Croatian Society
for Materials
Protection

Metalurgija/Metallurgy

Vol. 65 No. 4 / 2026

Research on Surface Defect Detection and Intelligent Identification Method of Hot-Rolled Strip Based on Deep Learning

S.C. Xie [1], Y.X. An [1], H. Wang [1], J.T. Yang [1], Y.H. Lin [1], and G.Z. Ren [1]*

https://doi.org/10.64486/m.65.4.15

[1] School of Mechanical Engineering and Automation, University of Science and Technology Liaoning, China
* Correspondence: a1041@sina.com

Type of the Paper: Article
Received: December 18, 2025
Accepted: May 7, 2026

Abstract: Hot-rolled steel strips play a crucial role in industrial settings, where the accurate identification of surface defects is essential to uphold product quality and safety. This study introduces an enhanced version of the YOLOv8 model by integrating an Efficient Multi-scale Attention (EMA) mechanism into the C2f module, thereby creating the C2f_EMA module. This integration aims to improve the adaptive feature representation in both channel and spatial dimensions. The EMA mechanism serves to emphasize critical defect areas, suppress irrelevant background details, and enhance the detection precision of intricate and small defects. Evaluation on the NEU-DET dataset reveals that the upgraded model exhibits superior detection accuracy across most of the six defect categories, resulting in an overall mean average precision boost from 76.1 % to 77.6 %. Particularly noteworthy is the substantial enhancement in detecting small and medium-scale defects like Inclusion, Scratches, and Crazing. These findings underscore the efficacy of the C2f_EMA module in augmenting multi-scale feature representation within the YOLOv8 framework, all while preserving its lightweight nature and real-time performance. Consequently, this approach proves to be well-suited for surface defect identification in hot-rolled steel strip production lines.

Keywords: steel strip; hot-rolled; surface defect; object detection; YOLOv8; attention mechanism

1. Introduction

Hot-rolled steel strips play a crucial role as fundamental materials in contemporary industrial production, extensively applied in industries like automotive, shipbuilding, construction, and equipment manufacturing. Nevertheless, the rapid production pace and elevated temperatures during manufacturing make the steel strip surfaces susceptible to diverse flaws such as scratches, inclusions, and pits. Failure to promptly identify and rectify these flaws can markedly undermine product quality and potentially result in serious safety risks during subsequent processing and application. Hence, the advancement of effective and precise defect identification techniques holds paramount importance for quality assurance and the smart advancement of the steel production sector.

Conventional manual inspection methods, being inefficient, costly, and heavily reliant on inspectors' expertise, are inadequate for large-scale industrial production requirements [1]. Recently, deep learning-based object detection technologies, notably the YOLO series models, have made significant progress [2]. These models have emerged as a preferred option for industrial defect detection due to their optimal balance between detection accuracy and real-time performance [3]. The latest version in this series, YOLOv8, brings enhancements to its backbone network, feature fusion structure, and detection head design, showcasing notable versatility and accuracy benefits [4]. Nevertheless, the C2f fusion module in YOLOv8 still exhibits limitations in feature interaction, particularly in its inadequate capacity to emphasize crucial information in multi-scale features [5]. This deficiency leads to restricted detection performance for small-scale defects within intricate backgrounds [6].

This paper suggests an enhanced YOLOv8 defect detection model by incorporating an Efficient Multi-scale Attention (EMA) mechanism into the C2f module. The EMA mechanism dynamically captures feature map correlations in channel and spatial dimensions to highlight crucial defect areas and reduce irrelevant background details, thus improving the efficiency and robustness of the fused feature representations. Through the integration of EMA with the C2f module, the model improves the accuracy of detecting multi-scale defects while preserving its lightweight architecture and real-time capabilities.

