2025 13th International Electrical Engineering Congress (iEECON) | 979-8-3315-4395-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/iEECON64081.2025.10987838

Automatic Crosswalk Detection and Counting Using YOLO and SAHI Techniques

Rapeepong Kunseewattana
Department of Electrical and Computer Engineering
Faculty of Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
s6601012610121@email.kmutnb.ac.th

Teerayut Sodaying
Department of Computer Engineering
Faculty of Engineering
Khon Kaen University
Khon Kaen, Thailand
teerayut.so@kkumail.com

Wilaiporn Lee
Department of Electrical and Computer Engineering
Faculty of Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
wilaiporn.l @eng.kmutnb.ac.th

Chatree Mahatthanajatuphat*
Department of Electrical and Computer Engineering
Faculty of Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
chatree.m@eng.kmutnb.ac.th

Vera Sa-ing*
Department of Electrical and Computer Engineering
Faculty of Engineering
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
vera.s@eng.kmutnb.ac.th

***Abstract*—The critical issue of pedestrian safety in large cities with high traffic density is focused on the effective detection and counting of pedestrians on crosswalks. This research proposes a deep learning model based on the You Only Look Once (YOLO) architecture, enhanced by the Slicing Aided Hyper Inference (SAHI) technique. The performance of several YOLO versions, including YOLOv8, YOLOv9, and YOLOv10, was evaluated with and without the integration of SAHI, using Accuracy, R-squared (*R^2*), and Mean Squared Error (*MSE*) as evaluation metrics. Our experimental results represent that, without SAHI, YOLOv9 achieved the highest accuracy of 37.52%, with an *R^2* value of 0.98 and an *MSE* of 19.36. When the SAHI technique was applied, YOLOv9+SAHI demonstrated a significant improvement by achieving an accuracy of 75.12%, maintaining an *R^2* value of 0.98, and reducing the *MSE* to 11.67. These findings highlight the effectiveness of SAHI in enhancing detection performance and reducing errors, particularly for YOLOv9, and improving the detection of small objects in complex environments. This study provides valuable insights into optimizing pedestrian detection systems, with YOLOv9 integrated with SAHI being the most effective model for pedestrian counting on crosswalks.**

Keywords*—Crosswalk detection, crosswalk counting, deep learning, convolutional neural network, YOLO

I. Introduction

Currently, pedestrian safety is an important issue that demands significant attention in large cities with high traffic density. One of the major concerns is the violation that occurs from the avoidance of traffic laws, such as vehicles driving over pedestrian crossings, which not only increases the risk of accidents but also reflects a lack of awareness regarding safety and respect for traffic regulations [1]. According to the World Health Organization (WHO) in 2023, traffic accidents involving vehicles are a leading cause of road fatalities, with cars being the most involved in such incidents. At the same time, 23% of global road traffic fatalities involve pedestrians, and the risk of pedestrians increases in restricted areas where traffic laws are frequently violated [2]. In recent years, object detection technology using the deep learning method has advanced significantly. The YOLO (You Only Look Once) model has indeed gained significant recognition methods for its ability to detect objects with high speed and accuracy. So, this deep learning model makes a popular choice for real-time applications, such as surveillance, autonomous driving, and pedestrian safety [3].

For this research, we see that the development of YOLO across its various versions demonstrates significant advancements in architecture and enhancements in capabilities, tailored to better suit specific applications. The researcher has chosen to use YOLOv8, which is suitable for general object detection with moderate computational requirements. YOLOv9 excels in resource-constrained environments, offering efficiency without sacrificing performance. YOLOv10 pushes the boundaries of real-time detection with its advanced architecture, making it the most versatile for complex applications. This approach effectively works alongside other models such as SAHI (Slicing Aided Hyper Inference) this is a technique designed to enhance the detection of small objects in large images, particularly when traditional object detection models struggle due to downscaling or resolution constraints [4]. SAHI works by using the original images, as shown in Fig. 1(a), and then slicing the input image into smaller overlapping patches. This allows the model to focus on finer details in each slice, as seen in Fig. 1(b) and 1(c). These slices are then processed individually through the object detection model, and the results are merged to provide improved detection accuracy Fig. 1(d).

