Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 2, No. 1, 2026

Research on a YOLOv8-Based Metal Residue Detection Algorithm

Qingqi He, Siyuan Zhao, Yunlong Liu, Yongxing Li, Yi Yang, Fangfang Hou

School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

Abstract: To address the recognition challenges caused by the tiny size of metal residues, high reflectivity, and severe background texture interference in industrial production environments, this paper proposes a YOLOv8-based metal residue recognition algorithm. First, to overcome the scarcity of high-quality samples in industrial sites, a metal residue dataset containing multiple complex backgrounds such as welds and oil stains is constructed by simulating actual working conditions. Targeted data augmentation techniques are employed to enhance the model's generalization ability in extreme environments. Second, to tackle the problems of extremely low pixel occupancy of metal targets and the difficulty of feature extraction, a lightweight coordinate attention mechanism is introduced into the backbone network. By capturing cross-channel directional and positional information, the model's perception and extraction capabilities of subtle metal edge features are strengthened. Experimental results show that the improved algorithm achieves a mean average precision (mAP) of 0.724 on the self-built dataset. While maintaining real-time inference speed, it can accurately identify metallic foreign objects with extremely small diameters, effectively reducing the miss rate and false alarm rate under complex working conditions, thus providing reliable technical support for the deployment of industrial precision inspection and automated rejection systems.

Keywords: Metal Residue Recognition, YOLOv8, Small Object Detection, Deep Learning, Feature Fusion.

1. Introduction

As industrial automation keeps improving, more and more factories are using automated equipment for production, and the demands on product quality inspection are getting higher. During machining, metal cutting, and other industrial processes, tiny metal residues often appear. If these residues are not detected and removed in time, they can affect product quality and might also interfere with equipment operation. So how to detect metal residues quickly and accurately has become an important issue in industrial inspection. Today, many industrial settings still rely on manual inspection for metal residues. Though manual inspection is simple to carry out, it is easily affected by the worker's experience, working condition, and environmental factors. Long hours of repeated inspection also cause fatigue, leading to missed detections and false alarms. [1] Meanwhile, manual inspection is slow, and in high-speed production lines it can hardly meet real-time demands. As a result, traditional manual inspection is becoming less and less suited to the needs of modern industrial development.

In recent years, deep learning has advanced rapidly in image recognition and object detection. Among them, convolutional neural network based detectors are now widely used in industrial vision inspection [2]. The YOLO (You Only Look Once) family of detectors is known for its fast speed and good real-time performance [3]. Compared with traditional methods, YOLO automatically extracts image features and still works well under complex conditions, so it is widely applied in industrial defect detection, security monitoring, autonomous driving, and other fields [4].

In this paper, we focus on detecting metal residues in industrial environments and study a YOLOv8-based recognition method. By building a metal residue dataset and training/testing the YOLOv8 model, we aim to automatically recognise and locate metal residues. We also analyse and optimise the detection results to improve accuracy and efficiency. This work has practical value for raising the level of automated industrial inspection, reducing the cost of manual detection, and promoting intelligent manufacturing. Theory of Thermal Deformation of Solid Lubricated Gears

2. The Algorithmic Model

2.1. YOLOv8 Network Model

YOLOv8 is an object detection algorithm proposed by the Ultralytics team in 2023 [5], and is a relatively new version in the YOLO series. Its network structure consists of three main parts: the backbone, the neck, and the head. In the backbone, the C2f module replaces the C3 module from earlier versions. The design of the C2f module is related to gradient flow and information reuse. In the old C3 module, input features went through two branch convolutions and were then concatenated, with only a simple skip connection. The C2f module introduces more branches: the feature map is split into several segments, each is convolved separately, and then they are fused together. This improves feature reuse. During backpropagation, gradients can bypass more non-linear layers, which alleviates the vanishing gradient problem in deep networks. At the same time, the C2f module reduces redundant computation while preserving the receptive field. Its parameter count is about 15% less than that of C3, yet its feature representation ability is better [6]. This is beneficial for small targets like metal residues, which have weak edge information — richer gradient information helps preserve high-frequency details.

The neck continues to use the FPN (Feature Pyramid Network) plus PAN (Path Aggregation Network) structure, fusing deep semantic information with shallow positional information to improve detection across different scales.

