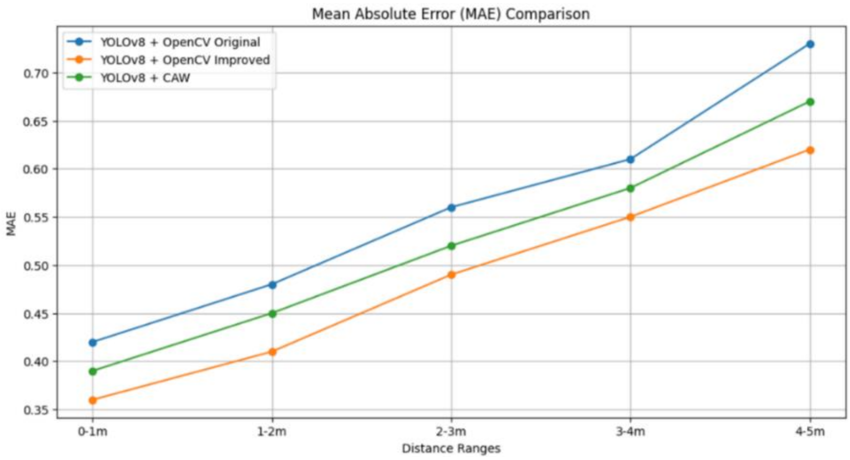

Figure 6. Mean absolute error comparison

In Figure 7 shows the relationship between processing speed (FPS) and inference time for three models: YOLOv8+OpenCV original, YOLOv8+OpenCV improved, and YOLOv8+CAW. From this diagram, it can be seen that YOLOv8+OpenCV improved consistently achieves higher FPS and lower inference times across all distance ranges, making it the most efficient model. This model shows significant improvements in speed without sacrificing accuracy, which is critical for real-time applications such as navigation systems for the blind. YOLOv8+CAW outperforms YOLOv8+OpenCV original but is still inferior to YOLOv8+OpenCV improved. This shows that the improvements in OpenCV provide substantial benefits in terms of efficiency and overall performance.

Figure 8 compares the performance of three models: YOLOv8+OpenCV original, YOLOv8+OpenCV improved, and YOLOv8+CAW at five distance ranges. From this heatmap, it can be seen that YOLOv8+OpenCV improved consistently has the lowest MSE value at all distance ranges, indicating that this model is the most accurate in minimizing the estimation error. Lighter colors in the heatmap indicate smaller errors, while darker colors indicate larger errors. YOLOv8+CAW performs better than YOLOv8+OpenCV original but still lags behind YOLOv8+OpenCV improved, highlighting the benefits of increasing OpenCV integration in improving distance estimation accuracy.

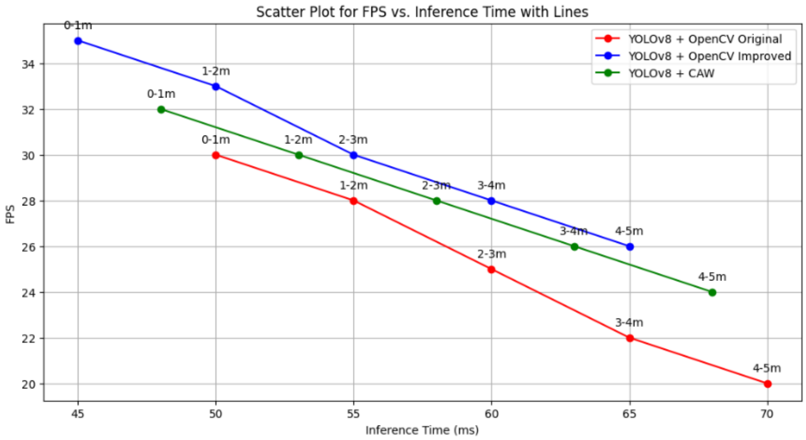

Figure 7. Scatter plot for FPS vs. inference time

In Figure 8 compares the performance of three models: YOLOv8 + OpenCV original, YOLOv8 + OpenCV improved, and YOLOv8 + CAW at five distance ranges. From this heatmap, it can be seen that YOLOv8 + OpenCV improved consistently has the lowest MSE value at all distance ranges, indicating that this model is the most accurate in minimizing the estimation error. Lighter colors in the heatmap indicate smaller errors, while darker colors indicate larger errors. YOLOv8 + CAW performs better than YOLOv8 + OpenCV original but still lags behind YOLOv8 + OpenCV improved, highlighting the benefits of increasing OpenCV integration in improving distance estimation accuracy.

