Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 2, November 2025, pp. 1129~1140
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i2.pp1129-1140 ❒ 1129

Improved YOLOv8 for rail squat detection and identification

Van-Dinh Do, Phuong-Ty Nguyen, Minh-Tuan Ha
Department of Electrical, Sao Do University, Hai Duong, Vietnam

Article Info

Article history:

Received Nov 11, 2024
Revised Sep 3, 2025
Accepted Oct 15, 2025

Keywords:

Deep learning
Railway inspection system
Object detection
Squat defect
Yolov8

ABSTRACT

Rail transport plays a vital part in the country's economy by ensuring the safe movement of both goods and passengers. Therefore, maintaining rail safety through consistent surface defect inspection is extremely importan. However, squat defect detection on rail surfaces faces considerable difficulties due to weather impacts, lighting changes, and variations in image contrast. These challenges hinder the accuracy and reliability of traditional inspection methods. To solve this problem, this study proposes an improved YOLOv8 model for the identification and classification of squat defects. The methodology involves capturing images of the rail track, preprocessing them to enhance image quality, labeling squat defects for training purposes, and training the proposed model using the labeled dataset. The improved YOLOv8 model incorporates enhancements such as multi-scale convolution modules and attention mechanisms to improve feature extraction and defect recognition. Experimental results demonstrate the effectiveness of the proposed method, achieving an impressive accuracy of 0.92 in detecting and categorizing squat defects. These findings highlight the potential of the proposed approach to enhance railway safety by providing a reliable and efficient solution for rail surface inspection.

Corresponding Author:

Van-Dinh Do
Department of Electrical, Sao Do University
Hai Duong, 170000, Vietnam
Email: dinh.dv@saodo.edu.vn

1. INTRODUCTION

Rail transportation is a vital part of modern life, offering an efficient, safe, and environmentally sustainable means of transporting both passengers and freight. However, maintaining the smooth and secure operation of railway systems requires regular inspection and detection of surface flaws. Typical rail defects include cracks, uneven textures, wave-like deformations, and squat defects. Of these, squat defects pose a particularly severe risk, as they can generate excessive vibrations, harm train wheels, and potentially trigger catastrophic derailments. Squat defects usually arise due to material fatigue, uneven forces on the rails, or harsh weather conditions. These defects are particularly challenging to detect due to their varying sizes, shapes, and low contrast against the rail surface, making traditional inspection methods less effective. Detecting and repairing squat defects in a timely manner is essential for maintaining the safety of both passengers and freight on trains.

Modern railway track inspection employs various advanced techniques to detect surface defects, including ultrasonic testing [1]-[3], eddy current testing [4]-[6], vibration-based analysis [7]-[9], and visual inspection methods [10]-[14]. Each technique has its own strengths and limitations: ultrasonic and eddy current testing offer high precision in identifying subsurface defects but are typically costly and require sophisticated equipment; vibration analysis can detect defects like squats from a distance but demands a high level of expertise in data interpretation; while visual inspection, particularly when using computer vision systems, excels in automation and can rapidly identify visible surface defects on the tracks.

When contrasting traditional image processing-based defect detection methods with deep learning techniques, conventional approaches such as edge detection and image segmentation, while easy to implement, frequently fall short in terms of accuracy and struggle with handling complex or obscured surfaces. Deep learning models, particularly convolutional neural networks (CNNs), outperform traditional methods in defect detection accuracy by leveraging their ability to extract and classify complex patterns from large datasets. Popular CNN-based object detection architectures include Faster R-CNN, single shot multibox detector (SSD), and you only look once (YOLO). Notably, YOLO—especially YOLOv8—offers exceptional speed and precision in identifying objects. YOLOv8's optimized neural network design enhances both detection performance and computational efficiency, making it highly effective for real-time applications [15]-[17].

The specific problem addressed in this paper is the detection and classification of squat defects on rail surface, which are particularly challenging to identify due to their varying sizes, shapes, and low contrast against the rail surface. Conventional computer vision approaches, including edge detection and image segmentation algorithms, often fail to handle these complexities effectively, especially under varying environmental conditions like lighting and weather. To overcome these limitations, we propose an improved YOLOv8 model that integrates multi-scale convolution modules and attention mechanisms to enhance feature extraction and defect recognition. This approach is designed to enhance the precision and robustness of squat defect identification, even in challenging scenarios.

The contribution of this paper lies in the development of a tailored YOLOv8 model specifically optimized for railway track inspection. By incorporating advanced modules for multi-scale feature extraction and attention mechanisms, the model achieves superior performance in detecting and classifying squat defects. Experimental results demonstrate that our method outperforms traditional techniques and other deep learning models, such as EfficientDet-D0 and SSD, particularly in identifying both large and small defects. This research not only provides a robust and efficient solution for rail surface inspection but also paves the way for further applications of deep learning in railway maintenance and safety, addressing critical challenges in the industry.

The remainder of this paper is structured as follows: section 2 surveys existing literaturein on railway defect detection. Section 3 describes the system configuration and data collection process. Section 4 details the proposed methodology, including the improved YOLOv8 model. Section 5 evaluates experimental outcomes and analyzes their implications. Finally, section 6 synthesizes key findings and proposes directions for future work.

2. RELATED WORKS

The detection of railway surface defects is critical for ensuring both operational safety and system efficiency in rail transport networks. Various methods are used for defect detection, spanning from conventional methods to cutting-edge technologies. Below are some commonly used approaches:

Several studies have demonstrated the effectiveness of guided ultrasonic wave techniques for rail defect detection. As presented in [1], researchers employed this approach to successfully detect and localize defects across different rail sections (head, web, and base). The methodology's key advantage is its capability to selectively excite specific guided wave modes that exhibit high sensitivity to defects in targeted rail regions. Experimental results confirmed both the successful capture of these wave modes and their defect sensitivity, with positioning errors falling within acceptable limits for practical maintenance purposes. Complementing these findings, [2] further applied ultrasonic guided wave (UGW) technology to identify internal defects specifically within the rail head. Numerical simulations in this study demonstrated that the selected wave mode not only showed high sensitivity to internal flaws but also achieved precise defect localization. Both approaches offer the advantage of high sensitivity and precise localization of defects in the rail, facilitated by thorough simulation analyses and complex calculations. Khalil *et al.* [3], a method using ultrasonic-guided Rayleigh waves and probability of detection (POD) techniques to assess the reliability of rail track inspection systems is presented. The method employs a specially designed sensor to generate Rayleigh waves for detecting defects, such as surface cracks and sub-surface through-thickness holes. A damage index (DI) is defined based on defect size and integrated into the POD model to calculate the probability of detection, thereby evaluating the inspection system's effectiveness.