More specifically, FPN follows a top-down path: semantic information from deep feature maps is gradually passed to shallow layers, giving shallow features stronger category discrimination ability [7], so that high-level features also retain precise edge localization cues. This two-way fusion is critical for detecting metal residues: in an image, a residue often occupies less than 0.5% of the pixels [8]. Relying only on shallow features makes it easy to blend with background noise, while deep features, after many down-sampling steps, barely respond to small targets. The FPN-PAN structure maintains a high-resolution response for small targets (with a down-sampling factor as low as 8x) while incorporating global context information, effectively mitigating feature misalignment.

The detection head adopts a decoupled design, handling classification and regression tasks separately. It also abandons anchor-based methods in favor of an anchor-free approach, which simplifies model parameters and speeds up convergence. While maintaining a fast detection speed, YOLOv8 performs better than earlier versions at recognizing small targets and targets under complex backgrounds. That is why it is widely used in industrial inspection, defect detection, and other real-world scenarios. Its strength lies in performing both object localization and category recognition in a single forward pass, giving it fast detection speed and good real-time performance.

Compared with traditional object detection methods, YOLO does not need to generate region proposals before classification; instead, it processes the whole image directly, thus improving detection efficiency. The coordinate attention mechanism embeds position information into channel attention, enhancing the model's localization capability with almost no extra computational cost [9]. As the algorithm continues to evolve, YOLO models have shown significant improvements in both detection accuracy and speed. Today, YOLO is widely applied in industrial inspection, security monitoring, autonomous driving, and other fields. In this paper, we use the YOLOv8 model to recognise metal residues, learning the target features from images to automatically detect and locate metal residues, as shown in Figure 1.