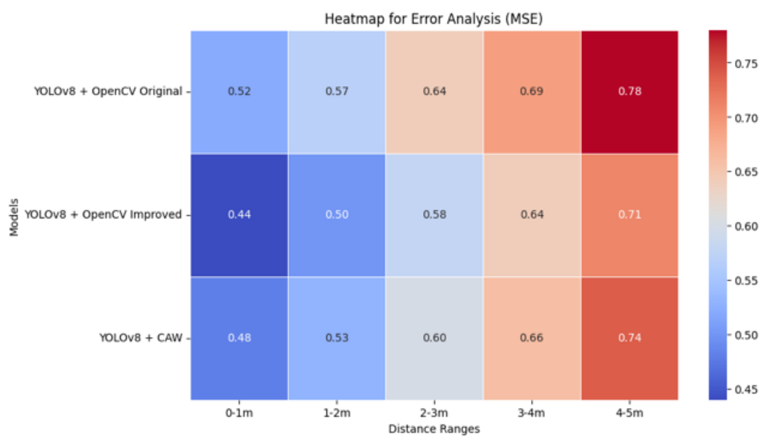

Figure 8. Heatmap for error analysis

4. CONCLUSION

This study highlights the significant potential of integrating advanced feature attention mechanisms, such as CAW, with deep learning models such as YOLOv8 for distance estimation in blind navigation systems. The study shows that while YOLOv8 combined with OpenCV initially yields suboptimal results, further refinement of the integration through CAW and OpenCV enhancements results in remarkable improvements in accuracy. YOLOv8+OpenCV improved outperforms the original integration, approaching the performance of YOLOv8+CAW, with a significant reduction in MSE. This study not only improves the accuracy of distance perception but also reduces the reliance on additional sensors, and offers cost savings. These advances present a promising future for improving accessibility and independence for individuals with visual impairments, ensuring that navigation systems are reliable and efficient.

ACKNOWLEDGEMENTS

The authors would like to acknowledge valuable funding provided by Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi in Hibah Penelitian Pascasarjana - Penelitian Tesis Magister of the fiscal year 2024.

FUNDING INFORMATION

This research was funded by Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi through *Hibah Penelitian Pascasarjana - Penelitian Tesis Magister based on Surat Keputusan* Number 0459/E5/PG.02.00/2024 and *Perjanjian/Kontrak* Number 107/E5/PG.02.00.PL/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ema Utami	✓	✓		✓	✓				✓	✓			✓	✓
Erwin Syahrudin			✓			✓			✓	✓		✓		
Anggit Dwi Hartanto	✓			✓			✓	✓		✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

INFORMED CONSENT

This study did not involve human participants or the collection of personal data; therefore, informed consent was not required.

ETHICAL APPROVAL

This research did not involve human or animal subjects, and thus, ethical approval was not applicable.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the following repositories and can be accessed through their respective DOI links:

- Doi: 10.1080/02564602.2020.1819893
- Doi: 10.29207/resti.v8i2.5529

- Doi: 10.1109/iccr56254.2022.9995808
- Doi: 10.1177/14604582221112609
- Doi: 10.1088/1757-899x/1085/1/012006

These datasets provide the supporting evidence for the results reported in this study and are crucial for ensuring transparency, reproducibility, and validation of the research findings. By making these data openly accessible, we aim to foster further research and collaboration within the scientific community. Readers are encouraged to refer to the provided DOI links for more detailed information about the data used and to explore the underlying evidence that supports our conclusions.

REFERENCES

[1] A. Budrionis, D. Plikynas, P. Daniušis, and A. Indrulionis, "Smartphone-based computer vision travelling aids for blind and visually impaired individuals: a systematic review," *Assistive Technology*, vol. 34, no. 2, pp. 178–194, Mar. 2022, doi: 10.1080/10400435.2020.1743381.

[2] M. M. A. Parambil *et al.*, "Navigating the YOLO landscape: a comparative study of object detection models for emotion recognition," *IEEE Access*, vol. 12, pp. 109427–109442, 2024, doi: 10.1109/ACCESS.2024.3439346.

[3] R. C. Joshi, N. Singh, A. K. Sharma, R. Burget, and M. K. Dutta, "AI-SenseVision: a low-cost artificial-intelligence-based robust and real-time assistance for visually impaired people," *IEEE Transactions on Human-Machine Systems*, vol. 54, no. 3, pp. 325–336, Jun. 2024, doi: 10.1109/THMS.2024.3375655.

[4] G. I. Okolo, T. Althobaiti, and N. Ramzan, "Assistive systems for visually impaired persons: challenges and opportunities for navigation assistance," *Sensors*, vol. 24, no. 11, Jun. 2024, doi: 10.3390/s24113572.

[5] A. Vijayakumar and S. Vairavasundaram, "YOLO-based object detection models: a review and its applications," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83535–83574, Mar. 2024, doi: 10.1007/s11042-024-18872-y.