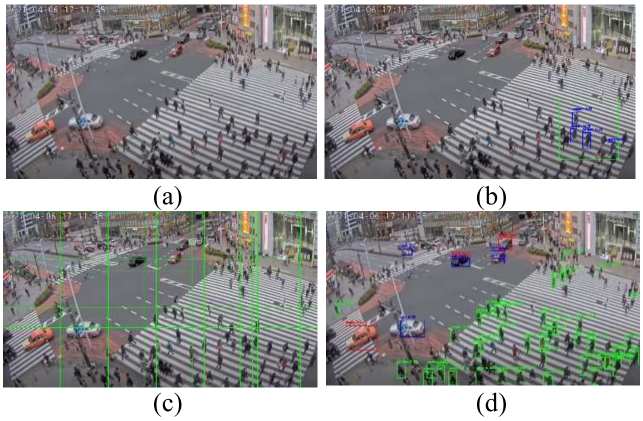
(a) (b)
(c) (d)

Fig. 1. These images represent the example of SAHI results by following (a) the original picture, (b) slicing the image, (c) overlapping patches, and (d) final detection.

* Corresponding author: vera.s@eng.kmutnb.ac.th, chatree.m@eng.kmutnb.ac.th

979-8-3315-4395-2/25/$31.00 ©2025 IEEE

This study will find the most of efficient deep learning model by comparing the performance of three models in detecting pedestrians crossing at zebra crossings using real-world data. The analysis will evaluate each compared model's capability to detect individuals, providing insights that could contribute to the development of more effective systems for identifying traffic violations in the future. Specifically, these insights could be used to improve traffic violation detection systems, making them more accurate and responsive. This includes enhancing the detection of traffic rule violations in complex and dynamic environments, as well as using technology to enable real-time warning or enforcement of traffic laws to improve safety in the future.

II. Methodology

The research was conducted with the objective of comparing the processing speed and accuracy efficiency of different versions of the YOLO model. In addition, the integration of SAHI techniques to enhance detection performance in high-resolution images will help identify the strengths and weaknesses of each version in terms of accuracy, speed, and practical applicability.

A. Object Detection Algorithm

YOLO is a real-time object detection system that revolutionized the field of computer vision. YOLO is designed to detect objects in images and videos with high speed and accuracy, making it particularly well-suited for applications requiring real-time performance. In our research, the compared algorithms will experiment by using the state-of-the-art of YOLO as follows.

1) YOLO Version 8

YOLOv8 is designed to enhance both speed and accuracy, utilizing the Cross Stage Partial Network (CSPNet) structure to reduce data loss during processing [5-6].

2) YOLO Version 9

YOLOv9 improves accuracy in detecting small objects and handling complex images. It introduces the Augmented Feature Pyramid Network (AFPN) to manage variations in lighting and camera angles [7-8].

3) YOLO Version 10

YOLOv10 integrates Vision Transformer (ViT) technology into its architecture, improving its ability to process and interpret intricate visual data with greater precision [9-11].

4) Slicing Aided Hyper Inference (SAHI)

SAHI is a technique designed to improve object detection in large images by slicing the image into smaller sections and processing each section separately. The results are then combined, which helps increase accuracy and reduce errors that arise from large images or objects that are hard to detect from a full image view [12-13].

B. Data Processing

This research began with using a 720p video clip (720 pixels in height and 1280 pixels in width) with a frame rate of 30 frames per second (fps). The frames from the acquired video were extracted by reducing the frame rate to 1 frame that was extracted by each 1 second to make it easier to extract clear images from each frame, which also facilitates the processing.

The video clip, now divided into many frames, was tested with different versions of the YOLO model to evaluate its performance and accuracy in counting objects present in each frame. This research tested various YOLO versions to find the one with the highest accuracy. In addition, our study will enhance the processing of large images and improve object detail recognition, the SAHI technique that was used with YOLO's processing. This technique allowed the proposed model to detect details in large images more clearly and accurately.