Alvarenga *et al.* [4], implemented a hybrid approach combining eddy current testing with a CNN model to enable real-time identification and categorization of rail surface defects, achieving a classification accuracy of approximately 98% with the embedded system. Park *et al.* [5], the eddy current technique (ECT) is utilized for non-destructive inspection of both surface and internal rail defects, employing a defect assessment algorithm to evaluate various parameters such as depth, length, and width. Furthermore, in [6], the ECT is employed to evaluate crack depth on rail rolling surfaces through comprehensive analysis of signal amplitude and phase characteristics; the eddy current transducer is capable of adjusting the excitation frequency and tilt angle to optimize sensitivity while minimizing interference caused by the curvature of the inspected surface.

Zuo *et al.* [7] and Tarawneh *et al.* [8], vibration analysis is used to detect rail defects by monitoring changes in vibration signals as trains pass over the tracks. Defects such as cracks, deformations, or weakened areas in the rail cause noticeable changes in vibration patterns, allowing for early detection and timely maintenance. Building on this, the method proposed in [9] decomposes nonstationary vibration signals using multi-channel singular spectrum analysis (MSSA). The results show that damage characteristics, such as sudden increases in the crest factor at a fixed angular position, can be identified through the reconstructed signal components.

Wavelet denoising techniques are used in railway inspection systems to eliminate noise artifacts from sensor-acquired inspection signals, improving analysis accuracy for vibration and eddy current signals often affected by noise [18], [19]. This technique effectively filters out irrelevant noise, preserves essential defect-related information, and enhances the detection and localization of small surface defects on rails. Building on this, [20] proposes an improved empirical wavelet transform (EWT) method to filter noise in magnetic flux leakage (MFL) signals specifically for detecting rail head defects. The method includes boundary optimization using mutual information (MI) to minimize boundary redundancy and component selection based on MI and kurtosis for more accurate signal reconstruction. Results show that this approach outperforms traditional methods like CEEMD and WT in noise reduction for MFL signals.

Beyond the previously discussed approaches, laser technology is also applied for rail surface inspection. Becker *et al.* [21], an advanced structured laser sensor is employed to gather precise data, reducing computational load when compared with traditional techniques; however, it requires specialized equipment and involves complex data processing. Similarly, Cao *et al.* [22] explores combining laser technologies with automated defect detection algorithms, such as covariance and normal intersection, which supports automation of the repair process and enhances transparency, though it is highly dependent on equipment quality and algorithm complexity. Ye *et al.* [23], a 3D pixel-level rail surface defect detection method is proposed using laser technology and deep learning. This method utilizes a low-cost 3D laser sensor to capture rail surface data, followed by a deep semantic segmentation network to detect defects at the pixel level. Experimental results show a defect detection accuracy of up to 87.9% mIoU, providing essential information about defect location, boundaries, and 3D characteristics.

Passos *et al.* [24], a CNN-based framework is employed to automatically detect the defects on the rail surface, evaluating ten widely used models to identify the most effective one for defect identification. The experimental results demonstrated an average detection accuracy of 83.7% for both light surface defects and squat cracks. Notably, the Inceptionv3 architecture outperformed other models, achieving 92% classification accuracy specifically for severe squat defect cases. Additionally, an innovative UAV-based inspection system for rail defect detectionis introduced in [25], focusing on image enhancement through the local weber-like contrast (LWLC) algorithm and defect segmentation using a threshold technique called gray stretch maximum entropy (GSME). Studies [12], [25], [26] further explore defect detection in railway systems using deep learning and computer vision. Zhang *et al.* [25] employs spiking neural networks for recognizing defects in train wheelsets, while [26] utilizes YOLOv5 and Mask R-CNN for detecting defects on the rail surfaces and fasteners. Finally, Belkhade [12] focuses on analyzing railway superstructures through image processing, enhancing safety and efficiency in railway transportation.

In addition, the technique of railway inspection from above using UAVs combined with deep learning offers several advantages, including the ability to access difficult areas and enhance defect detection efficiency through automated and accurate image processing [27]-[29]. The use of UAVs helps to reduce inspection time and costs compared to traditional methods. However, the limitations of this technique include reliance on the quality of the captured images, as well as the need for proper calibration and training of the deep learning model to ensure high accuracy. Additionally, adverse weather conditions can affect the UAV's operational capabilities, reducing inspection effectiveness.

Currently, numerous studies explore methods for detecting and inspecting defects on railway tracks and surfaces through image processing and deep learning [12]-[16], [30], [31]. The works in [12]-[14], [30] analyze and implement machine learning algorithms for railway maintenance, aiming to diagnose and predict defects while offering a classification of existing machine learning approaches to help the railway industry respond in a timely manner. With their outstanding advantages, YOLO models have been extensively improved by researchers worldwide. These models are not only applied in detecting and tracking human activities or pedestrians, as demonstrated in [32], [33], but are also widely utilized in industrial applications such as defect inspection on solar cell surfaces and solar panels, as highlighted in [34], [35]. Recently, research on enhanced YOLO models has demonstrated high efficiency in defect detection on railway tracks.