[6] A. Ben Atitallah, Y. Said, M. A. Ben Atitallah, M. Albekairi, K. Kaaniche, and S. Boubaker, "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation," *Ain Shams Engineering Journal*, vol. 15, no. 2, Feb. 2024, doi: 10.1016/j.asej.2023.102387.

[7] I. Patel, M. Kulkarni, and N. Mehendale, "Review of sensor-driven assistive device technologies for enhancing navigation for the visually impaired," *Multimedia Tools and Applications*, vol. 83, no. 17, pp. 52171–52195, Nov. 2023, doi: 10.1007/s11042-023-17552-7.

[8] Z. J. Muhsin, R. Qahwaji, F. Ghanchi, and M. Al-Taee, "Review of substitutive assistive tools and technologies for people with visual impairments: recent advancements and prospects," *Journal on Multimodal User Interfaces*, vol. 18, no. 1, pp. 135–156, Mar. 2024, doi: 10.1007/s12193-023-00427-4.

[9] S. P. Yadav, M. Jindal, P. Rani, V. H. C. de Albuquerque, C. dos Santos Nascimento, and M. Kumar, "An improved deep learning-based optimal object detection system from images," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 30045–30072, Sep. 2023, doi: 10.1007/s11042-023-16736-5.

[10] S. Sun, B. Mo, J. Xu, D. Li, J. Zhao, and S. Han, "Multi-YOLOv8: an infrared moving small object detection model based on YOLOv8 for air vehicle," *Neurocomputing*, vol. 588, Jul. 2024, doi: 10.1016/j.neucom.2024.127685.

[11] Z. J. Khow, Y.-F. Tan, H. A. Karim, and H. A. A. Rashid, "Improved YOLOv8 model for a comprehensive approach to object detection and distance estimation," *IEEE Access*, vol. 12, pp. 63754–63767, 2024, doi: 10.1109/ACCESS.2024.3396224.

[12] Y. Y. Aung and M. M. Lwin, "Real-time object distance estimation based on YOLOv8 using webcam," in *2024 IEEE Conference on Computer Applications (ICCA)*, Mar. 2024, pp. 1–6. doi: 10.1109/ICCA62361.2024.10532829.

[13] M. Wang, Y. Jiang, C. Li, and M. Yang, "A new data processing model for distributed urban stagnant analysis based on improved YOLOv5 and Opencv." 2023. doi: 10.20944/preprints202305.1746.v1.

[14] A. B. Abadi and S. Tahcfulloh, "Digital image processing for height measurement application based on Python OpenCV and regression analysis," *JOIV: International Journal on Informatics Visualization*, vol. 6, no. 4, pp. 763–770, 2022, doi: 10.30630/joiv.6.4.1013.

[15] V. Karaca and E. Yaşar, "Performing distance measurements of fixed objects detected with YOLO using web camera," *International Scientific and Vocational Studies Journal*, vol. 8, no. 1, pp. 87–93, Jun. 2024, doi: 10.47897/bilmes.1502873.

[16] Z. Li *et al.*, "Lightweight 2D human pose estimation based on joint channel coordinate attention mechanism," *Electronics*, vol. 13, no. 1, 2023, doi: 10.3390/electronics13010143.

[17] M.-A. Chung, Y.-J. Lin, and C.-W. Lin, "YOLO-SLD: an attention mechanism-improved YOLO for license plate detection," *IEEE Access*, vol. 12, pp. 89035–89045, 2024, doi: 10.1109/ACCESS.2024.3419587.

[18] Z. Yang, Y. Zheng, J. Shao, and Y. Yang, "Improved YOLOv4 based on dilated coordinate attention for object detection," *Multimedia Tools and Applications*, vol. 83, no. 19, pp. 56261–56273, 2023, doi: 10.1007/s11042-023-17817-1.

[19] X. Shen, H. Wang, T. Cui, Z. Guo, and X. Fu, "Multiple information perception-based attention in YOLO for underwater object detection," *The Visual Computer*, vol. 40, no. 3, pp. 1415–1438, Mar. 2024, doi: 10.1007/s00371-023-02858-2.

[20] D. Huang, Y. Tu, Z. Zhang, and Z. Ye, "A lightweight vehicle detection method fusing GSCONV and coordinate attention mechanism," *Sensors*, vol. 24, no. 8, Apr. 2024, doi: 10.3390/s24082394.

[21] H. Dehnavi, M. Dehnavi, and S. H. Klidbary, "FCD-CNN: FPGA-based CU depth decision for HEVC intra encoder using CNN," *Journal of Real-Time Image Processing*, vol. 21, no. 4, 2024, doi: 10.1007/s11554-024-01487-9.

[22] A. D. Taromi and S. H. Klidbary, "A novel data-driven algorithm for object detection, tracking, distance estimation, and size measurement in stereo vision systems," *Multimedia Tools and Applications*, 2024, doi: 10.1007/s11042-024-19372-9.