C. Evaluation Methods

For the purpose of evaluating the performance of the designed model in this research [14-15], the following metrics were utilized:

1) Accuracy (%)

Accuracy measures a model's ability to correctly classify instances, Calculated by the percentage of correct outcomes relative to the total number of instances.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Instances} \times 100\%$$

This metric represents the proportion of accurate predictions and is particularly suitable for evaluating the classification performance of object detection models like YOLO, as it reflects the overall correctness of predictions.

2) R-squared (R^2)

R^2 quantifies the extent to which the independent variables account for the variability observed in the dependent variable. It has a range between 0 and 1, with higher values signifying improved model accuracy.

$$R^2 = 1 - \frac{The\ residual\ sum\ of\ squares}{The\ total\ sum\ of\ squares}$$

This metric evaluates how well the model's predictions explain the variability in the ground truth. An R^2 value near 1 suggests that the predictions closely align with actual values, while lower values indicate weaker predictive accuracy. It is particularly effective for assessing regression tasks, including those enhanced by the SAHI technique.

3) Mean Squared Error (MSE)

MSE was used to measure the error between actual values (y_i) and predicted values ($\hat{y}_i$) in regression models. A lower *MSE* indicates higher model accuracy.

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

The variable n stands for the total number of data points. Mean Squared Error measures the average squared disparity between predicted and observed values, providing a clear representation of prediction error. Lower values of *MSE* denote better accuracy, as they indicate minimal deviations from actual results. This metric is indispensable for refining and evaluating regression performance across different models.

III. Experimental Results

This result presents the experimental results of the YOLO10, YOLO9, and YOLO8 models that were integrated with the SAHI technique to assess their predictive capabilities. These metrics include Accuracy (%), R-squared (R^2), and Mean Squared Error (*MSE*), each serving a distinct purpose in the analysis.

A. Comparing the YOLO Models

The overall experimental results as presented in Table I show the evaluation of three YOLO models (YOLOv8, YOLOv9, and YOLOv10) under two different conditions without using the SAHI technique and with using the SAHI technique. The models were measured using Accuracy, R-squared (R^2), and Mean Squared Error (*MSE*).

1) YOLOv8: Without SAHI, this model achieved an accuracy of 34.33%, an R-squared value of 0.98, and an *MSE* of 19.16. With SAHI, the performance improved significantly, achieving an accuracy of 64.43%, maintaining an R-squared value of 0.98, and reducing the *MSE* to 14.09.

2) YOLOv9: Without SAHI, this model achieved the highest Accuracy among the models at 37.52%, with an R-squared value of 0.98 and an *MSE* of 19.36. With SAHI, the Accuracy increased to 75.12%, which is the highest improvement across all models. The R-squared remained stable at 0.98, and the *MSE* reduced significantly to 11.67.

3) YOLOv10: Without SAHI, this model had the lowest performance in terms of accuracy, at only 17.80%. The R-squared value dropped to 0.72, and the *MSE* was significantly higher at 308.75. With SAHI, the accuracy improved substantially to 65.57%, and the R-squared value increased to 0.98. The *MSE* was greatly reduced to 13.87, indicating a significant performance gain.

From the overall experimental results, these results demonstrate that applying the SAHI technique enhanced the performance of all YOLO models in terms of accuracy and reduced the *MSE* while maintaining a consistent value of 0.98 for YOLOv8 and YOLOv9. Moreover, YOLOv9 applied SAHI that represented *MSE* less than the compared models.

B. The Proposed Model

From experimental results demonstrate that integrating SAHI techniques with all YOLO models enhances accuracy and reduces *MSE*. The SAHI technique plays a critical role in improving the models' ability to detect small or complex objects, leading to superior performance in object detection tasks. This advancement is particularly evident in YOLOv9, which exhibits enhanced capabilities in image processing and analysis, especially in applications requiring high resolution and precision in objects, as shown in Fig. 2.

YOLOv9 combined with the SAHI technique shows high potential for improving object detection performance. This approach is particularly effective in applications requiring precise differentiation between valid and invalid objects, making it valuable for research in advanced computer vision systems in Fig. 3.

TABLE I. The Overall Results of The Compared YOLO Models by Measuring The Accuracy, R-Squared, And Mean Squared Error Between Without And With SAHI Technique.

