ATCSJ Open Access
Applied Technology and Computing Science Journal
p-ISSN 2621-4458 | e-ISSN 2621-4474

UNUSA
Available online to journal2.unusa.ac.id
UNUSA
S4-Accredited – SK No.79/M/KPT/2023
Journal Page is available to https://journal2.unusa.ac.id/index.php/ATCSJ

A Dual-Stage Hybrid Vision Framework Using YOLOv8n-Canny Edge Detection for Real-Time Railway Trespassing and Intrusion Monitoring

Adiratna Ciptaningrum[1,*], Mohammad Erik Echsony[2], R. Akbar Nur Apriyanto[3]

[123]*State Polytechnic of Madiun, Indonesia*
Jl. Serayu No.84, Pandean, Taman, Madiun
[1*]adiratna@pnm.ac.id , [2]erik_sony@pnm.ac.id , [3]akbar@pnm.ac.id

Article history:

Received 20 January 2026
Revised 10 February 2026
Accepted 20 February 2026
Available online 24 February 2026

Keywords:

Railway Safety,
Intrusion,
Trespasser,
YOLOv8n,
Canny Edge Detection.

Abstract

Intrusion and trespasser detection on railway tracks is a crucial safety measure to prevent accidents and maintain operational reliability. This study proposes a hybrid vision-based approach that integrates YOLOv8n, a lightweight real-time object detection model, with the Canny edge detection algorithm to identify and classify unauthorized objects and individuals on railway tracks. In this context, intrusions refer to inanimate objects such as rocks, fallen trees, or construction materials obstructing the tracks, whereas trespassers refer to humans or other living beings engaging in unauthorized activities near or on the railway line. YOLOv8n is employed as a single-stage detector to localize and classify objects, while Canny edge detection is applied to enhance object contours and improve shape-based differentiation between intrusion and trespasser categories. Experimental results show an average accuracy of 52.37%, indicating moderate detection performance. Although the accuracy remains limited, the findings demonstrate the potential of combining deep learning and traditional image processing techniques to develop an automated monitoring system that supports railway safety and surveillance applications. Further optimization of the dataset, model tuning, and feature enhancement are recommended to improve detection performance.

I. INTRODUCTION

Rail transportation is a key component of Indonesia's mobility and logistics system. Passenger numbers reached approximately 119.8 million in 2022, representing a 42% increase from 2021 [1], leading to higher train operation frequency and a corresponding rise in rail-related incidents [2]. Indonesian railway accident reports (KNKY) identify derailments as the most frequent incidents, primarily caused by infrastructure

[1*] Corresponding author

deficiencies [3]. External factors, including trespassing and tracks intrusion by humans, animals, or foreign objects, further increase derailment risks, particularly on open-track sections such as the Madiun line where physical barriers are limited [4] [5]. Current monitoring relies on manual surveillance and passive measures, which are ineffective across Indonesia's 6,300 km wailway network, underscoring the need for automated, real-time vision-based intrusion detection systems [6] [7] [8] [9] [10].

Several intelligent vision-based methods have been proposed for detecting trespassing and intrusion along railway tracks, particularly in high-risk open-access areas. Among these, real-time object detection using deep learning especially the Only You Look Once (YOLO) architecture has shown strong potential [11]. The lightweight YOLOv8n model is well suited for edge deployment (e.g., Jetson Nano) due to its high inference speed and low computational cost, while maintaining reliable detection of humans, animals, and other intruding objects in railway environments. Zhang et al. demonstrated the effectiveness of YOLOv8n for monitoring pedestrian activity at unmanned crossing under varying light conditions [12]. Additionally, Canny Edge Detection is often integrated to enhance contour extraction and reduce false positives, particularly for thin or partially occluded objects on the track [13].

Railway intrusion detection has shifted from manual patrols to vison-based monitoring using mobile inspection robots. In this study, continuous video from an onboard camera is processed in real time using the lightweight YOLOv8n model to detect unauthorized objects along vulnerable tracks segments. To mitigate visual noise in open-track environments, Canny Edge Detection is integrated to enhance object boundaries and reduce false positives. Compared with infrared or radar systems, the proposed YOLOv8n-Canny pipeline provides a cost-effective and scalable solution for continuous intrusion monitoring in derailment-prone areas.

