ITC 4/53

Information Technology and Control
Vol. 53 / No. 4/ 2024
pp. 1028-1041
DOI 10.5755/j01.itc.53.4.37632

Enhanced Feature Extraction with AL-YOLOv9s Lightweight Model: Application in Key Component Recognition Within Highly Integrated Device Environments

Received 2024/06/14 | Accepted after revision 2024/09/13

HOW TO CITE: Wang, Y., Pan, W., Wang, L., Zhang, P. (2024). Enhanced Feature Extraction with AL-YOLOv9s Lightweight Model: Application in Key Component Recognition Within Highly Integrated Device Environments. *Information Technology and Control*, 53(4), 1028-1041. https://doi.org/10.5755/j01.itc.53.4.37632

Enhanced Feature Extraction with AL-YOLOv9s Lightweight Model: Application in Key Component Recognition Within Highly Integrated Device Environments

Yang Wang
School of Electrical Engineering, Naval University of Engineering, Wuhan 430030, China

Wei Pan
School of Power Engineering, Naval University of Engineering, Wuhan 430030, China

Liming Wang
School of Electrical Engineering, Naval University of Engineering, Wuhan 430030, China

Peng Zhang
People's Liberation Army Unit 92808, Haikou 570100, China

Corresponding author: panwei19860418@126.com (W.P.)

In environments containing highly integrated devices, accurately monitoring the status of circuit breaker lockouts is essential for maintaining the stability of power systems. Traditional detection methods are often inadequate due to complex equipment configurations and severe operational challenges. This paper presents an enhanced detection model, the AL-YOLOv9s, which improves the efficiency and accuracy of detecting circuit breaker lockouts. The AL-YOLOv9s model is based on the advanced YOLOv9s algorithm and incorporates an enhanced efficient multi-scale attention module to boost feature extraction capabilities. It also integrates channel and spatial attention mechanisms to optimize the feature fusion process, thereby improving detection performance. Additionally, the model has been optimized to a size of 4.7M, making it suitable for lightweight field applications without compromising accuracy. Experimental results demonstrate that the AL-YOLOv9s model achieves high standards in accuracy and portability, thus offering an effective and practical solution for lockout detection.

KEYWORDS: Circuit breaker lockout, YOLOv9s, AL-YOLOv9s, Multi-scale attention mechanism, Channel attention, Spatial attention, Lightweight model, Object detection.

1. Introduction

In the era of rapid advancements in industrial and information technology, the integration of smart devices with Internet of Things (IoT) technologies is increasingly driving the progression of industrial automation and intelligent manufacturing [2, 4]. Modern industrial systems are characterized by highly integrated production lines, intelligent monitoring systems, and remote operation platforms [8]. Within these complex systems, core components such as sensors, actuators, and control units are pivotal for ensuring stable system operations, enhancing production efficiency, and maintaining workplace safety. Consequently, the accurate identification and real-time monitoring of these critical components are significant technological challenges in the industrial sector [9].

Particularly in the maintenance and operation of highly integrated equipment, effective monitoring of the status of electrical equipment locks is crucial for ensuring both equipment stability and personnel safety [3, 21]. Failures in accurately detecting these states can result in equipment malfunctions and serious safety incidents. In such settings, the role of target detection technology becomes critical. Various target detection algorithms have been widely applied to monitor the status of electrical equipment, providing precise and timely monitoring data [22, 23]. Advancements in artificial intelligence have led to the evolution of target detection from manual feature-based methods like HOG [16] to sophisticated deep learning-based algorithms such as Faster R-CNN, which utilizes Region Proposal Networks (RPN) for rapid and accurate target detection while maintaining high real-time performance [6, 7]. In highly integrated industrial environments, target detection confronts challenges including device diversity, complex backgrounds, varying target sizes, and demands for real-time processing. However, applying generic target detection models to specific scenarios, like electrical lock identification in substations, often encounters obstacles such as large model sizes and poor portability, which can significantly hinder detection efficiency and accuracy [1, 19].