	Accuracy	***R²***	***MSE***
YOLOv8	34.33	**0.98**	19.16
YOLOv9	37.52	**0.98**	19.36
YOLOv10	17.80	0.72	308.75
YOLOv8 + SAHI	64.43	**0.98**	14.09
YOLOv9 + SAHI	**75.12**	**0.98**	**11.67**
YOLOv10 + SAHI	65.57	**0.98**	13.87

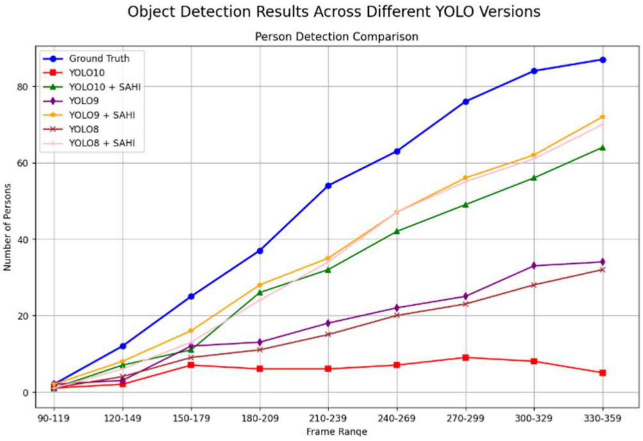

Fig. 2. Comparison results of person detection and counting in compared YOLO versions and YOLO + SAHI with the ground truth.

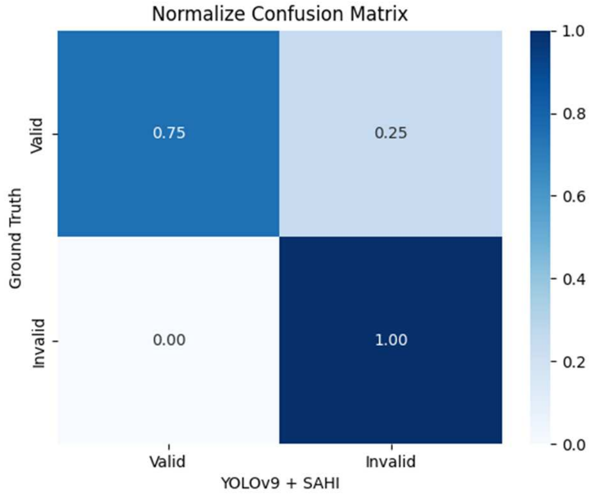

Fig. 3. Confusion Matrix of the best experimental result was achieved by the YOLOv9 model integrated with the SAHI technique.

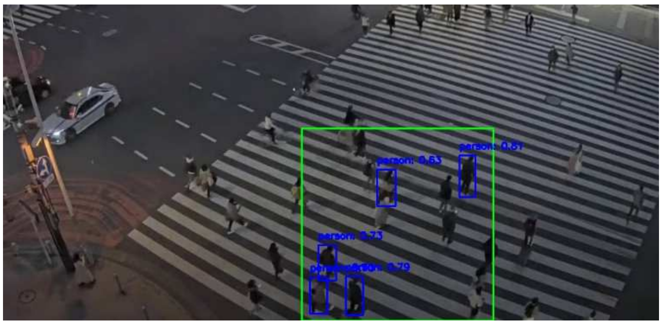

Fig. 4. The proposed model can detect and count the two people overlapping in the image.

The experimental results indicate that the two individuals overlap in the image shown in Fig. 4. The model often detects only one person. This occurs due to the Intersection over Union (IoU) value between the model's predictions and the ground truth being too high. As the experimental result, the original YOLO model fails to distinguish between the two individuals and instead treats the overlapping predictions as a single detection. This leads to a decrease in the accuracy of person detection when significant overlap is present. However, the model is designed to improve the effectiveness of object detection, focusing on situations involving small-sized or overlapping objects. By utilizing the SAHI technique, which divides large images into smaller, analyzable tiles, the model achieves higher accuracy while reducing *MSE*, ensuring precise and reliable detection.