One method of monitoring unauthorized access and foreign object presence on railway tracks can be achieved through real-time object detection in image frames. Convolutional Neural Network (CNN)-based models are widely used due to their high accuracy and ability to extract spatial features from visual data [14]. Among these, YOLO (You Only Look Once) stands out as a single-stage object detection algorithm that balances accuracy and speed effectively. Several versions of YOLO have been developed, such as YOLOv3, YOLOv5, and the more recent YOLOv8 [15]. YOLOv8n, a lightweight variant, is specifically designed to operate efficiently in constrained environments while maintaining real-time performance. Studies have shown YOLOv8n to perform competitively in surveillance applications, even when deployed on mobile platforms [16].

To improve object contour extraction under noisy visual conditions, Canny Edge Detection is integrated as a pre-processing step to enhance boundary clarity and reduce false detections between railway structures and potential intrusions such as pedestrians or animals [17]. Compared to two-stage detectors (e.g., Faster R-CNN), which incur high computational cost and latency due to region proposal mechanism [18], YOLOv8n employs a single-stage architecture with a single forward pass, making it well suited for real-time deployment on mobile rail inspection robots. Prior studies have demonstrated the effectiveness of YOLO-based models in infrastructure monitoring, including surface damage detection on bridges [21], defects identification in sewer pipes [22], and crack detection on concrete structures under visual noise [23]. Building on these findings, YOLOv8n due to its compactness and inference speed offers a practical solution for real-time railway intrusion monitoring. When combined with Canny Edge Detection, the system becomes more robust against illumination variation, background clutter, and partial occlusion.

Despite that growing adoption of deep learning-based detectors, real-time railway safety applications remain limited due to diverse intrusion object characteristics and the trade-off between detection accuracy and real-time performance. The proposed YOLOv8n-Canny pipeline addresses these challenges by enabling efficient and continuous monitoring in intrusion-prone and derailment-sensitive track segments [19] [20].

Although deep learning-based object detection has been widely applied in infrastructure monitoring, most railway safety studies focus on fixed surveillance systems or accuracy-oriented models that are not optimized for real-time deployment on mobile inspection robots. Research addressing narrow-gauge railways (1,067 m) which are more vulnerable to derailments due to limited clearance remains scarce, particularly in rural open-track environments such as the Madiun section, where physical barriers are minimal and trespassing is frequent. Furthermore, existing approaches generally rely on end-to-end detectors without explicitly leveraging edge-based enhancement to mitigate visual clutter caused by illumination variation, motion blur, and complex backgrounds. In addition, prior studies rarely distinguish between static intrusions and living trespassers, and the interaction between object density, inference latency, and detection accuracy in real-time railway monitoring has not been systematically analyzed.

II. METHODS

This research proposes a hybrid vision-based framework that integrates the lightweight YOLOv8n object detector with Canny Edge Detection for real-time railway intrusion monitoring. YOLOv8n is selected for its compact architecture and fast inference, enabling robust detection in dynamic environments. Canny Edge Detection is applied in a dual-stage pipeline during preprocessing to enhance contour information and during postprocessing on detecting intrusions and trespassers under complex visual conditions. The system architecture is illustrated in Figure 1.

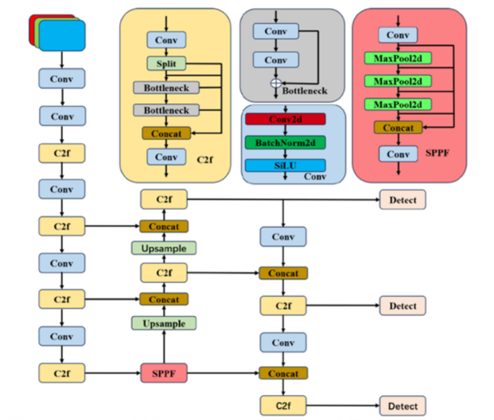

Figure 1. YOLOv8n model architecture.

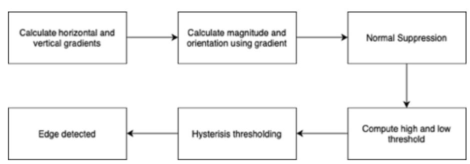

Figure 2. Canny Edge detection architecture.