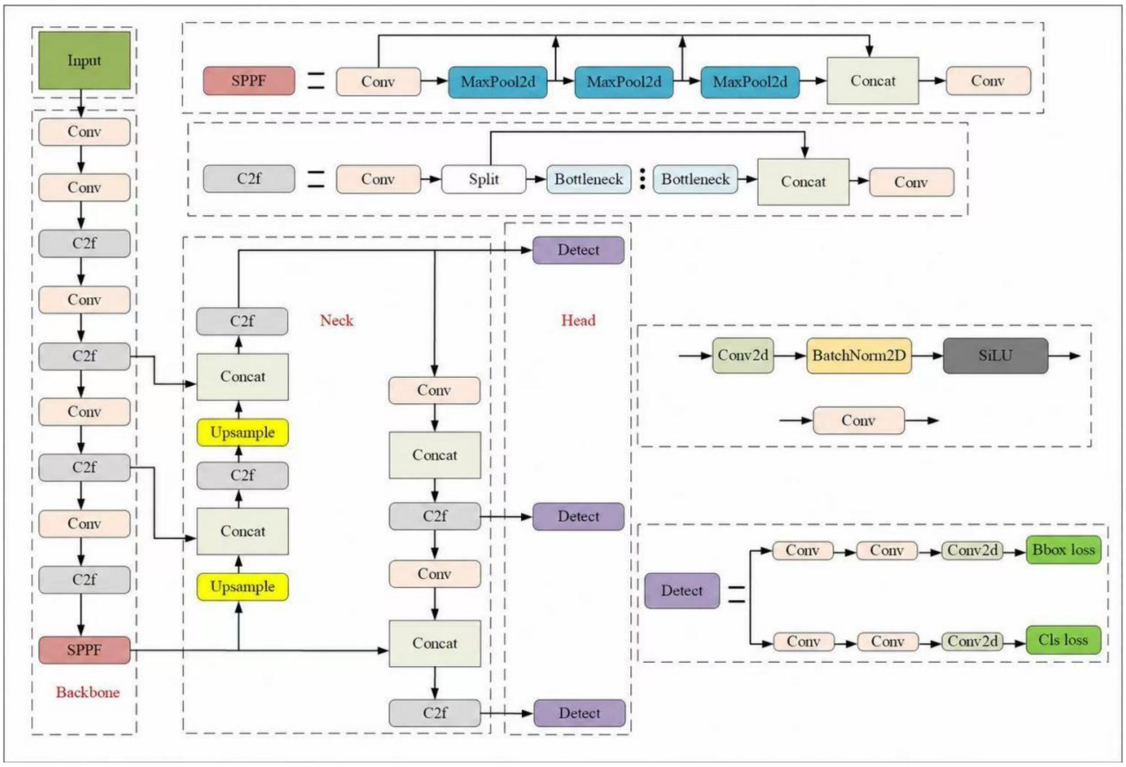

Figure 1. YOLOv8 network architecture diagram

3. Experiments and Analysis

3.1. Experimental Dataset

In industrial manufacturing, the inclusion of metal residues (such as metal shavings, broken nails, gaskets, etc.) is one of the main hazards leading to product recalls and safety accidents. These foreign objects are often tiny in size and varied in shape, and they can be easily occluded by the product itself or confused with complex backgrounds. Traditional manual visual inspection and metal detectors struggle to achieve both speed and accuracy, posing a great challenge to online quality control on production lines. To improve inspection efficiency and automation, this study uses an industrial camera to build an image acquisition system above the conveyor belt and employs the YOLO algorithm for real-time recognition and localization of metal residues [10]. Compared with conventional physical detection methods, the deep vision-based detection approach offers advantages such as non-contact operation, localization capability, and rich information, which is of great significance for ensuring product safety and reducing labor costs.

To evaluate the generalization ability of the model, the dataset was divided into a training set and a validation set at a ratio of 5:1. About 83% of the data (1,540 images) was used as the training set for model training, and the remaining 17% (300 images) served as the validation set to evaluate model performance. Specifically, two new folders named "train" and "val" were created in the "mydata" folder, and the corresponding images were placed into them. Random sampling was used to split the dataset, ensuring a relatively balanced distribution of the two label categories between the training and validation sets, thereby avoiding the impact of class imbalance on model performance.

3.2. Experimental Environment and Experimental Evaluation Metrics

Regarding the selection of the deep learning framework, this paper ultimately adopts the open-source PyTorch framework. The underlying architecture of the PyTorch framework provides comprehensive CUDA acceleration support for NVIDIA GPUs, thereby maximally unleashing the computational performance of the hardware. The core YOLOv8 detection algorithm employed in this paper has been deeply developed and maintained based on the PyTorch ecosystem since the initial stage of its official design [11]. Selecting this framework ensures perfect compatibility with the underlying code of the algorithm, guarantees operational stability during model training and inference, and consequently furnishes a reliable software foundation for the experimental implementation of this study.

To objectively evaluate the detection performance of the YOLOv8 model on the metal residue recognition task, this paper uses Precision, Recall, and mean Average Precision (mAP) as evaluation metrics [12]:

(1) Precision: the proportion of correctly predicted positive samples among all samples predicted as positive, reflecting the model's ability to avoid false positives. The formula is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3-1}$$

(2) Recall: the proportion of actual positive samples that are correctly predicted as positive, reflecting the model's ability to avoid false negatives. The formula is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3-2}$$

Where:

TP (True positive)-the number of training samples that actually belong to the positive class and are correctly classified as positive during detection.

TN (True Negative)-the number of training samples that actually belong to the negative class and are correctly classified as negative during detection.

FP (False Positive)-the number of training samples that actually belong to the negative class but are incorrectly classified as positive during detection.

FN (False Negative)-the number of training samples that actually belong to the positive class but are incorrectly classified as negative during detection.

(3) Average Precision (AP): an evaluation metric used to assess the performance of object detection or image classification models. It measures the average accuracy of the model across different classes. The formula is:

$$\text{AP} = \int_0^1 \text{P(R)dR} \tag{3-3}$$

(4) Mean Average Precision (mAP): the average of precision values at different recall levels. It is a commonly used performance metric in object detection algorithms, taking into account both recall and precision. The formula is:

$$\text{mAP} = \frac{\sum_{i=1}^{n} \text{AP(i)}}{\text{n}} \tag{3-4}$$

(5) Frames Per Second (FPS): the number of frames the model processes per second, which directly affects the feasibility of the algorithm in real-time applications. A higher FPS value indicates that the model can detect more targets per second, meaning faster detection speed and better real-time performance. FPS measurement is affected by the test hardware: the stronger the GPU performance, the faster the processing speed. Therefore, when evaluating FPS, it should be measured under the same hardware configuration to ensure the accuracy and comparability of the test results.

(6) F1-score: the harmonic mean of precision and recall. It is used to comprehensively evaluate the performance of a binary classification model and aims to balance the trade-off between precision and recall. The formula is:

$$\text{F1} = \frac{\text{2TP}}{\text{2TP} + \text{FP} + \text{FN}} \tag{3-5}$$

Precision indicates the proportion of correctly identified targets among the model's predictions, measuring the accuracy of the detection results. Recall indicates the proportion of real targets that are correctly detected, reflecting the model's ability to detect targets. mAP comprehensively considers both classification and localization performance and is a commonly used evaluation metric in object detection. The higher the Precision, the fewer false positives the model produces. The higher the Recall, the fewer false negatives the model produces. The higher the mAP, the better the overall detection performance of the model. Therefore, these metrics can provide a relatively comprehensive evaluation of the model's detection effectiveness on the metal residue recognition task.