YOLO (You Only Look Once) is a widely recognized target detection algorithm that stands out for its rapid and efficient performance [12]. A key advantage of the YOLO algorithm is its ability to achieve high detection rates rapidly while maintaining accuracy, thereby making it highly suitable for real-time video surveillance applications. Since its debut, the YOLO framework has evolved through multiple iterations [13], culminating in the latest version, YOLOv9 [11]. Each iteration has brought enhancements such as optimized network structures, increased detection accuracy and speed, and improved capability to detect objects of varying sizes. Zhu et al. introduced the C2DEM-YOLO model, which integrates the C2Dense and EMA modules to significantly boost detection precision [24]. However, the practical applicability of this model in real-world settings remains to be validated, and its performance within deeper network architectures is still constrained. Hao et al. developed the YOLOv5-EMA model, which incorporates an EMA module to better detect small and occluded objects, though the model still lacks optimization for lightness and robustness [5]. Yan et al. created a lightweight multi-object joint training model that utilizes complex network feature mappings to enhance information entropy and employs thermal pixel methodology to address the imbalance between foreground and background, using pixel temperature to indicate the likelihood of object presence [20]. This model is well-suited for devices with limited computing resources and surpasses traditional lightweight models and knowledge distillation methods in accuracy and reliability. Despite these benefits, there is room for improvement, especially in managing the added complexity of feature mapping, which could impact overall model performance. Wang et al. designed the GF-YOLOv7 network model specifically for the recognition of substation bouncing locks [14]. This model integrates the MobileViT module and CBAM attention mechanism to achieve both lightness and enhanced performance.

Xiao et al. built the DCFormer framework, replaced the multi-head attention module (MHA), the core component of Transformer, and proposed dynamically combinable multi-head attention (DCMHA). DCMHA releases the fixed binding of the search selection loop and the transformation loop of the MHA attention head, allowing them to be dynamically combined according to the input, fundamentally improving the expressive power of the model [18]. Wang et al. proposed a content-aware mixer (CAMixer) to meet the needs of large image (2K-8K) super-resolution (SR) and overcome the shortcomings of existing methods [15]. CAMixer allocates convolution and deformable window attention in a content-aware manner, and uses a learnable predictor to generate offsets, classification masks, and convolutional attention for window distortion, thereby adaptively including more useful textures and improving the representation power of convolution. A global classification loss is introduced to improve the accuracy of the predictor. By stacking CAMixers, CAMixerSR is formed, which performs well in large image SR, lightweight SR and omnidirectional image SR.

Despite these advancements in specific tasks, these models still confront common challenges such as managing model complexity, validating adaptability to real-world environments, and generalization of the model.

This study employs the most recent installment from the YOLO series, YOLOv9, as the base model. It enhances the feature extraction capabilities of the backbone network through the integration of an improved spatial learning efficient attention module and employs a channel attention mechanism to guide feature fusion. In response to the needs for task portability and edge deployment, the network has been modified to achieve significant lightweight improvements.

The main contributions of this paper are summarized as follows:

- The EMA attention module has been enhanced and integrated into the YOLOv9 network to improve the feature extraction capabilities of the backbone network.
- Feature fusion in the proposed network is guided by both spatial and channel attention mechanisms.
- Addressing the challenge of lock recognition, an improved AL-YOLOv9s network is proposed. Experimental results indicate that the proposed network achieves higher accuracy while significantly reducing model weight.
- The structure of this paper is organized as follows: Section 2 provides a detailed explanation of the YOLOv9 model and the self-attention mechanisms. Section 3 elaborates on the AL-YOLOv9s algorithm, including the introduction of improved, efficient multi-scale, channel, and spatial attention modules to enhance feature extraction and optimize feature fusion, achieving a lightweight design. Section 4 presents a comparative analysis of the AL-YOLOv9s algorithm with other detection methods. Finally, Section 5 concludes the paper and outlines future research directions.