Wen *et al.* [17], the YOLOv5-CGBD model is employed to detect damage to fasteners on railway tracks. Nakhaee *et al.* [13] enhanced YOLOv5s to detect surface defects on railway tracks. Talib *et al.* [15], the YOLOv8-CAB model was introduced, incorporating the context attention block (CAB) and the coarse-to-fine (C2F) block to improve the detection of small objects, achieving a mAP of 97% on the COCO dataset, underlining its real-time detection precision. Finally, Tan *et al.* [16], YOLOv8 was applied to detect bolts and nuts for robotic arm vision systems, achieving a mAP of 90% in scenarios resembling the training data, with particular effectiveness at distances between 20 cm and 60 cm. While these studies have achieved notable results, none have specifically focused on detecting and classifying squat defects. Following a similar approach, the papers [17], [35]-[38] have in common the research and development of methods for detecting defects on the rail surface based on the YOLOv8 model or its improved versions. These studies focus on optimizing the detection and classification of different types of defects, such as squats, cracks and other defects on train wheels and rails. The above methods have significantly improved the detection speed and accuracy for rail surface inspection.

It is evident that the primary disadvantages of the abovementioned methods include high costs, the need for specialized equipment, and advanced data analysis experience. Furthermore, certain techniques like eddy current and ultrasonic testing require direct contact with the rails, making it difficult to perform continuous and rapid inspections over extended distances. And vibration analysis also requires large volumes of data and significant processing time. Compared to these methods, the benefit of using YOLOv8 for non-contact visual inspection of rail defects lies in its capability to quickly and automatically detect surface flaws, due to computer vision models. YOLOv8 can directly recognize defects from images, eliminating the need for sophisticated equipment, and avoiding direct contact with the rails. This method not only cuts expenses but is also simple to deploy in real-time, delivering high precision in identifying visible flaws. This substantially boosts the effectiveness and applicability of periodic inspection operations, securing the railway network's safety.

3. SYSTEM CONFIGURATION

For image collection in the railway visual inspection system, a photographic setup is configured, as shown in Figure 1. The arrangement consists of a camera oriented to collect images of the track along its length, complemented by two illumination units installed symmetrically on both sides of the camera's viewpoint. To obtain sharp and high-quality track images, an industrial-grade camera with 1024×1024 pixel resolution is employed. This imaging unit is positioned 50 cm above the rail surface.

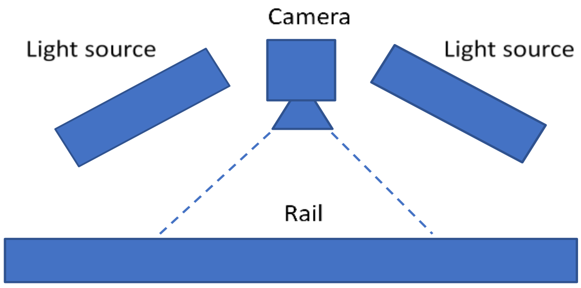
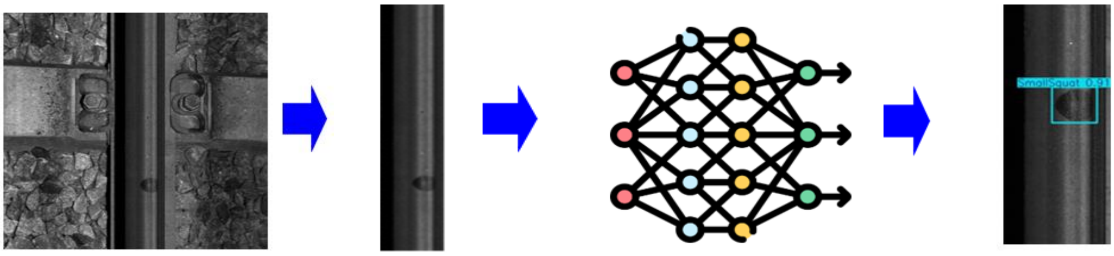


Figure 1. Structure of the imaging system

This system is mounted on the track inspection car operated by Hunan Intercity Railway Company, serving the railway line between Changsha and Zhuzhou in China's Hunan Province. During operation, the equipment captures images of every segment of the track as the vehicle progresses, with each image encompassing a 1-meter section. The captured images are sequentially numbered to enable accurate mapping of each track segment's position along the inspection path.

For accurate detection of squat defects on rail surfaces, we propose a defect identification method employing an enhanced YOLOv8 model as shown in Figure 2. The process involves: (i) processing camera-captured images to reduce noise as shown in Figure 2(a), (ii) extracting rail sections as 200×1024 pixel images to optimize defect detection in Figure 2(b), and (iii) training the improved YOLOv8 model in Figure 2(c). Once trained, the model is deployed to identify squat defects on rail surfaces in Figure 2(d).



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

Figure 2. Pipeline of our method: (a) original image, (b) rail image, (c) improved YOLOv8 model, and (d) predicted image

4. METHODOLOGIES AND DATA

Figure 3 presents sample images extracted from the original set captured by the imaging system installed on the inspection vehicle. These images reveal that the photos taken by the system are affected by several environmental and material factors, including atmospheric conditions, illumination intensity, the reflective characteristics of the rail, and contrast variations between the crushed stone ballast bed and cement roadbed surfaces. Additionally, the squat defects produce different grayscale levels between different images. These challenges significantly hinder the accurate detection of squat defects, particularly for traditional image processing-based methods.

Our dataset includes 1000 grayscale images with a size of 1024×1024 pixels. For convenience in training and testing the model, the area containing the rail is extracted from original images as rail images with a resolution of 200×1024 pixels using the rail image extraction technique introduced in our previous study [39].

The images were subsequently annotated, as illustrated in Figure 4. A key observation is that rail surface squat defects exhibit variations in both dimensions and grayscale intensity. Here, we categorize the defects into two groups according to their size: big squats in Figure 4(a) and small squats in Figure 4(b). The LabelImg tool was used to annotate each image.

After all original images were annotated, the dataset was constructed by pairing the images with their corresponding labels. This dataset was then divided into three subsets: 80% of the rail images and labels were assigned to the training set, 10% to the validation set, and the remainder to the test set. This allocation strategy ensures the model receives sufficient training data while retaining adequate samples for evaluating its performance on both validation and test sets.