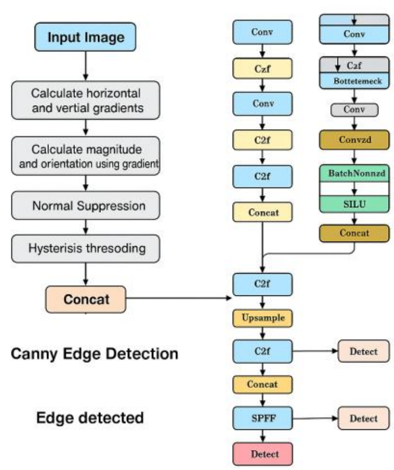

Figure 3. Hybrid vision-based of YOLOv8n and Canny Edge Detection.

A. Data Acquisition and Canny Edge Detection

The primary dataset consists of video sequences captured using a mobile rail inspection robot deployed along the Madiun railway line in East Java, Indonesia. The proposed hybrid method begins with Canny Edge Detection applied as a preprocessing step. This technique enhances edge feature to extract structural boundary information from input images, which can later enhance the object localization capabilities of the YOLO detector. The process involves a series of steps that convert raw pixel intensity variations into a refined edge map.

Initially, the input image is smoothed using a Gaussian filter to reduce noise. Following this, the horizontal and vertical intensity gradients of the smoothed image are computed using convolutional operators such as Sobel filters. Let G_x and G_y represent the horizontal and vertical gradients, respectively. The gradient magnitude M and orientation θ at each pixel location are then calculated as:

$$M(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (1)$$

$$\theta = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \quad (2)$$

To refine the detected edges, non-maximum suppression is applied to thin the edges by retaining only local maxima in the direction of the gradient. This step eliminates spurious responses and ensures that edges are crisp and one-pixel wide.

Next, double thresholding is employed to classify edge pixels into strong, weak, and non-edge categories using two threshold values, a high threshold $T_H$and a low threshold T_L, where $T_H > T_L$. Finally, hysteresis thresholding connects weak edge pixels to strong ones if they are part of a continuous edge, resulting in the final binary edge map E. This edge map captures salient object boundaries, which are then used in the subsequent fusion step with the YOLO feature extractor.

B. Edge-Feature Fusion Mechanism

To exploit structural information from the Canny edge map, edge features are fused with the raw input or intermediate YOLOv8n backbone features, enabling joint learning of low-level edge cues and high-level semantic representations.

Let the input image be denoted as $I \in \mathbb{R}^{H\times W\times C}$, where $H,\ W,\ C$ are the height, width, and number of channels, respectively. The Canny edge detector produces a binary edge map $I \in \mathbb{R}^{H\times W\times C}$, which highlights the boundaries within the image. To integrate this information, the edge map is concatenated with the input image along the channel dimension to form a fused tensor:

Let the input image be denoted as $I \in \mathbb{R}^{H\times W\times C}$, where $H,\ W,\ C$ are the height, width, and number of channels, respectively. The Canny edge detector produces a binary edge map $I \in \mathbb{R}^{H\times W\times C}$, which highlights the boundaries within the image. To integrate this information, the edge map is concatenated with the input image along the channel dimension to form a fused tensor:

$$F_0 = Concat(I, E) \quad (3)$$

Alternatively, if the fusion is applied within the network architecture, the edge map can be concatenated with early convolutional features $F_1 \in \mathbb{R}^{H\times W\times C}$, yielding:

$$F_{fused} = Concat\ (F_1, E') \quad (4)$$

where $'$ is a resized version of the edge map to match the spatial dimensions of F_1 using bilinear interpolation or nearest-neighbor scaling.

The fused feature tensor F_{fused} is then passed through the subsequent layers of the YOLO backbone. This integration enables the model to exploit edge-specific localization cues alongside conventional feature extraction processes, improving detection performance especially in scenarios with poor contrast, occlusion, or fine structural details.

C. YOLOv8n-Based Object Detection

YOLOv8n is the lightweight variant of the YOLOv8 family, is employed as the backbone of the object detection module due to its efficiency and high inference speed, making it suitable for edge deployment.

The backbone consists of an arrangement of convolutional layers and Cross-Stage Partial (C2f) modules that extract hierarchical features from the input tensor. The C2f module enhances gradient flow and reduces computational complexity by splitting and merging feature paths.

The neck includes a combination of up-sampling, down-sampling, and concatenation operations that aggregate features at multiple spatial scales, preserving both coarse and fine details. This structure is crucial for detecting objects of varying sizes.