Understanding the metallurgical mechanisms underlying various surface defects is crucial for developing effective detection algorithms. Hot-rolled steel strips are susceptible to different types of surface defects during the manufacturing process, each with distinct formation mechanisms [7]. Inclusion defects typically originate from non-metallic inclusions embedded in the steel matrix during steelmaking, appearing as dark patches that disrupt surface continuity [8]. Scratches and crazing often result from mechanical interactions between the steel surface and rolling equipment, manifesting as linear patterns parallel or perpendicular to the rolling direction [9]. Rolled-in scale defects occur when oxide scales formed during heating fail to be completely removed and become pressed into the steel surface during rolling [10]. Patches and pitted surface defects are frequently associated with uneven cooling, thermal stress concentration, or surface reactions with the cooling environment [11]. These defects not only affect the aesthetic appearance but also compromise mechanical properties such as fatigue strength and corrosion resistance. Therefore, accurate detection and classification of these defects using deep learning approaches can provide valuable insights for process optimization and quality control in steel production.

2. Related Works

Historically, the early detection of surface defects in hot-rolled steel strips predominantly depended on conventional physical detection techniques like acoustic detection, ultrasonic testing, and contact-based methods [12]. These approaches ascertain possible defects by examining the acoustic characteristics or surface reactions of materials during transmission, providing a degree of detection proficiency. Nonetheless, constrained by accuracy, adaptability, and efficiency, these techniques encounter challenges in fulfilling the instantaneous detection requirements of extensive production processes amid the expanding capacities of steel manufacturing.

The rise of deep learning has significantly enhanced defect detection, particularly with the widespread adoption of the YOLO series models for their real-time performance and accuracy benefits. YOLOv1 pioneered the single-stage detection framework [13]. YOLOv3 integrated Feature Pyramid Networks (FPN) to improve small object detection [14], and YOLOv5 further refined performance by implementing the CSP structure and adaptive anchor box mechanisms [15]. Expanding on these advancements, YOLOv8 introduces an anchor-free detection head, a decoupled detection head, and incorporates the C2f module to enhance feature extraction and optimize gradient propagation efficiency [16].

The C2f module in YOLOv8 currently lacks adaptive modeling capabilities during feature fusion, which leads to inadequate representation of small-scale defects. To mitigate this limitation, this paper proposes the integration of the Efficient Multi-scale Attention (EMA) mechanism into the C2f module. This enhancement aims to improve the network's feature selection and representation capabilities, thereby boosting detection performance for multi-scale defects, especially small-scale defects [17]. The NEU-DET dataset used in this study provides a comprehensive benchmark for evaluating steel surface defect detection algorithms [18].

https://doi.org/10.64486/m.65.4.15

Additionally, Kou et al. developed a YOLO-V3-based model for detecting defects on steel strip surfaces, demonstrating the effectiveness of YOLO series models in industrial defect detection [19].

3. Methodology

To enhance global dependency modeling in feature representation while preserving the lightweight advantages of the C2f structure, this paper introduces an Efficient Multi-branch Attention (EMA) mechanism within the original C2f module framework, constructing an improved structure called C2f-EMA. This module achieves collaborative recalibration across spatial and channel dimensions through grouped modeling and dual-path response interaction, while maintaining manageable computational complexity.

In the standard C2f structure, input features are first channel-expanded via 1×1 convolution and split into two parts along the channel dimension. One part directly participates in cross-stage concatenation, while the other sequentially passes through several Bottleneck units to enhance local representation capabilities. This design inherits the CSP (Cross Stage Partial) concept, effectively reducing redundant gradient propagation [20]. However, traditional Bottleneck structures primarily rely on convolution for local modeling, and their ability to characterize long-range spatial dependencies and direction-sensitive structures remains limited [21]. To address this, this paper embeds the EMA module within the Bottleneck to improve the global consistency and direction-selective response capabilities of features.

The core idea of EMA is to construct dual-branch interactive attention within a grouped representation space. Let the input feature be $X \in \mathbb{R}^{B \times C \times H \times W}$. First, it is divided into G groups along the channel dimension, with each group having C/G channels. The grouping strategy dynamically ensures $C/G \geq 1$, maintaining stable modeling granularity across different scales. Subsequently, feature recalibration is performed independently within each group.