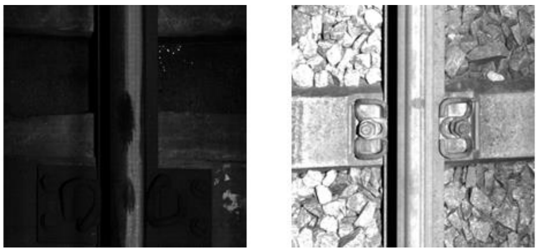
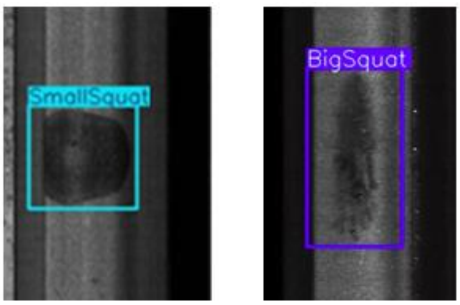


Figure 3. Illustration of some original images



(a) (b)

Figure 4. Annotating technique with two types of defect (a) small squat and (b) big squat

In this study, we employ an improved YOLOv8 model to address the object detection challenge. Compared with YOLOv8 network, the proposed model has the following improvements as shown in Figure 5, while the details of these improvements are illustrated in Figure 6. We replaced the traditional convolutions in C2f with a more efficient and lightweight multi-scale convolution module called SEF as shown in Figure 6(a). This module incorporates convolutions of different kernel sizes to capture spatial features at various scales, while also expanding the receptive field using larger kernels to enhance the model's ability to recognize long-range dependencies. In the SEConv2d operation as shown in Figure 6(b), input channels are divided into four parts; two of them are left unchanged, and the others undergo 3×3 and 5×5 convolutions. A final 1×1 convolution merges these channels, reducing computational load and the number of parameters while preserving key features. This approach allows SEF to maintain local details and essential semantic information as the network deepens.

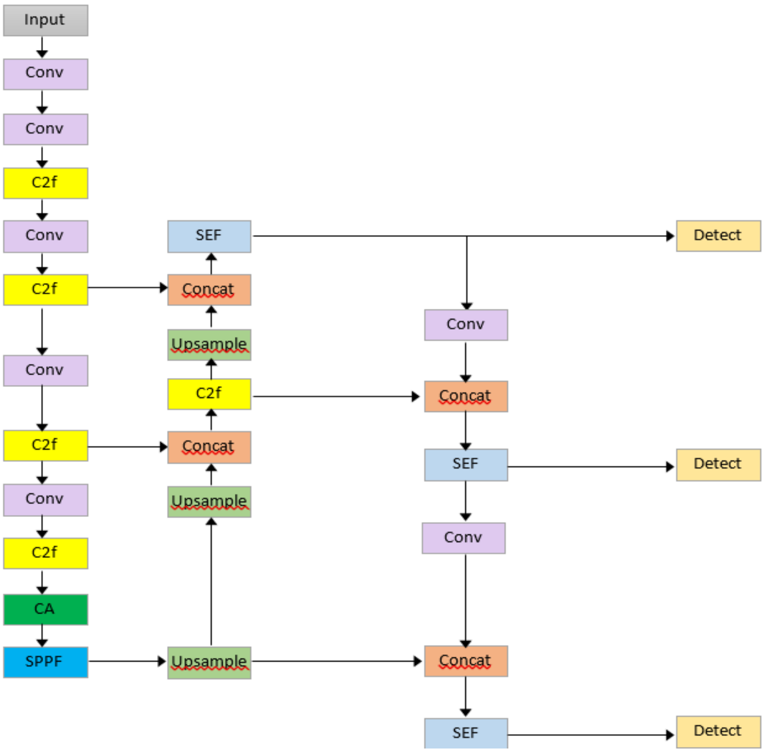


Figure 5. The structure of our model

To improve feature extraction for squat defect, our model integrates the coordinate attention (CA) module strategically positioned before the SPPF layer in the backbone network, as depicted in Figure 5. This placement was carefully chosen to maximize spatial feature integration before the final downsampling stage, ensuring both global context preservation and local defect detail retention. The significance of this attention mechanism lies in its ability to boost the accuracy of the neural network while incurring minimal computational costs. As shown in Figure 6(c), the CA module separates the attention process into two parallel one-dimensional encoding pathways along the x and y axes. This configuration effectively consolidates spatial coordinate information into the attention map, leading to significant improvements in model performance. By embedding the CA module into the architecture, our method further enhances the feature extraction capabilities, resulting in more precise detection and recognition of squat defects. The module's lightweight design makes it particularly suitable for deployment in resource-constrained railway inspection systems, achieving real-time performance.

We selected a total of 200 epochs to guarantee sufficient training time for the model to learn the data's features, while avoiding excessive epochs that could lead to overfitting. The batch size was configured to 16, which enhances the efficiency of the training process in terms of computational resources. To enable stable learning without significant fluctuations between iterations, we set the learning rate to 0.001. Finally, we employed the Adam optimizer, which effectively integrates the advantages of momentum optimization and RMSProp, enabling faster convergence while maintaining stable model parameter updates. Figure 7 shows the training results of the proposed model.

After training, the proposed model was deployed to automate squat defect detection and classification through the following steps: (i) camera-captured images were preprocessed to reduce noise, (ii) rail surfaces were extracted from the original images using the rail image extraction technique described in [39], and (iii) the processed rail images were predicted by our improved YOLOv8 model. The system classified detected defects into two predefined categories (big and small squats), enabling operators to receive alerts and implement appropriate corrective measures.