The head ouputs the final predictions, which consist of bounding box coordinates, object scores, and class probabilities. Each predicted bounding box b_i is parameterized by its center coordinates (x_i, y_i), width w_i, height h_i, and confidence score c_i. The confidence score c_i is defined as the product of the object probability and the highest class probability:

$$c_i = P_{obj}(b_i)\ .\max\left(P_{cls}(b_i)\right) \tag{5}$$

The network is optimized using composite loss function:

$$\mathcal{L} = \lambda_{box}\ .\mathcal{L}_{box} + \lambda_{obj}\ .\mathcal{L}_{obj} + \lambda_{cls}\ .\mathcal{L}_{cls} \tag{6}$$

Where $\mathcal{L}_{box}$ represents the bounding box regression loss (typically CIoU or DIoU loss), $\mathcal{L}_{obj}$ is the binary cross-entropy loss for objectness, and $\mathcal{L}_{cls}$ is the multi-class classification loss. The scalar weights λ_{box}, λ_{obj}, and λ_{cls}, control the contribution of each loss component.

Through this architecture, YOLOv8n enables efficient and accurate detection, particularly when enhanced with edge-based feature cues as proposed in the earlier fusion module.

D. Evaluation Metrics

To quantitatively assess the performance of the proposed hybrid edge-aware object detection system, standard object detection evaluation metrics are utilized. The primary metrics include Prcission, Recall, F-1 Score, and mean Average Precission (mAP), which offer comprehensive insights into detection accuracy and robustness. Precission measures the proportion of correctly predicted positive instances among all predicted positives:

$$Precision = \frac{TP}{TP+TF} \tag{7}$$

Where TP denotes True Positive and FP denotes False Positive.

Recall, on the other hand, measures the completeness of the detector. It is the ratio of correctly identified positive samples TP to the total number of actual positive instances, including the ones the model missed (False Negative, FN).

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

High Recall indicates that most objects are detected, though it may increase false positives. To balance precision and recall, the F-1 Score is used as their harmonic mean, providing a single measure of detection accuracy and completeness:

$$F1 = 2\ .\frac{Precision\ .\ Recall}{Precision+Recall} \tag{9}$$

III. RESULTS AND DISCUSSIONS

The hybrid detection framework combining Canny Edge Detection and YOLOv8n was evaluated on its ability to detect two key event categories: intrusions and trespassers. Performance was tested on a dataset featuring varying object counts and scene complexities. Detection accuracies ranged from 63.25% to 75.22%, reflecting sensitivity to object quantity and environmental factors, as detailed in Table 1.

Table 1. Perform Indicator of Testing.

No	Perform Indicator	
	Testing Result	***Accuracy***
0.	Intrusion =1 and Trespasser = 0	65,4%
1.	Intrusion = 0 and Trespasser = 1	67,42%
2.	Intrusion = 0 and Trespasser = 1	63,25%
3.	Intrusion = 1 and Trespasser = 3	73,71%
4.	Intrusion = 2 and Trespasser = 1	75,22%
5.	Intrusion = 3 and Trespasser = 2	64,65%
6.	Intrusion = 1 and Trespasser = 3	71,29%

As we can see in Figure 4, the hybrid detection framework combining Canny Edge Detection and YOLOv8n was evaluated based on its performance in detecting two primary event categories, intrusions and trespassers. The results were assessed and plotted using a dataset that included variations in the number of these objects across even scenarios.

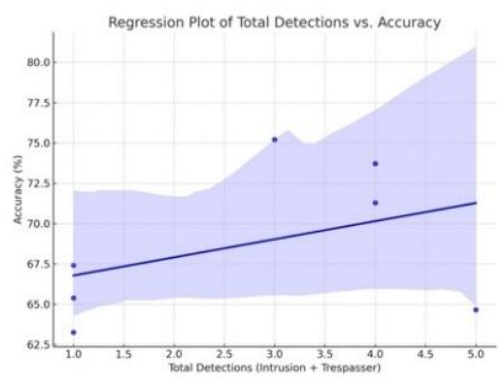

Figure 4. Regression plot of total detection vs accuracy.

This research confirms that hybrid detection methods enhance railway safety by combining deep learning-based object detection with edge-based feature extraction. The Pearson correlation coefficient (r) is used to quantify the linear relationship between the number of detected objects and model accuracy in intrusion and trespasser detection.

Table 2. Perform Indicator of testing.