In spatial modeling, EMA extracts statistical information in the height and width directions through direction-aware pooling operations:

$$X_h = \mathrm{Pool}_h(X), X_w = \mathrm{Pool}_w(X) \tag{1}$$

where $\mathrm{Pool}_h(X)$ and $\mathrm{Pool}_w(X)$ perform adaptive average pooling along the height and width directions, respectively. After concatenation, they are fused via 1×1 convolution to generate direction-related modulation weights, which are activated by Sigmoid to form a spatial gating map. This process enables the network to explicitly model structural patterns with directional distributions, such as strip-shaped and crack-like defects.

In channel interaction, EMA constructs a dual-path response mechanism. One path performs global average pooling on spatially recalibrated features and obtains channel weight distribution through Softmax; the other path performs the same operation on locally convolution-enhanced features. The two paths are respectively multiplied with the corresponding feature matrices to obtain complementary response maps. Finally, the two sets of response results are summed and activated by Sigmoid to form adaptive weights W:

$$W = \sigma(A_1 X_2 + A_2 X_1) \tag{2}$$

where A_1 and A_2 represent attention coefficients generated from global statistics. This design achieves bidirectional coupling of "global guidance of local" and "local feedback to global," effectively mitigating the expression bias that may arise from a single attention path.

In Bottleneck-EMA, input features sequentially undergo 1×1 convolution for dimension reduction, EMA recalibration, and 3×3 convolution to restore channel dimensions; when input and output channels are consistent, residual connections are introduced to stabilize gradient propagation. This structure retains the local modeling capabilities of the standard Bottleneck while providing cross-spatial scale adaptive modulation through EMA.

Furthermore, in C2f-EMA, multiple Bottleneck-EMA units are serially embedded into the C2f backbone path and cascaded with the initial branch features, finally completing feature fusion via 1×1 convolution. Since attention modeling is performed independently within the grouped space, computational complexity increases linearly with C/G, significantly reducing parameter and FLOPs overhead compared to full-channel

https://doi.org/10.64486/m.65.4.15

self-attention mechanisms. Therefore, this structure improves the discriminative ability for complex-shaped targets while preserving the lightweight characteristics of the YOLOv8 backbone.

After introducing C2f-EMA into the P3-P5 multi-scale output layers of the detection head, feature maps at different resolutions all receive enhancements in direction awareness and global consistency, thereby improving the collaborative representation between small-scale details and large-scale structural targets. This grouped multi-path recalibration mechanism provides a stable and efficient feature enhancement paradigm for multi-scale object detection. As presented in Figure 1.