3.3. Model Performance Evaluation

As shown in Table 1, at the initial stage of training (Epoch 1), the model performance was very low, with Precision, Recall, and mAP50 reaching only 0.310, 0.306, and 0.187, respectively. At this point, the network weights were randomly initialized, so the model could not effectively extract features of metal residues, resulting in many false positives and false negatives. After 50 epochs of iterative optimization, all three metrics increased significantly: Precision rose to 0.793, Recall to 0.611, and mAP50 notably improved to 0.690, indicating that the model had preliminarily learned the discriminative features of metal residues and its detection capability had basically taken shape [13]. When training continued to 100 epochs, Precision reached 0.813, Recall reached 0.689, and mAP50 increased to 0.724. Although the growth rate slowed compared with the first 50 epochs, the metrics still showed a steady upward trend, and the model gradually converged. The final Precision and Recall are at relatively high levels, and the mAP50 exceeds 0.72, demonstrating that the model effectively reduces false detections while ensuring a high true detection rate. Overall, the detection performance is good, and the model is already capable for practical use in metal residue recognition tasks.

Table 1. Comparison of performance metrics before and after model training

Metric	Epoch1	Epoch50	Epoch100
precision	0.310	0.793	0.813
recall	0.306	0.611	0.689
mAP50	0.187	0.690	0.724

3.4. Model Detection Performance Demonstration

To further verify the model’s detection capability in real-world scenarios, typical samples from the validation set were selected for testing. The detection results show that YOLOv8 can accurately identify metal residues in the images and localize them using bounding boxes [14]. Most targets are correctly detected, and the detection boxes generally align well with the target positions, indicating good detection accuracy. This paper selects several representative images for performance demonstration, including images of cracks, oil stains, pits, etc.

The generated metal residue detection results are shown in Figures 2-4.

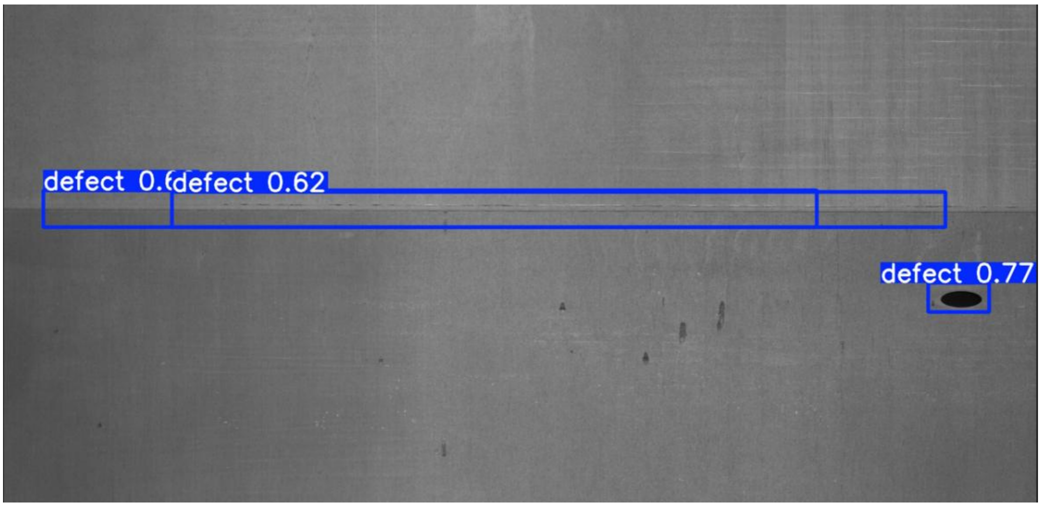

Figure 2. Crack detection

As shown in Figure 2, the algorithm can identify and locate cracks on the metal surface, with confidence scores of 0.62 and 0.77, indicating that it can effectively detect surface damage during the production process.