Jumlah Intrusion	Jumlah Trespasser	Akurasi (%)	Preprocess (ms)	Inference (ms)	Postprocess (ms)
1	0	65.4	8	804.4	8.1
0	1	67.42	6.3	800.1	6.4
0	1	62.25	11.7	823.7	7.9
1	2	73.71	7.6	844.4	10.8
2	1	75.22	9.7	789.6	11.8
3	3	64.65	9.7	1037.8	19.6
1	3	71.29	5.7	788	14.4

The Pearson correlation coefficient r is calculated using the following equation:

$$r = \frac{\sum(X_i - X)(Y_i - Y)}{\sqrt{\sum(X_i - X)^2}\sqrt{\sum(Y_i - Y)^2}} \tag{10}$$

Table 3. The Pearson Correlation Between The Number of Intrusion, The Number of Trespasser, And Accuracy.

Total Intrusion	Total Trespasser	Accuracy (%)	Correlation with Accuracy (Pearson)
1	0	65.4%	0.17 (weak)
0	1	67.42%	0.45 (moderate)
0	1	63.25%	0.45 (moderate)
1	2	73.71%	0.41 (moderate)
2	1	75.22%	0.41 (moderate)
3	3	64.65%	0.41 (moderate)
1	3	71.29%	0.41 (moderate)

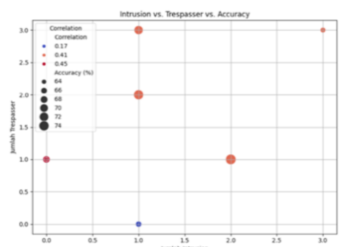

Figure 5. Intrussion vs trespasser vs accuracy.

Figure 5 illustrates the relationship between object composition and detection performance. The YOLOv8n-Canny framework improves detection accuracy to 63.25 – 75.22%. correlation analysis shows a moderate positive relationship between the number of trespassers and accuracy (r = 0.45), reflecting YOLOv8n’s strong human recognition capability. The total object count also exhibits a moderate correlation with accuracy (r = 0.41), indicating that richer spatial contexts support more robust pattern recognition. The correlation between intrusion count and accuracy is weak (r = 0.17), indicating that static objects often lack strong edge features for reliable detection. While Canny edge detection assists initial segmentation, low-contrast contours limit YOLOv8n’s ability to accurately identify static intrusion.

Static intrusion often exhibit low-contrast edges, making them harder for YOLOv8n to detect with high confidence. Inference time ranges from 788 to 1037.8 ms and increases with object complexity, while pre- and post-processing remain relatively stable with slight overhead under dense edge conditions. Overall, the results indicate that the hybrid model improves railway safety by detecting trespassers more reliably than static objects, though intrusion detection can be further enhanced through adaptive edge optimization or depth-based segmentation. These findings confirm that combining edge-based features with deep learning outperforms single-method approaches and offers a promising foundation for robust real-time railway surveillance systems.

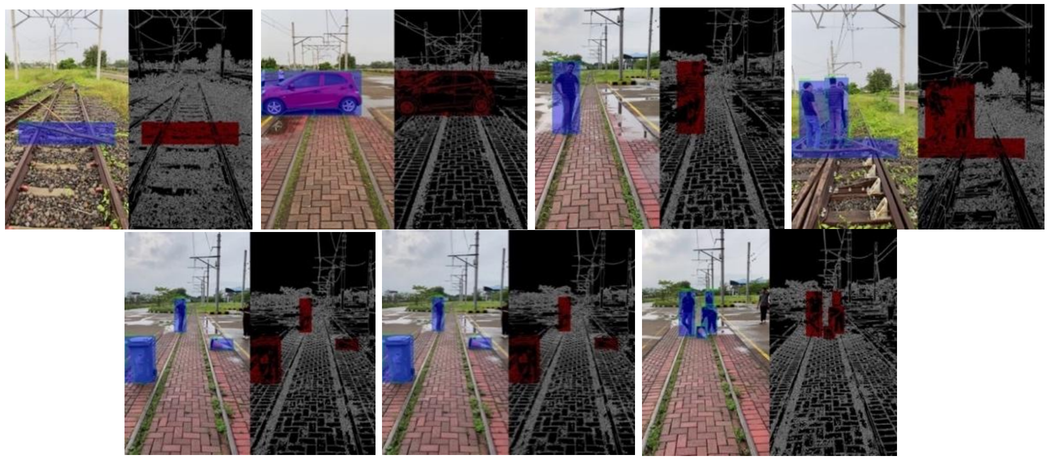

Figure 6. Fused of YOLOv8n and Canny Edge Detection from left to right (Intrusion = 1; Trespasser =1; Trespasser =1; Intrusion=1 Trespasser=3; Intrusion=2 Trespasser=1; Intrusion=3 Trespasser=2; Intrusion=1 Trespasser=3).