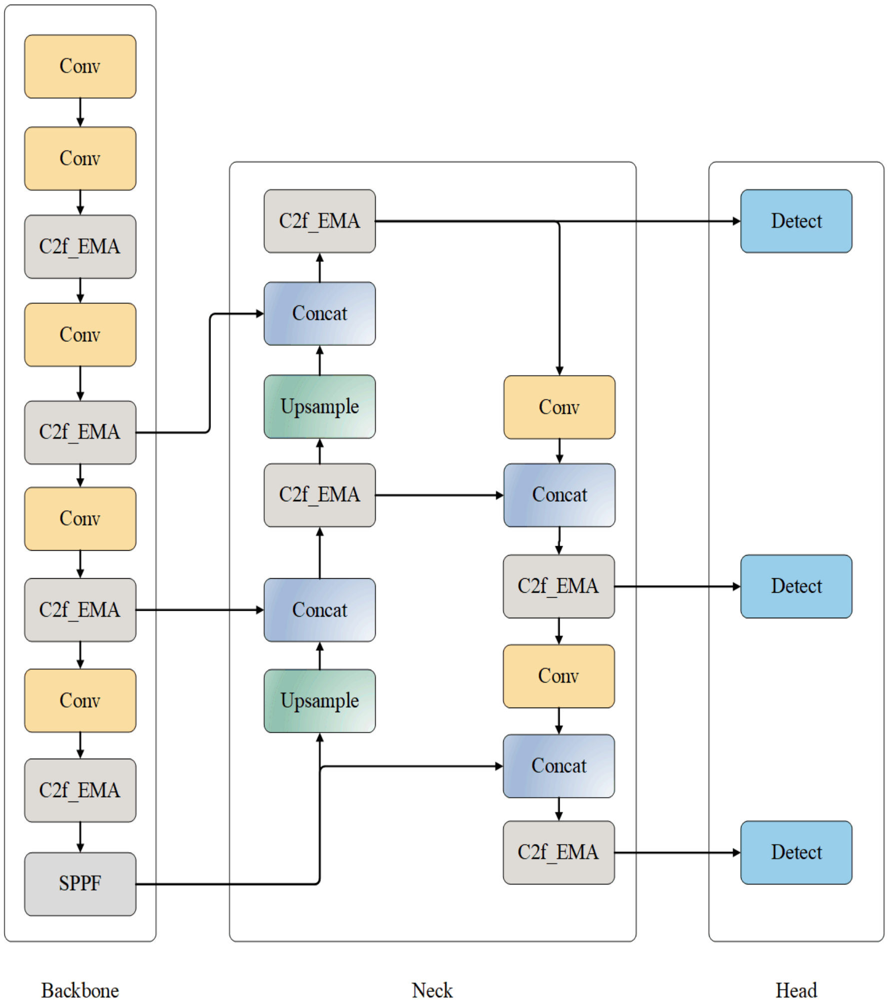

Figure 1. Improved YOLOv8 Network

4. Method of Implementation

The study utilized the NEU-DET steel strip surface defect dataset, which consists of 1,800 images taken from real industrial settings, encompassing six defect categories: Inclusion, Patches, Pitted Surface, Scratches, Crazing, and Rolled-in Scale. The dataset was divided into training, validation, and test sets in an 8:1:1 ratio to maintain data authenticity and diversity. Model parameters were initialized randomly, and the attention weights in the C2f-EMA module were trained concurrently with the network. During training, input images underwent batch preprocessing, involving normalization, random flipping, and scaling for data augmentation, before being inputted into the model for forward computation and error backpropagation to optimize parameters across all layers. Following each training epoch, the current network parameters were frozen, and the interim model's performance was assessed on the validation set to evaluate the impact of the EMA attention mechanism on multi-scale defect feature extraction. Upon completing training, the final model was tested on the test set to gauge the detection accuracy of the improved C2f-EMA module across defects of various scales.

 https://doi.org/10.64486/m.65.4.15

This evaluation continued until the model's performance stabilized, ensuring reliable recognition of small scratches and intricate defects. As presented in Figure 2.