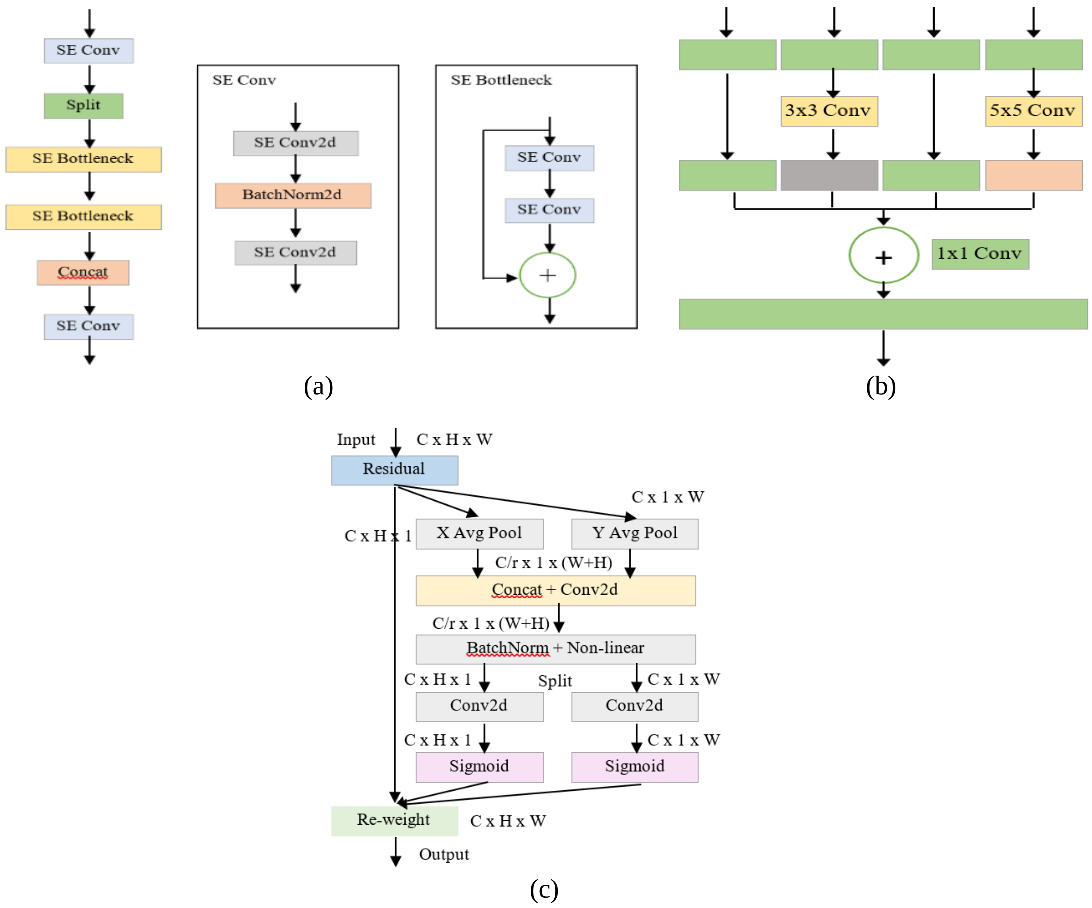


Figure 6. The structure of (a) SFE, (b) SEConv2d, and (c) CA module

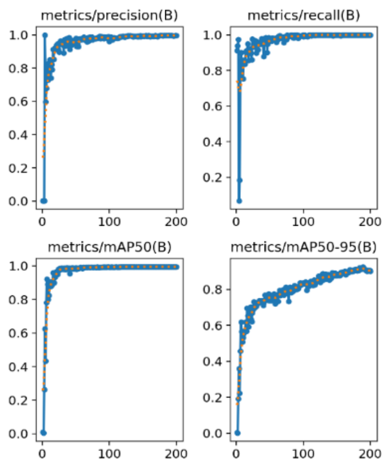
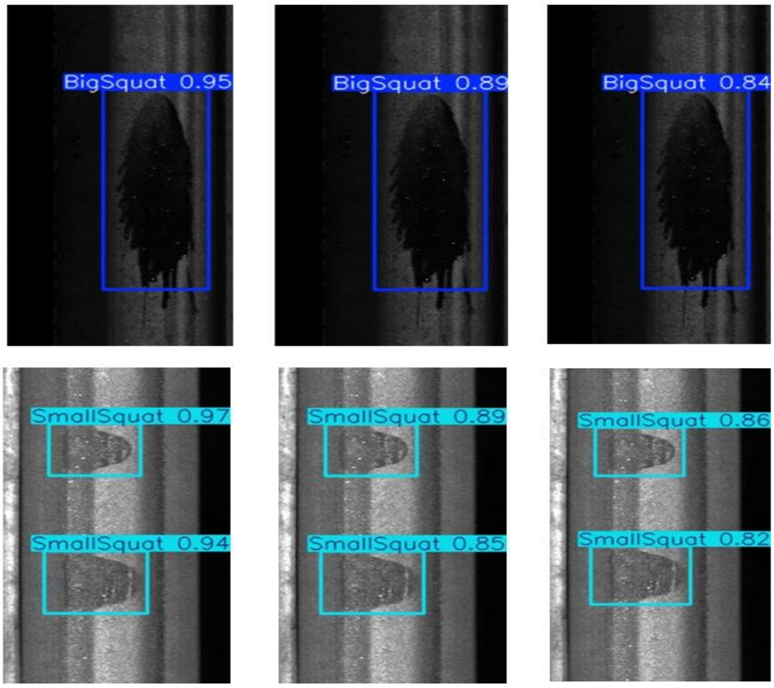


Figure 7. The proposed model training results with our dataset

5. EXPERIMENTAL RESULTS

The results of the squat defect detection with different methods are shown in Figure 8. It can be seen that all of the defects in each rail image were identified. At the same time, they are categorized into two groups based on their size, including Big squat and Small squat.



(a) (b) (c)

Figure 8. Experimental results of squat detection using different methods: (a) the proposed method, (b) EfficientDet-D0, and (c) SSD

To evaluate the proposed method's performance in detecting squat defects, we employed precision (Pre) and recall (Re) as evaluation metrics. They can be computed by the (1) and (2).

$$Pre = \frac{PT}{TP+FP} \tag{1}$$

$$Re = \frac{TP}{TP+PN} \tag{2}$$

Where true positive (TP) indicates the defects that have been accurately identified; false positive (FP) refers to defects that were mistakenly detected; and false negative (FN) signifies the defects that were missed.