Table 4. Metrics Evaluation.

No	TP	FP	FN	Precision	Recall	F1 Score
0.	0.6540	0.3460	0.3460	0.6540	0.6540	0.6540
1.	0.6742	0.3258	0.3258	0.6742	0.6742	0.6742
2.	0.6225	0.3775	0.3775	0.6225	0.6225	0.6225
3.	2.2113	0.7887	0.7887	0.7371	0.7371	0.7371
4.	2.2566	0.7434	0.7434	0.7522	0.7522	0.7522
5.	3.8790	2.1210	2.1210	0.6465	0.6465	0.6465
6.	2.8516	1.1484	1.1484	0.7129	0.7129	0.7129

Tabel 5. Performance Comparison and Correlation Summary

Metric	YOLOv8n	Hybrid (Fused Canny + YOLOv8n)	Remarks
Average Precision	0.63	0.69	Hybrid improves precision (fewer false positives)
Average Recall	0.80	0.45	YOLOv8 captures more true positives (but with more false positive)
F1 Score Range	0.65 – 0.78	0.65 – 0.78	Stable and consistent across both methods
Correlation (F1 vs Inference Time)	-	r = -0.37	Longer inference time tends to reduce accuracy
Correlation (F1 vs Total Detection)	-	r = +0.37	Slight improvement with higher object counts

Figure 7 shows that the YOLOv8n-Canny fused model consistently outperforms the baseline YOLOv8n in precision across most confidence thresholds. The fused method achieves a peak precision of ~0.75 and a mean precision of -.69, compared to 0.67 and 0.63 for the baseline, respectively. The improvement is most pronounced at mid-level confidence scores (0.5 – 0.8), indicating reduced false positives and enhanced boundary discrimination from edge feature integration.

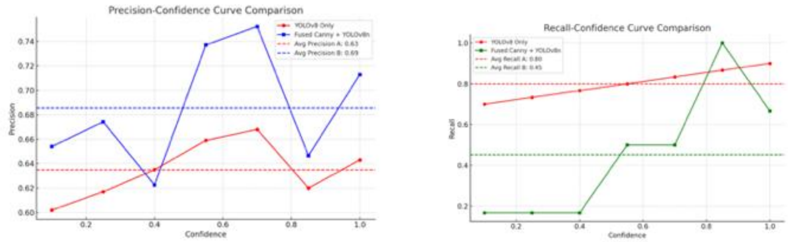

Figure 7. Precision-Confidence curve comparison and Recall-Confidence curve comparison.

In Figure 7, the Recall-Confidence Curve compares the performance of YOLOv8n alone (red line) with the fused method using Canny Edge + YOLOv8n (green line). The red line shows a steady and high recall, averaging 0.80, which means YOLOv8n detects most objects correctly. In contrast, the fused method has a lower and more unstable recall, averaging only 0.45. Although it reaches a perfect recall (1.0) at one point, it drops sharply at other confidence levels. This suggests that the Canny Edge fusion may cause the model to miss some objects, especially when the object edges are not clearly detected. In short, while the fusion improves precision, it reduces the ability to detect all objects.

The F1 Score per Detection Case chart shows how well the fused Canny + YOLOv8n method balances precision and recall across different test samples. The F1 scores are consistently high, mostly ranging between 0.65 and 0.78, indicating stable and reliable detection performance overall. This consistency suggests that, despite lower recall, the fused method still maintains a good trade-off between detecting objects correctly and avoiding false positives. It proves especially useful in scenarios where both accuracy and reliability are important across various cases.

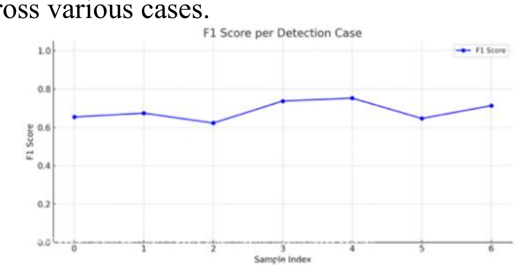

Figure 8. F-1 Score per detection case.