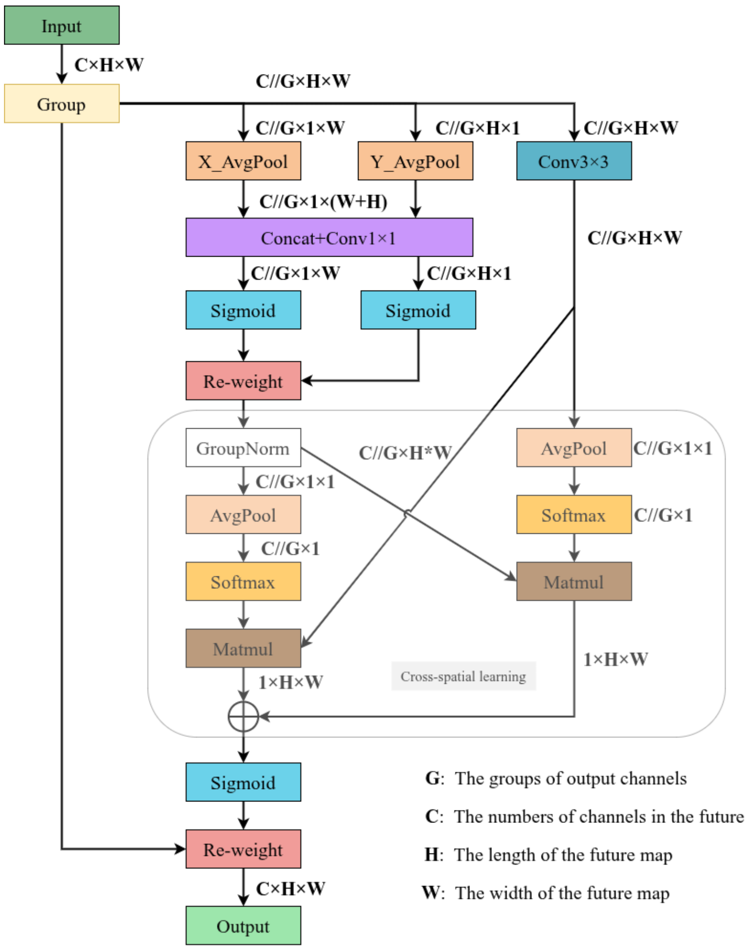

Figure 2. Efficient Multi-scale Attention Network Structure

5. Experiment and Analysis

To further verify the effectiveness and stability of the proposed C2f-EMA structure in improving detection performance, systematic experiments were conducted on the NEU-DET dataset. All models were trained on an Ubuntu 24.04 system with an NVIDIA RTX 4060 GPU, input resolution of 640×640, batch size of 16, and 500 training epochs. Other training strategies remained consistent to ensure experimental fairness.

First, to eliminate performance fluctuations caused by random initialization, five independent repeated trainings were performed for both the baseline YOLOv8 and the improved model. Under identical data partitioning and hyperparameter settings, the mAP@0.5 values of the baseline model in five experiments were 75.9 %, 76.3 %, 76.2 %, 76.0 %, and 76.1 %, with an average of 76.1 % and standard deviation of 0.15 %; the results of the improved model were 77.5 %, 77.7 %, 77.6 %, 77.8 %, and 77.4 %, with an average of 77.6 % and standard deviation of 0.16 %. It can be seen that the standard deviations of both models are below 0.2 %, indicating good convergence stability during training. Further statistical analysis using paired t-test yielded $p = 0.00026$ ($p < 0.01$), indicating that the 1.5 % performance improvement is statistically significant and not caused by random perturbations. This result statistically validates the effectiveness of the C2f-EMA structure in enhancing feature expression. The results are presented in Table 1 and Figure 3.

 https://doi.org/10.64486/m.65.4.15

Table 1. Results of Five Independent Trainings and Statistical Analysis for Baseline and Improved Models (mAP@0.5, %)

Model	Run1	Run2	Run3	Run4	Run5	Mean	Std
YOLOv8	75.9	76.3	76.2	76.0	76.1	76.1	0.15
C2f-EMA	77.5	77.7	77.6	77.8	77.4	77.6	0.16