From the results in Table 1, we can observe that all defects belonging to the big squat group were detected, although some of them were incorrectly categorized as small squat. Regarding the small squat defects, some of them were undetected. This outcome can be attributed to the diminutive shapes and sizes of these defects, as well as their minimal contrast against the surrounding rail image. These findings underscore the model's strength in detecting larger, more visible defects while highlighting a key limitation in addressing small or low-contrast squats—a challenge consistent with prior studies using visual inspection methods [17], [20]. The misclassification of some large defects as small may stem from overlapping features in the training data or the model's sensitivity to local texture variations. Notably, the model's performance aligns with the study's goal of automating rail inspections, achieving a balance between precision (0.92) and recall (0.96) that surpasses traditional techniques like ultrasonic testing [3] in speed and scalability.

In this study, due to a lack of technical infrastructure, we cannot deploy some previously mentioned rail track assessment techniques, including ultrasonic waves, laser, or eddy current. To demonstrate the efficacy of our approach, some experiments were conducted with two additional deep learning models, i.e, EfficientDet-D0 and SSD. The choice of these comparative models was strategic: EfficientDet-D0 represents state-of-the-art efficiency in object detection [30], while SSD serves as a widely adopted baseline for real-time applications [26]. This comparative framework directly addresses our hypothesis that the improved YOLOv8 architecture would outperform conventional approaches in both accuracy and computational efficiency for rail defect detection. The experimental outcomes illustrated in Figure 8 reveal that our method attains a higher confidence value for each type of squat defect compared to the alternative methods. This confidence metric serves as a filter to eliminate unreliable predictions exhibiting values markedly less than 1. Notably, our model achieved an average confidence score of 0.94 for big squats and 0.89 for small squats, representing 12% and 15% improvements over SSD and EfficientDet-D0 respectively. Furthermore, the comparative results of the proposed method alongside the two other approaches are presented in Table 2. These experimental findings show that our method is more dependable, exhibiting greater accuracy than those from the other techniques.

Table 1. Performance of our method in squat detection

Defect type	TP	FP	FN	Pre	Re
Big squat	32	3	0	0.92	1.00
Small squat	91	8	6	0.92	0.94

Table 2. Comparison results of the proposed method with two other ones

Algorithm	Pre	Re	mAP@0.5
SSD	0.88	0.89	0.88
EfficientDet-D0	0.89	0.93	0.90
Ours	0.92	0,96	0.93

From the experimental results, it can be seen that our method achieved high performance in detecting and identifying squat defects on railway tracks. Specifically, the model accurately classified the squat defects into two groups based on their size, with precision and recall values of 0.92 and 0.96, respectively. A key piece of evidence supporting these results is that the model detected all large defects (Recall=1.00), confirming its robust recognition capability for objects with large sizes and clear contrast.

When compared to other deep learning models, i.e, EfficientDet-D0 and SSD, our method demonstrated superior performance in both precision and recall. Specifically, our model can achieve an mAP@0.5 value of 0.93, higher than the 0.90 of EfficientDet-D0 and 0.88 of SSD. This study's innovation primarily derives from the combination of improvements such as the SEF module and CA, which enhance feature extraction and defect recognition capabilities. However, the study also has some limitations, such as the omission of some small squat defects due to their size and low contrast.

This research aims to develop an effective framework for the detection and classification of squat defects in railway track, thereby supporting maintenance efforts and ensuring railway safety. The experimental results not only demonstrate the feasibility of the proposed method but also open new avenues for applying deep learning to practical problems. However, some questions remain unresolved, such as the technique for improving the detection of small defects and enhancing the model's reliability under varying environmental conditions. In the future, researchers can expand the dataset, integrate data augmentation techniques, and develop specialized attention modules to further improve the model's performance.

6. CONCLUSION

Railway transportation serves as a fundamental pillar of worldwide economic systems, guaranteeing the secure and reliable transit of both cargo and travelers. However, the presence of squat defects on rail surfaces poses significant safety risks, making their timely detection and classification essential for maintaining operational safety. This paper addresses this challenge by proposing an improved YOLOv8 model specifically optimized to identify and categorize squat defects. The model incorporates multi-scale convolution modules and attention mechanisms to enhance feature extraction and defect recognition, achieving high accuracy even under varying environmental conditions such as lighting and weather. Extensive experimental validation confirms the superior performance of our approach compared to conventional methods and state-of-the-art deep learning architectures, such as EfficientDet-D0 and SSD, particularly in identifying both large and small defects. Therefore, this study has provided an effective solution for detecting defects with complex characteristics on rail surfaces, while also highlighting the potential of the improved YOLO model in railway inspection and maintenance. Future research directions will prioritize enhancing detection sensitivity for low-contrast small defects and improving model generalizability across varied operational environments.

ACKNOWLEDGEMENTS

Authors wish to express sincere thanks to Saodo University, Vietnam for providing constant support and encouragement during paper preparation.

FUNDING INFORMATION

Authors state no funding involved.

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
Van-Dinh Do	✓	✓	✓	✓		✓		✓	✓	✓		✓	✓	
Phuong-Ty Nguyen		✓		✓	✓	✓	✓	✓	✓		✓	✓		
Minh-Tuan Ha	✓		✓	✓	✓	✓	✓			✓	✓		✓	

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

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

Ethical approval is not applicable as this paper as not talk about using people or animals.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] H. Shi, L. Zhuang, X. Xu, Z. Yu, and L. Zhu, “An ultrasonic guided wave mode selection and excitation method in rail defect detection,” *Applied Sciences (Switzerland)*, vol. 9, no. 6, p. 1170, Mar. 2019, doi: 10.3390/app9061170.

[2] J. Xie, W. Ding, W. Zou, T. Wang, and J. Yang, “Defect detection inside a rail head by ultrasonic guided waves,” *Symmetry*, vol. 14, no. 12, p. 2566, Dec. 2022, doi: 10.3390/sym14122566.