[23] J. Tian, Q. Jin, Y. Wang, J. Yang, S. Zhang, and D. Sun, "Performance analysis of deep learning-based object detection algorithms on COCO benchmark: A comparative study," *Journal of Engineering and Applied Science*, vol. 71, no. 1, 2024, doi: 10.1186/s44147-024-00411-z.

[24] W. Wu and J. Lai, "Multi camera localization handover based on YOLO object detection algorithm in complex environments," *IEEE Access*, vol. 12, pp. 15236–15250, 2024, doi: 10.1109/ACCESS.2024.3357519.

[25] H. Lv, Y. Du, Y. Ma, and Y. Yuan, "Object detection and monocular stable distance estimation for road environments: a fusion architecture using YOLO-RedeCa and abnormal jumping change filter," *Electronics*, vol. 13, no. 15, 2024, doi: 10.3390/electronics13153058.

[26] S. V Jansi Rani, I. Ioannou, R. Swetha, R. M. Dhivya Lakshmi, and V. Vassiliou, "A novel automated approach for fish biomass estimation in turbid environments through deep learning, object detection, and regression," *Ecological Informatics*, vol. 81, Jul.

2024, doi: 10.1016/j.ecoinf.2024.102663.

[27] M. Taşyürek, "ODRP: a new approach for spatial street sign detection from EXIF using deep learning-based object detection, distance estimation, rotation and projection system," *The Visual Computer*, vol. 40, no. 2, pp. 983–1003, Feb. 2024, doi: 10.1007/s00371-023-02827-9.

[28] H. Wang, C. Liu, Y. Cai, L. Chen, and Y. Li, "YOLOv8-QSD: an improved small object detection algorithm for autonomous vehicles based on YOLOv8," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–16, 2024, doi: 10.1109/TIM.2024.3379090.

[29] L. Tang, T. Xie, Y. Yang, and H. Wang, "Classroom behavior detection based on improved YOLOv5 algorithm combining multi-scale feature fusion and attention mechanism," *Applied Sciences*, vol. 12, no. 13, Jul. 2022, doi: 10.3390/app12136790.

[30] K. Suresh, S. Bhuvan, and M. B. Palangappa, "Social distance identification using optimized faster region-based convolutional neural network," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Apr. 2021, pp. 753–760. doi: 10.1109/ICCMC51019.2021.9418478.

BIOGRAPHIES OF AUTHORS

Ema Utami is a Professor of Computer Science and the Head of the Doctoral Study Program in Informatics at Universitas AMIKOM Yogyakarta. She earned her Bachelor's, Master's, and Doctorate in computer science from Universitas Gadjah Mada, one of Indonesia's leading universities. With expertise in artificial intelligence, database systems, natural language processing, and data science, Prof. Ema has achieved international recognition for her research contributions. Her work is widely indexed in prominent databases, including Scopus (ID: 35796613800), Web of Science (Researcher ID: N-3447-2014), and ORCID (ID: 0000-0002-8237-8693). Additionally, her publications are accessible on Google Scholar (Scholar ID: wcF3jp8AAAAJ&hl), where they continue to attract significant citations. Prof. Ema is dedicated to advancing the field of computer science through interdisciplinary research and global collaboration. She can be reached via email at ema.u@amikom.ac.id.

Erwin Syahrudin is a graduate student in the magister of informatics program at Universitas Amikom Yogyakarta. His academic journey is marked by a strong passion for research and innovation in the fields of computer vision, artificial intelligence, software engineering, image processing, and machine learning. As an aspiring researcher, Erwin has actively engaged in various academic and collaborative projects aimed at addressing real-world challenges through technology. His research contributions are recognized on Google Scholar. In addition to his studies, he leverages his technical expertise and critical thinking skills to contribute to the advancement of these fields. He is also building his professional network and sharing his insights through LinkedIn https://id.linkedin.com/in/erwin-syahrudin-0a39131a5). With a dedication to continuous learning and a commitment to excellence, Erwin is poised to make significant contributions to the tech industry. He can be contacted at email: erwinsyahrudin@students.amikom.ac.id.

Anggit Dwi Hartanto is a lecturer at the Faculty of Computer Science, Universitas Amikom Yogyakarta. He holds a master's degree in informatics engineering from Universitas Amikom Yogyakarta and is currently pursuing his doctoral studies at the same institution. His research interests focus on natural language processing (NLP), data mining, and information systems. A dedicated researcher, he is an active member of IEEE and IEEE Systems. His academic profile can also be found on Google Scholar (Scholar ID: FnAiB08AAAAJ), where his contributions to the field continue to receive recognition. He can be contacted via email at anggit@amikom.ac.id.