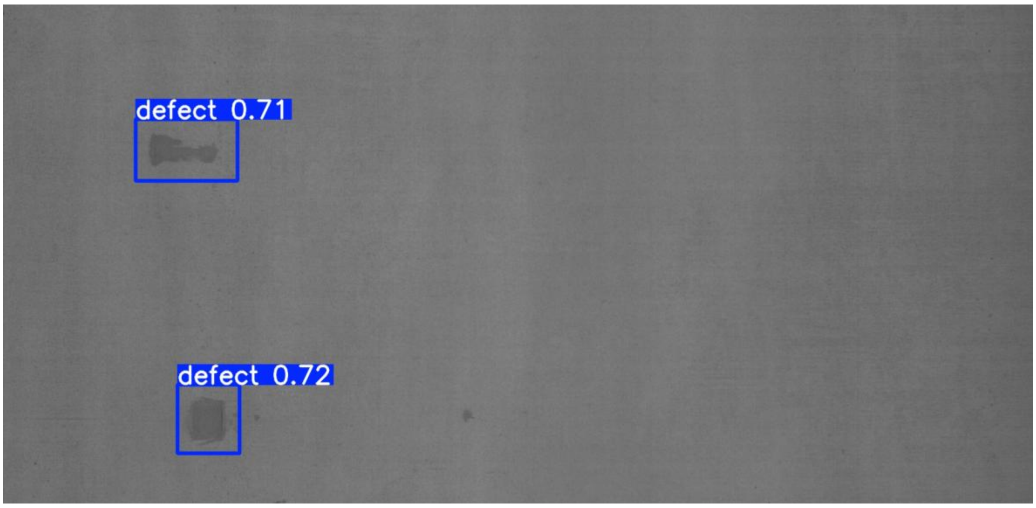

Figure 3. Oil stain detection

As shown in Figure 3, the algorithm can identify and locate oil stains on the metal surface, with confidence scores of 0.71 and 0.72, indicating that it can effectively detect surface stains during the production process.

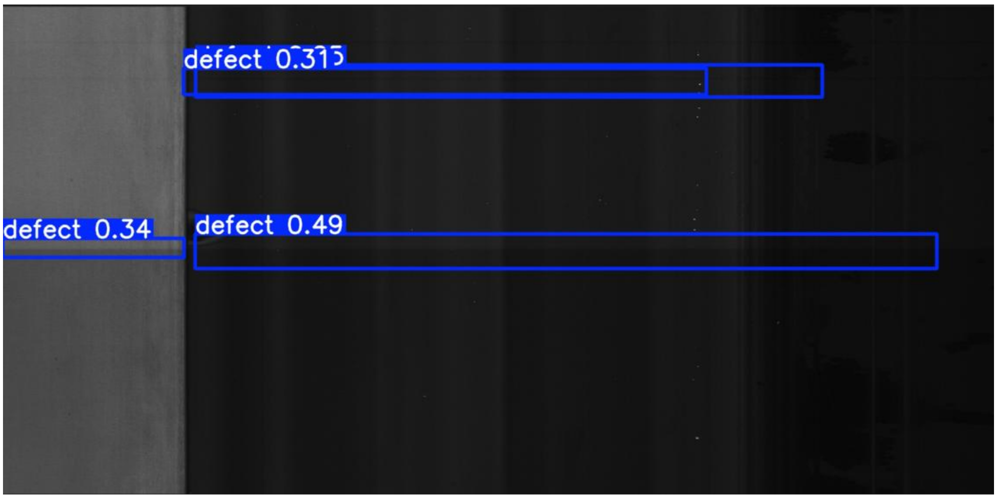

Figure 4. Crescent-shaped pit detection

As shown in Figure 4, the algorithm can identify and locate pits on the metal surface, with confidence scores of 0.31, 0.34, and 0.49. In very low-contrast environments, the confidence scores are limited due to blurred features, resulting in lower values, but the system can still localize the positions of the scratches.

4. Conclusion

This paper focuses on YOLO-based metal residue recognition. Given the small size, complex backgrounds, and susceptibility to lighting interference of metal residue targets in industrial scenarios, we constructed a multi-target recognition scheme for tiny metal residue detection and implemented a YOLO-based object detection method to effectively identify and locate metal residues. The main achievements of this study are as follows:

(1) Construction of a metal residue detection dataset and completing model detection: The acquisition and annotation of metal residue images in industrial scenarios were accomplished, and data augmentation methods were employed to enhance data diversity, thereby providing an effective data foundation for model training [15]. The model is capable of rapidly identifying and localizing targets within input images, demonstrating favorable real-time performance and engineering application value.

(2) Implementation of a YOLO-based detection model: Based on the YOLO framework, we built a metal residue detection system capable of rapidly identifying and localizing targets in input images, offering good real-time performance and engineering application value.