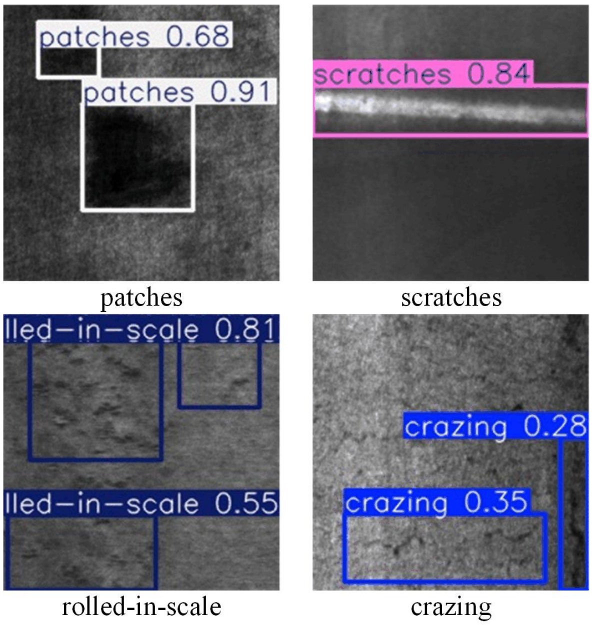

Figure 3. Detection Effect Diagram

When conducting comparative analysis at the category level, the source of the improved model's advantages can be observed more clearly. Compared to the overall mAP of 76.1 % for the original YOLOv8, the introduction of C2f-EMA increased it to 77.6 %. Among them, the Inclusion category improved from 80.5 % to 83.2 %, Pitted surface from 86.2 % to 87.7 %, Scratches from 86.2 % to 87.7 %, and the Rolled-in scale category significantly improved from 66.5 % to 78.0 %. These categories generally have features of small scale or obvious texture directionality, indicating that the direction-aware spatial modeling mechanism and global dependency modulation mechanism introduced by EMA have positive effects on fine-grained structure expression. Although there are slight fluctuations in the Patches and Crazing categories, the overall average precision still achieved stable improvement, indicating that the attention mechanism mainly enhanced target expression capabilities in direction-sensitive and complex background scenarios. As presented in Table 2.

Table 2. Detection Accuracy Comparison between Original YOLOv8 and C2f-EMA on NEU-DET Categories

Model	mAP	Inclusion	Patches	Pitted surface	Scratches	Crazing	Rolled-in scale
YOLOv8	76.1 %	80.5 %	91.4 %	86.2 %	86.2 %	90.0 %	66.5 %
C2f-EMA	77.6 %	83.2 %	89.3 %	87.7 %	87.7 %	90.6 %	78.0 %

To further evaluate the model's competitiveness, comparative experiments were conducted with mainstream detection models under the same dataset and training settings. The two-stage detection framework Faster R-CNN achieved an mAP of 73.4 % on this dataset, Retina Net reached 74.2 %, lightweight single-stage models YOLOv5s and YOLOv7-tiny reached 75.3 % and 76.8 % respectively. The detection accuracy of the original YOLOv8 was 76.1 %, while after introducing C2f-EMA, it reached 77.6 %, achieving the highest detection accuracy among lightweight structures while maintaining real-time detection capabilities. This indicates that the proposed structural improvement can effectively enhance feature representation quality

https://doi.org/10.64486/m.65.4.15

while keeping the main framework unchanged. The performance comparison with different SOTA models is presented in Table 3.

Table 3. Performance Comparison of Different SOTA Models on NEU-DET Dataset

Model	Detection Framework Type	Detection accuracy (mAP@0.5, %)
Faster R-CNN	Two-stage	73.4
Retina Net	Two-stage	74.2
YOLOv5s	Single-stage	75.3
YOLOv7-tiny	Single-stage	76.8
YOLOv8	Single-stage	76.1
C2f-EMA (Ours)	Single-stage	77.6

In response to concerns about model complexity and inference efficiency in the review comments, systematic evaluations of parameter scale, computational complexity, and inference time were conducted. The baseline YOLOv8 has 3.27M parameters, 9.0 GFLOPs of computation, and an average inference time of 6.2 ms per image on the RTX 4060 platform, corresponding to 161 FPS. After introducing C2f-EMA, the model parameters increased to 3.38M, computation rose to 9.3 GFLOPs, and the inference time per image was 6.5 ms, corresponding to 154 FPS. It can be seen that with only about 3 % increase in parameters and computation, the model achieved a 1.5 % accuracy improvement, while the inference speed remained above 150 FPS, meeting the requirements of industrial real-time detection. This indicates that the introduced grouped multi-branch attention modeling strategy enhances expression capabilities without significantly compromising the original lightweight characteristics. As presented in Table 4.

Table 4. Comparison of Model Complexity and Inference Efficiency between Baseline YOLOv8 and C2f-EMA

Model	Parameters (M)	Computation (GFLOPs)	Inference Time (ms)	FPS
YOLOv8	3.27	9.0	6.2	161
C2f-EMA	3.38	9.3	6.5	154

Considering that C2f-EMA is embedded in the P3-P5 multi-scale detection layers, further statistical analysis was conducted on targets of different scales. Experimental results show that the detection accuracy for small-scale targets improved from 63.8 % to 66.2 %, for medium-scale targets from 78.4 % to 79.6 %, while for large-scale targets it remained basically stable. The improvement for small-scale targets is more obvious, indicating that the direction-aware spatial recalibration mechanism can more effectively capture fine-grained defect structures and enhance the expression consistency of feature maps in high-resolution branches. As presented in Table 5.