[3] A. Khalil, F. Masurkar, and A. Abdul-Ameer, “Estimating the reliability of the inspection system employed for detecting defects in rail track using ultrasonic guided waves,” in *Lecture Notes in Civil Engineering*, vol. 473 LNCE, Springer Nature Switzerland, 2024, pp. 190–202.

[4] T. A. Alvarenga, A. L. Carvalho, L. M. Honorio, A. S. Cerqueira, L. M. A. Filho, and R. A. Nobrega, “Detection and classification system for rail surface defects based on eddy current,” *Sensors*, vol. 21, no. 23, p. 7937, Nov. 2021, doi: 10.3390/s21237937.

[5] J. W. Park *et al.*, “Rail surface defect detection and analysis using multi-channel eddy current method based algorithm for defect evaluation,” *Journal of Nondestructive Evaluation*, vol. 40, no. 3, Sep. 2021, doi: 10.1007/s10921-021-00810-9.

[6] S. P. Shlyakhtenkov, D. B. Nekrasov, S. V. Palagin, O. V. Bessonova, A. A. Popkov, and S. A. Bekher, “Possibilities of manual eddy current testing for depth gaging of contact-fatigue cracks on rail rolling surface,” *Russian Journal of Nondestructive Testing*, vol. 59, no. 4, pp. 447–455, Apr. 2023, doi: 10.1134/S1061830923700328.

[7] Y. Zuo, J. Lundberg, T. Najeh, M. Rantatalo, and J. Odelius, “Squat detection of railway switches and crossings using point machine vibration measurements,” *Sensors*, vol. 23, no. 7, p. 3666, Mar. 2023, doi: 10.3390/s23073666.

[8] C. Tarawneh, J. Montalvo, and B. Wilson, “Defect detection in freight railcar tapered-roller bearings using vibration techniques,” *Railway Engineering Science*, vol. 29, no. 1, pp. 42–58, Feb. 2021, doi: 10.1007/s40534-020-00230-x.

[9] M. Salehi, S. A. Bagherzadeh, and M. Fakhari, “Experimental detection of train wheel defects using wayside vibration signal processing,” *Structural Health Monitoring*, vol. 22, no. 5, pp. 3286–3301, Jan. 2023, doi: 10.1177/14759217221149614.

[10] Y. Wu, Y. Qin, Z. Wang, and L. Jia, “A UAV-based visual inspection method for rail surface defects,” *Applied Sciences (Switzerland)*, vol. 8, no. 7, p. 1028, Jun. 2018, doi: 10.3390/app8071028.