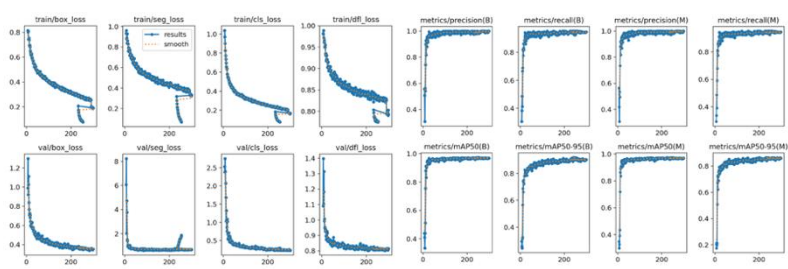

Figure 9. Result of the training model YOLOv8n.

The training results indicate excellent performance of the YOLOv8n model enhanced with Canny Edge Detection for intrusion and trespasser detection tasks. High precision and recall values (close to 1.0) across both bounding box (B) and mask (M) evaluations demonstrate the model's strong capability to accurately detect and classify intruders while minimizing false alarms. The steadily decreasing loss curves and near-perfect mAP scores (mAP50 and mAP50-95) further confirm that the model learned effectively without overfitting. The integration of Canny Edge Detection likely contributed by enhancing object boundaries, allowing YOLOv8n to focus better on areas of interest, especially in complex or low contrast scenes. Overall, the results suggest that this hybrid approach is highly reliable for real-time security and surveillance applications.

 E-ISSN: 2621-4474, ISSN:2621-4458.

IV. Conclusions And Recommendations

Ensure In this research, a hybrid object detection framework was proposed by integrating Canny Edge Detection with the YOLOv8n deep learning model to improve robustness of visual surveillance system in detecting intrusions and trespassing events. The experimental evaluation revealed that the fused method improved average precision from 0.63 (YOLOv8n only) to 0.69, indicating a reduction in false positive detections. This improvement suggests the edge-based preprocessing contributes to more accurate object localization and filtering of irrelevant background features.

Ensure In this research, a hybrid object detection framework was proposed by integrating Canny Edge Detection with the YOLOv8n deep learning model to improve robustness of visual surveillance system in detecting intrusions and trespassing events. The experimental evaluation revealed that the fused method improved average precision from 0.63 (YOLOv8n only) to 0.69, indicating a reduction in false positive detections. This improvement suggests the edge-based preprocessing contributes to more accurate object localization and filtering of irrelevant background features.

Correlation analysis showed a moderate negative relationship (r = -0.37) between inference time and F1 Score, suggesting that prolonged inference may slightly degrade performance. A positive correlation (r = +0.37) between the number of detected objects and F1 Score indicates a favorable response in more populated scenes. Overall, the integration of Canny Edge Detection with YOLOv8n enhance precision while preserving stable performance metrics, making it viable solution for applications requiring accurate and efficient detection with minimized false positives.

V. Acknowledgments

This research was supported by the RIIM LPDP Grant and BRIN, grant number 103/IV/KS/07/2025 and 5851/PL19/HK.04.00/2025. We also thanks to the Politeknik Negeri Madiun, Politeknik Perkapalan Negeri Surabaya, and PT Bilzatama Inovasi Teknologi.

VI. References

[1] P. KAI, "Train User Volume Increases 42% in Semester I 2022 [Press Release – in Indonesian]. Indonesia Railway Company PT KAI," KAI, [Online]. Available: https://www.kai.id/information/full_news/5379- volume-pelanggan-kereta-api-naik-42-pada-semester- i-2022. [Accessed May 2025].

[2] Luthfi, K. M., Sugiana, A., & Suratman, F. Y., "Broken rail detection system using laser," in IOP Conference Series: Materials Science and Engineering.

[3] Buku Statistik Investigasi Kecelakaan Transportasi KNKT 2021, Pelayanan Investigasi dan Kerjasama Bagian Data, Informasi, dan Humas Komite Nasional Keselamatan Transportasi., Komite Nasional Keselamatan Transportasi, 2021.

[4] Indonesia Transportation Safety Committee (KNKT), "Book of Transportation Accident Investigation Statistics in 2021," 2022.