Table 5. Detection Accuracy Comparison for Targets of Different Scales (mAP@0.5, %)

Model	Small-scale Targets	Medium-scale Targets	Large-scale Targets
YOLOv8	63.8	78.4	89.6
C2f-EMA	66.2	79.6	89.9

In summary, through multiple independent experiments, statistical significance tests, category-level analysis, comparisons with mainstream models, and evaluations of complexity and inference efficiency, it can be confirmed that C2f-EMA significantly enhances multi-scale feature representation capabilities while preserving the lightweight and real-time characteristics of YOLOv8, and has stronger discriminative advantages for small-scale and direction-sensitive defects. This improved structure achieves a relatively balanced effect between performance gain and computational overhead, and has practical industrial deployment value.

https://doi.org/10.64486/m.65.4.15

6. Conclusion

This study focuses on enhancing surface defect detection in hot-rolled steel strips by incorporating an Efficient Multi-scale Attention (EMA) mechanism into the C2f module of the YOLOv8 model, resulting in the C2f-EMA module. This integration improves the adaptive representation of features in both channel and spatial dimensions. Experimental results indicate that the enhanced model achieves higher detection accuracy across the six defect categories, with the overall mean Average Precision (mAP) rising from 76.1 % to 77.6 %. Particularly, there are notable enhancements in detecting small- and medium-scale defects like Inclusion, Scratches, and Crazing, highlighting the effectiveness of the EMA attention mechanism in improving feature representation for intricate defects. In conclusion, the incorporation of the EMA attention mechanism into the C2f module significantly enhances the discriminative ability of multi-scale features, providing a streamlined and effective solution for detecting surface defects in hot-rolled steel strips. Moreover, this study offers valuable insights and guidance for the future utilization and refinement of deep attention mechanisms in industrial defect detection applications.

Acknowledgments: This work was supported by innovation and entrepreneurship training program for college students of University of Science and Technology Liaoning 2025.

References

[1] R. Gelaky, M. Usher, and K. Warwick, "Industrial inspection-the ease-of-use/flexibility trade-off," In IEE Colloquium on HCI: Issues for the Factory, Landon, UK, 1991, pp. 5/1-5/3.

[2] B. Chen, "Research overview of YOLO series object detection algorithms based on deep learning," Journal of Computing and Electronic Information Management, Vol.15, No.3, pp. 84-92, 2024, https://doi.org/10.54097/p81rtv77.

[3] T. Yang, L. Chang, J. Yan, J. Li, Z. Wang, and K. Zhang, "A Survey on Foundation-Model-Based Industrial Defect Detection," arXiv preprint arXiv:2502.19106, 2025, https://doi.org/10.48550/arXiv.2502.19106.

[4] N.M. Muriyah, J. H. Sim, and A. Yulianto, "Evaluating YOLOv5 and YOLOv8: Advancements in Human Detection," Journal of Information Systems and Informatics, Vol. 6, No. 4, pp. 2999-3015, 2024, https://doi.org/10.51519/journalisi.v6i4.944.

[5] H. Li, A. Wu, Z. Jiang, F. Liu, and M. Luo, "Improving Object Detection in YOLOv8n with the C2f-f Module and Multi-Scale Fusion Reconstruction," In 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference, Chongqing, China, 2024, Vol. 6, pp. 374-379, https://doi.org/10.1109/imcec59810.2024.10575292.

[6] J. Lehr, M. Pape, J. Philipps, F. Scholler, and J. Krüger, "Limitations of anomaly detection: beyond which size defects can be reliably recognized," In Sixteenth International Conference on Machine Vision, Yerevan, Armenia, 2024, Vol. 13072, pp. 132-137, https://doi.org/10.1117/12.3023615.