[11] P. Aela, J. Cai, G. Jing, and H. L. Chi, "Vision-based monitoring of railway superstructure: A review," *Construction and Building Materials*, vol. 442, p. 137385, Sep. 2024, doi: 10.1016/j.conbuildmat.2024.137385.
[12] A. Belkhade, "Automatic vision based inspection of railway track: a review," *International Journal of Research in Engineering and Technology*, vol. 03, no. 02, pp. 12–15, Feb. 2014, doi: 10.15623/ijret.2014.0302003.
[13] M. Chenariyan Nakhaee, D. Hiemstra, M. Stoelinga, and M. van Noort, "The recent applications of machine learning in rail track maintenance: a survey," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11495 LNCS, Springer International Publishing, 2019, pp. 91–105.
[14] P. M. Bhatt *et al.*, "Image-based surface defect detection using deep learning: a review," *Journal of Computing and Information Science in Engineering*, vol. 21, no. 4, Feb. 2021, doi: 10.1115/1.4049535.
[15] M. Talib, A. H. Y. Al-Noori, and J. Suad, "YOLOv8-CAB: Improved YOLOv8 for real-time object detection," *Karbala International Journal of Modern Science*, vol. 10, no. 1, pp. 56–68, Jan. 2024, doi: 10.33640/2405-609X.3339.
[16] Y. K. Tan, K. M. Chin, T. S. H. Ting, Y. H. Goh, and T. H. Chiew, "Research on YOLOv8 application in bolt and nut detection for robotic arm vision," in *KST 2024 - 16th International Conference on Knowledge and Smart Technology*, Feb. 2024, pp. 126–131, doi: 10.1109/KST61284.2024.10499651.
[17] Y. Wen, X. Gao, L. Luo, and J. Li, "Improved YOLOv8-based target precision detection algorithm for train wheel tread defects," *Sensors*, vol. 24, no. 11, p. 3477, May 2024, doi: 10.3390/s24113477.
[18] Y. Zuo, F. Thiery, P. Chandran, J. Odelius, and M. Rantatalo, "Squat detection of railway switches and crossings using wavelets and isolation forest," *Sensors*, vol. 22, no. 17, p. 6357, Aug. 2022, doi: 10.3390/s22176357.
[19] S. Taheri, B. Moslehi, V. Sotoudeh, and B. M. Hopkins, "Rail defect detection using fiber optic sensors and wavelet algorithms," Apr. 2018, doi: 10.1115/JRC2018-6109.
[20] Y. Jia, J. Lin, P. Wang, and Y. Zhu, "An improved empirical wavelet transform filtering method for rail-head surface-defect magnetic-flux leakage signal," *Applied Sciences (Switzerland)*, vol. 14, no. 2, p. 526, Jan. 2024, doi: 10.3390/app14020526.
[21] D. De Becker, J. Dobrzanski, L. Justham, and Y. M. Goh, "A laser scanner based approach for identifying rail surface squat defects," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 235, no. 6, pp. 763–773, Oct. 2021, doi: 10.1177/0954409720962252.
[22] X. Cao, W. Xie, S. M. Ahmed, and C. R. Li, "Defect detection method for rail surface based on line-structured light," *Measurement: Journal of the International Measurement Confederation*, vol. 159, p. 107771, Jul. 2020, doi: 10.1016/j.measurement.2020.107771.
[23] J. Ye, E. Stewart, Q. Chen, C. Roberts, A. M. Hajiyavand, and Y. Lei, "Deep learning and laser-based 3-D pixel-level rail surface defect detection method," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023, doi: 10.1109/TIM.2023.3272033.
[24] R. A. da S. L. Passos, M. P. Ferreira, B.-H. de A. e Silva, L. A. S. Lopes, H. Ribeiro, and R. P. dos Santos, "An in-depth assessment of convolutional neural networks for rail surface defect detection," *Research, Society and Development*, vol. 11, no. 8, p. e12211830252, Jun. 2022, doi: 10.33448/rsd-v11i8.30252.
[25] C. Zhang, C. Huang, and J. He, "Defects recognition of train wheelset tread based on improved spiking neural network," *Chinese Journal of Electronics*, vol. 32, no. 5, pp. 941–954, Sep. 2023, doi: 10.23919/cje.2022.00.162.
[26] D. Zheng *et al.*, "A defect detection method for rail surface and fasteners based on deep convolutional neural network," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/2565500.
[27] I. Aydin, M. Sevi, K. Sahbaz, and M. Karakose, "Detection of rail defects with deep learning controlled autonomous UAV," in *2021 International Conference on Data Analytics for Business and Industry, ICDABI 2021*, Oct. 2021, pp. 500–504, doi: 10.1109/ICDABI53623.2021.9655796.
[28] Q. Yu, A. Liu, X. Yang, and W. Diao, "An improved lightweight deep learning model and implementation for track fastener defect detection with unmanned aerial vehicles," *Electronics (Switzerland)*, vol. 13, no. 9, p. 1781, May 2024, doi: 10.3390/electronics13091781.
[29] M. Eylence, M. Yücel, M. M. Özmen, and B. Aksoy, "Railway security system design by image processing and deep learning unmanned aerial vehicle," *Türk Doğa ve Fen Dergisi*, vol. 11, no. 3, pp. 150–154, Sep. 2022, doi: 10.46810/tdfd.1112957.
[30] L. Wang, Q. Zang, K. Zhang, and L. Wu, "A rail fastener defect detection algorithm based on improved YOLOv5," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 238, no. 7, pp. 851–862, Feb. 2024, doi: 10.1177/09544097241234380.
[31] H. Luo, L. Cai, and C. Li, "Rail surface defect detection based on an improved YOLOv5s," *Applied Sciences (Switzerland)*, vol. 13, no. 12, p. 7330, Jun. 2023, doi: 10.3390/app13127330.
[32] S. Rajakumar and R. B. Azad, "A novel YOLOv8 architecture for human activity recognition of occluded pedestrians," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 5, pp. 5244–5252, Oct. 2024, doi: 10.11591/ijece.v14i5.pp5244-5252.
[33] D. Jayaram, S. Vedagiri, and M. Ramachandra, "Framework for multiple person identification using YOLOv8 detector: a transfer learning approach," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 2790–2802, Jun. 2024, doi: 10.11591/ijece.v14i3.pp2790-2802.
[34] M. Zhang, and L. Yin, "Solar cell surface defect detection based on improved YOLO v5," *IEEE access*, vol. 10, p. 80804-80815, 2022, doi: 10.1109/ACCESS.2022.3195901.
[35] A. Ghahremani, S.D. Adams, M. Norton, S.Y. Khoo, and A.Z. Kouzani, "Detecting defects in solar panels using the YOLO v10 and v11 algorithms," *Electronics*, vol. 14, no. 2, p. 344, 2025, doi: 10.3390/electronics14020344.
[36] J. Du, R. Zhang, R. Gao, L. Nan, and Y. Bao, "RSDNet: a new multiscale rail surface defect detection model," *Sensors*, vol. 24, no. 11, p. 3579, Jun. 2024, doi: 10.3390/s24113579.
[37] S. Yang, Z. Zhang, B. Wang, and J. Wu, "DCS-YOLOv8: an improved steel surface defect detection algorithm based on YOLOv8," in *ACM International Conference Proceeding Series*, Jan. 2024, pp. 39–46, doi: 10.1145/3647649.3647656.
[38] Y. Ding, Q. Zhao, T. Li, C. Lu, L. Tao, and J. Ma, "A rail defect detection framework under class-imbalanced conditions based on improved you only look once network," *Engineering Applications of Artificial Intelligence*, vol. 138, p. 109351, Dec. 2024, doi: 10.1016/j.engappai.2024.109351.
[39] D. Pham, M. Ha, and C. Xiao, "A novel visual inspection system for rail surface spalling detection," *IOP Conference Series: Materials Science and Engineering*, vol. 1048, no. 1, p. 12015, Jan. 2021, doi: 10.1088/1757-899X/1048/1/012015.

BIOGRAPHIES OF AUTHORS

Van-Dinh Do 1998: Graduated from University with major in automation, Hanoi University of Science and Technology; 2005: Graduated with Master's degree in automation engineering, Hanoi University of Science and Technology; 2018: Graduated with a doctorate in control and automation engineering, Hanoi University of Science and Technology; current job: lecturer, administrator, Sao Do University; the main researches are the application of artificial intelligence in measurement, control and automation solutions, and intelligent measuring devices. He can be contacted at email: dinh.dv@saodo.edu.vn.

Phuong-Ty Nguyen 2001: Graduated from University with major in electrical and electronic equipment, Hanoi University of Science and Technology; 2005: Graduated with Master's degree in automation engineering, Hanoi University of Science and Technology; 2020: Graduated with a doctorate in electrical and thermal engineering, Platov South-Russian State Polytechnic University, NPI; current job: lecturer, Sao Do University; the main researches are electrical equipment and industrial robot control, nonlinear control systems, fuzzy control, neural control. He can be contacted at email: tynp2109@gmail.com.

Minh-Tuan Ha graduated with a Bachelor's degree in transport electrical engineering from Maritime University in 2005, a Master's degree in instrumentation and control from Hanoi University of Science and Technology in 2010, and a Doctor of engineering in control science and engineering from Hunan University in China in 2022. Currently, he is a lecturer at Sao Do University, focusing his research on deep learning, image processing techniques, 3D reconstruction, and AI systems. He can be contacted at email: minhtuanha031@gmail.com.