[5] Lai, J., Xu, J., Wang, P., Yan, Z., Wang, S., Chen, R., & Sun, J, "Numerical investigation of dynamic derailment behavior of railway vehicle when passing through a turnout," Engineering Failure Analysis, no. 121, 105132, 2021.

[6] Yuliang Zhao, Zhiqiang Liu, Dong Yi, Xiaodong Yu, Xiaopeng Sha, Lianjiang Li, Hui Sun, Zhikun Zhan, and Wen Jung Li, "A Review on Rail Defect Detection Systems Based on Wireless Sensors," Sensors, vol. 22, no. 6409, 2022.

[7] Aydin, I., Akin, E., & Karakose, M., "Defect classification based on deep features for railway tracks in sustainable transportation," Applied Soft Computing, no. 111(107706), p. 1–14, 2021.

[8] Fahmi, "New Measuring Train from Switzerland Arrives in Jakarta, Here's What It Looks Like [Internet-in Indonesian]," 2022.

[9] Rizaty MA, "evelopment of Railway Track Length in Indonesia for 2016-2020 [Internet-in Indonesian]. 2021 [cited 2023 Apr 1]," [Online]. Available:

https://databoks.katadata.co.id/datapublish/2021/10/26 /indonesia-miliki-rel-kereta-sepanjang-632-juta- meter-pada-2020.

[10] "Wikipedia. List of countries by rail transport network size - Wikipedia [Internet]," [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_byrail_transport_network_size#cite_note-40.

[11] Adiratna Ciptaningrum, Andhika Putra Widyadharma, Imam Junaedi, Rahayu Mekar Bisono, R. Akbar Nur Apriyanto, R. Gaguk Pratama Yudha, Mohammad Erik Echsony, Larissa Kartika Putri, Edo Zulmi Faikhsan, "Design of YOLOv5 Medium as Unmanned Rail Inspection in Braking Control System Based on Computer Vision," International Journal of Science, Engineering and Information Technology, vol. 8, no. 2, pp. 467-474, 2024.

[12] X. Zhang, Y. Li, J. Wang, and L. Chen, "Real-time pedestrian monitoring at unmanned railway crossings using YOLOv8 under dynamic lighting," IEEE Trans. Intell. Transp. Syst, vol. 24, no. 7, p. 5678–5689, July 2023.

[13] H. Liu, Y. Zhang, and K. Zhou, "Enhancing object detection performance for railway safety using Canny edge-assisted deep learning models," . Rail Transp. Plan. Manag, vol. 22, p. 100354, 2022.

[14] A. Kumar, R. Singh, and D. Patel, "Real-time detection of foreign objects and intrusions on railway tracks using CNN-based models," IEEE Access, vol. 9, p. 123456–123467, 2021.

[15] Adiratna Ciptaningrum, Wahyu Pribadi, Dirvi Eko Juliando, Rakhmad Gusta, Edo Zulmi Faikhsan, "Design of smart human following on rail inspection using human pose estimation marker-less motion capture based on blazepose," Journal Geuthee of Engineering and Energy (JOGE), vol. 2, no. 2, pp. 106 - 118, 2023.

[16] J. Redmon and A. Farhadi, "YOLOv8: Real-time object detection in constrained environments," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2023.

[17] T. H. Nguyen, M. N. Bui, and Q. D. Tran, "Edge-enhanced object detection in railway surveillance using Canny-based pre-processing," Expert Syst. Appl., vol. 23, p. 117634, 2022.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Inf. Process. Syst. (NeurIPS), vol. 28, 2015.

[19] M. Ali, L. Zhao, and W. Tan, "Lightweight object detection for railway intrusion monitoring using YOLOv8n and edge enhancement techniques," Sensors, vol. 23, no. 12, pp. 5123, 2023, 2023.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.

[21] L. Zhang, Y. Chen, and H. Wang, "Bridge surface damage detection using an improved YOLOv3 with batch normalization and focal loss,," Autom. Constr, vol. 128, p. 103764, 2021.

[22] Z. Yin, Y. Liu, and T. Huang, "utomatic sewer defect detection using YOLOv3-based deep learning," Tunnelling and Underground Space Technol, vol. 98, p. 103249, 2020.

[23] J. Deng, L. Xu, and Y. Han, "Surface crack detection on concrete structures with graffiti using YOLOv2," Struct. Control Health Monit, vol. 26, no. 9, p. e2405, 2019.