[7] D. N. Klepov, V. V. Yashin, I. A. Latushkin, E. V. Aryshenskii, and Y. A. Erisov, "Determination of Conditions for the Occurrence of Delamination, a Surface Defect Occurring during Hot Rolling of Aluminum Alloy Strips," Russian Metallurgy, Vol.2024, No.7, pp. 1694-1699, 2024, https://doi.org/10.1134/s0036029524702963.

[8] Z. Zhao, "A review of deep learning-based detection and segmentation of steel surface defects: advances in resolving intra-class differences and inter-class similarities," Advances in Engineering Innovation, Vol.16, No.6, pp. 130-134, 2025, https://doi.org/10.54254/2977-3903/2025.24461.

[9] H. Dyja, K. Sobczak, and A. Kawałek, "The influence of the shape of grooves on the behavior of internal material discontinuities in continuous S355J2G3 steel strands during rolling," Metalurgija, Vol.53, No.4, pp. 501-504, 2014.

[10] W. Shen, G. Cheng, G. Wang, "Formation and Deformation Mechanism of Mountain-Like Surface Crack on the Hot-Rolled Thick Plate of HSLA Steel," Steel Research International, Vol.94, No.1, pp. 2200498, 2023, https://doi.org/10.1002/srin.202200498.

[11] И. П. Щербаков, А. Г. Кадомцев, и А. Е. Чмель, "Временной паттерн накопления микротрещин при ударном повреждении пористой керамики SiC," Письма в журнал технической физики, Vol.48, No.11, pp. 45-48, 2022, https://doi.org/10.21883/pjtf.2022.11.52614.19140.

https://doi.org/10.64486/m.65.4.15

[12] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, "Advanced post-processing for scanned ultrasonic arrays: Application to defect detection and classification in non-destructive evaluation," Ultrasonics, Vol.48, No.6-7, pp. 636-642, 2008, https://doi.org/10.1016/j.ultras.2008.07.019.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 779-788, https://doi.org/10.1109/cvpr.2016.91.

[14] S. V. Viraktamath, M. Yavagal, and R. Byahatti, "Object detection and classification using YOLOv3," International Journal of Engineering Research & Technology, Vol.10, No.2, pp. 197-202, 2021.

[15] R. Khanam, and M. Hussain, "What is YOLOv5: A deep look into the internal features of the popular object detector." arXiv preprint arXiv:2407.20892, 2024.

[16] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, “A review on yolov8 and its advancements,” In International conference on data intelligence and cognitive informatics, Springer, Singapore, 2024, pp. 529-545, https://doi.org/10.1007/978-981-99-7962-2_39.

[17] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning." In ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing, Rhodes Island, Greece, 2023, pp. 1-5, https://doi.org/10.1109/icassp49357.2023.10096516.

[18] M. Wu, J. Peng, X. Yu, H. Xu, and H. Sun, "EFEN-YOLOv8: Surface defect detection network based on spatial feature capture and multi-level weighted attention," PLoS One, Vol.21, No.1, pp. e0339617, 2026, https://doi.org/10.1371/journal.pone.0339617.

[19] L. Li, Y. Ma, and L. Su, “CM -YOLO: A Highly Accurate YOLOv5 S-Based Model for Steel Surface Defect Detection,” In 2025 8th International Conference on Advanced Algorithms and Control Engineering, Shanghai, China, 2025, pp. 972-976, https://doi.org/10.1109/ICAACE65325.2025.11018999.

[20] T. Chaffey, F. Forni, and R. Sepulchre, “Loop Shaping with Scaled Relative Graphs,” arXiv preprint arXiv:2208.04880, 2022, https://doi.org/10.48550/arXiv.2208.04880.

[21] H. Wu, J. Liu, Z. Zha, Z. Chen, and X. Sun,“Mutually Reinforced Spatio-Temporal Convolutional Tube for Human Action Recognition,” In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 2019, pp. 968-974, https://doi.org/10.24963/ijcai.2019/136.

 https://doi.org/10.64486/m.65.4.15