(3) Validation of the model's effectiveness and feasibility: Through the analysis of experimental results, the model achieves a Precision of 0.813 and a Recall of 0.689, enabling it to perform target detection tasks with reasonable accuracy. This provides directions for future improvements, such as further optimizing the feature extraction architecture and enhancing multi-scale detection capabilities. It demonstrates that the proposed method possesses certain practical application value.

In summary, in response to the demand for tiny metal residue recognition, this paper has specifically optimized and improved existing algorithmic models, proposing the YOLOv8 model. This model demonstrates excellent performance in small object recognition and significantly advances the recognition and classification accuracy of various tiny metal residues such as cracks, pits, and oil stains. The proposed solution further improves work efficiency in industrial production and provides important theoretical and practical support for the development of related technologies.

References

[1] Zhang, Z., Zhang, J., Chen, Q., et al. (2026). Unification of closed-open industrial detection scenarios: New large-scale benchmarks, challenges and baselines. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2026.3680569

[2] Duan, G., Pan, X., Wang, D., et al. (2026). Surface defect detection of mosaic ceramic walls based on deep learning. Journal of Computing in Civil Engineering, 40(5). https://doi.org/10.1061/JCCEE5.CPENG-7309

[3] Chen, W., Deng, Q., Xiang, Z., et al. (2026). FDAYOLOv8: refined small object detection in unmanned aerial vehicle imagery. Scientific Reports. https://doi.org/10.1038/s41598-026-51902-3

[4] Gupta, V., & Mangal, D. (2026). A deep learning approach based small object detection for dynamic digital twins in smart manufacturing. Computers & Electrical Engineering, 136, 111246. https://doi.org/10.1016/j.compeleceng.2026.111246

[5] Zhang, W., Du, Y., & Wu, Z. (2026). Helmet-wearing detection with small object recognition via improved YOLOv8 in complex environments. Digital Signal Processing, 176, 106066. https://doi.org/10.1016/j.dsp.2026.106066

[6] Yang, W., Zhang, B., Guo, S., et al. (2026). Small target detection of floating objects in river channels based on improved YOLOv7. Scientific Reports, 16(1), 11423. https://doi.org/10.1038/s41598-026-40688-z

[7] Hu, P., Dong, L. (2023). A multi-scale pyramid feature fusion-based object detection method for remote sensing images. International Journal of Remote Sensing, 44(24), 7790–7807. https://doi.org/10.1080/01431161.2023.2288947

[8] Ma, J., Ge, S., Dong, Y., et al. (2023). Instance segmentation algorithm based on fine-grained feature perception and cross-path aggregation. Knowledge-Based Systems, 276. https://doi.org/10.1016/j.knosys.2023.110754

[9] Macías, C. M. F., Maldonado, S. J. F., Álvarez, M. P., et al. (2026). Torchmil: A PyTorch-based library for deep multiple instance learning. Neurocomputing, 680, 133286. https://doi.org/10.1016/j.neucom.2026.133286

[10] Ling, J., Wu, H., & Peng, J. (2026). Bio-inspired small object motion detection visual network based on motion and local entropy information. Engineering Applications of Artificial Intelligence, 177, 114877. https://doi.org/10.1016/j.engappai.2026.114877

[11] Raj, E. C. E., & Shaji, S. R. (2026). Energy Harvesting in Mobile Networks: A Microstrip Circular Patch Antenna Design for Efficient Radio Frequency Signal Capture. International Journal of Communication Systems, 39(9), e70502. https://doi.org/10.1002/dac.70502

[12] Tian, H. (2025). A comprehensive review of YOLO object detection algorithms. Electronics Science Technology and Application, 12(2). https://doi.org/10.59429/esta.v12i2.10579

[13] Muhammad, S., & Guo, Z. (2022). Deep Residual Learning for Image Recognition: A Survey. Applied Sciences, 12(18), 8972. https://doi.org/10.3390/app12188972

[14] Wu, S., Kong, H., & Xu, Z. (2026). HYOLO: A lightweight and efficient detection framework for small objects in aerial images. Expert Systems with Applications, 328, 132794. https://doi.org/10.1016/j.eswa.2026.132794

[15] Zeng, W. (2024). Image data augmentation techniques based on deep learning: A survey. Mathematical Biosciences and Engineering, 21(6), 6190–6224. https://doi.org/10.3934/mbe.2024272