IV. Conclusion

From the experimental results, this study demonstrated a significant enhancement in object detection performance by integrating the SAHI technique with various YOLO models that consist of YOLOv8, YOLOv9, and YOLOv10. The performance evaluation, based on key metrics such as accuracy, R-squared, and *MSE*, showed a notable improvement in detection accuracy and error reduction after incorporating SAHI. Without SAHI, YOLOv9 performed the best in terms of accuracy, achieving 37.52%, with an R-squared value of 0.98 and an *MSE* of 19.36. However, SAHI was applied to YOLOv9 and the accuracy increased significantly to 75.12%, with the R-squared value maintaining 0.98 and the *MSE* dropping to 11.67. This highlights the critical role of SAHI in enhancing model performance, especially in detecting small objects and complex image scenarios. In contrast, YOLOv8 and YOLOv10 also showed improvements when using SAHI, achieving accuracy levels of 64.43% and 65.57%, respectively, with corresponding reductions in *MSE*. Despite YOLOv10's lower baseline performance without SAHI, the integration of SAHI greatly improved its detection capabilities, emphasizing its potential for real-world applications where image resolution and precision are crucial. Overall, the study underscores the potential of combining the SAHI technique with YOLO models to optimize object detection systems, particularly in applications involving high-resolution images and complex object scenarios. These findings provide valuable insights for further research in pedestrian safety, traffic monitoring, and related fields where real-time and accurate object detection is essential.

Acknowledgment

This research budget was allocated by National Science, Research and Innovation Fund (NSRF), and King Mongkut's University of Technology North Bangkok (Project no. KMUTNB-FF-68-B-15). In addition, authors would like to express our gratitude to the Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok for supporting this research.

References

[1] National Association of City Transportation Officials (NACTO), "Urban Street Design Guide," 2013. [Online]. Available: https://nacto.org/publication/urban-street-design-guide/

[2] World Health Organization, "Global status report on road safety 2023," 2013. [Online]. Available: https://www.who.int/publications/b/68866

[3] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *in the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.

[4] F. C. Akyon et al., "Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection," *in the 2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 966-970.

[5] R. Varghese and S. M., "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," *in the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1-6.

[6] A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," *in IEEE Access*, vol. 11, pp. 35479-35516, 2023.

[7] M. Bakirci and I. Bayraktar, "YOLOv9-Enabled Vehicle Detection for Urban Security and Forensics Applications," *in the 2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, 2024, pp. 1-6.

[8] M. Bakirci and I. Bayraktar, "Refining Transportation Automation with Convolutional Neural Network-Based Vehicle Detection via UAVs," *in the 2024 International Russian Automation Conference (RusAutoCon)*, 2024, pp. 150-155.

[9] F. Y. A'la, N. Firdaus, Hartatik and H. Imaduddin, "Precision in Safety: YOLOv9 vs. YOLOv10 for Helmet Image Detection," *in the 2024 International Visualization, Informatics and Technology Conference (IVIT)*, 2024, pp. 159-164.

[10] N. Jegham et al., "Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors," *in arXiv*, 2024, pp. 1-6.

[11] A. Wang et al., "YOLOv10: Real-Time End-to-End Object Detection," *in the 38th Conference on Neural Information Processing Systems (NeurIPS 2024)*, 2024, pp. 1-21.

[12] Y. Zheng et al., "YOLOv5s FMG: An Improved Small Target Detection Algorithm Based on YOLOv5 in Low Visibility," *in IEEE Access*, vol. 11, pp. 75782-75793, 2023.

[13] M. Muzammul et al., "Enhancing UAV Aerial Image Analysis: Integrating Advanced SAHI Techniques With Real-Time Detection Models on the VisDrone Dataset," *in IEEE Access*, vol. 12, pp. 21621-21633, 2024.

[14] V. Sa-ing et al., "Multiscale adaptive regularisation Savitzky–Golay method for speckle noise reduction in ultrasound images," *in the IET Image Processing*, vol. 12, pp. 105-112, 2018.

[15] S. Pingali and J. Segen, "Performance evaluation of people tracking systems," *in the Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96*, 1996, pp.