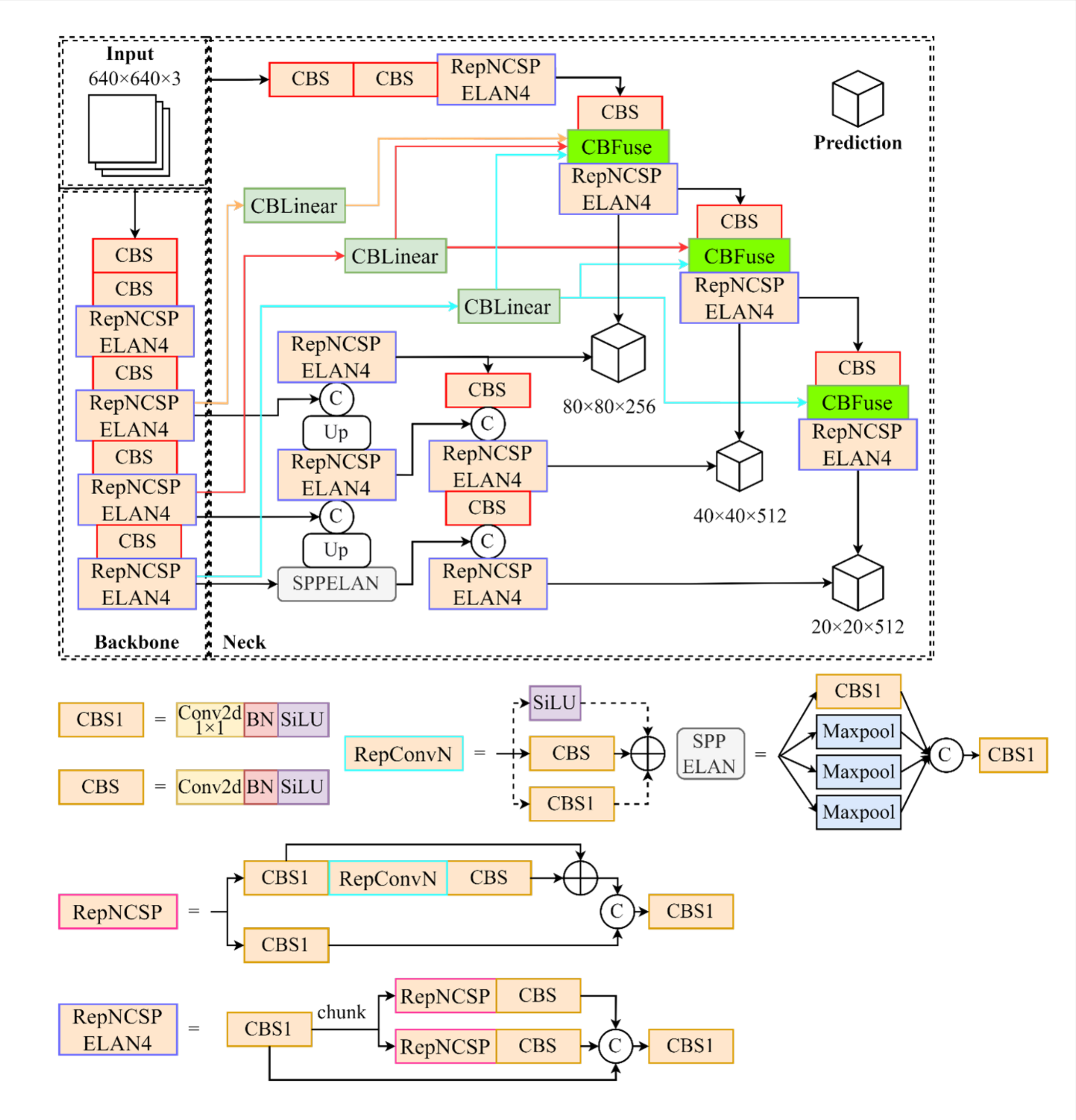
2. Related Works

2.1. YOLOv9 Object Detection Network

YOLOv9, released on February 21, 2024, by Chien-Yao Wang, the author of YOLOv4 and YOLOv7, represents the latest advancement in detection networks. YOLOv9 addresses the issue of information loss during data transmission within deep networks by introducing programmable gradient information. This innovative concept has led to the development of a Generalized Efficient Layer Aggregation Network (GELAN) architecture for feature extraction. YOLOv9 preserves and extracts critical information for mapping data to targets with remarkable efficiency, achieving detection performance that equals or surpasses previous YOLO models while utilizing fewer parameters and reducing computational demands. Figure 1 provides a schematic of the YOLOv9 network structure.

At the input stage, YOLOv9 maintains consistency with YOLOv7 [13], continuing to utilize Mosaic data augmentation, adaptive anchor box calculations, and adaptive image scaling to enhance data quality for the input model. Modifications to the backbone network include the integration of CSPNet and ELAN features to design the GELAN as the primary feature extraction unit. The use of RepConv as the foundational convolution module, coupled with reparameteriza-

Figure 1
YOLOv9 Network Structure


Input
640×640×3
Backbone
Neck
Prediction
CBS
RepNCSP ELAN4
CBLinear
CBFuse
C
Up
SPPELAN
80×80×256
40×40×512
20×20×512
CBS1 = Conv2d 1×1 BN SiLU
CBS = Conv2d BN SiLU
RepConvN = SiLU CBS CBS1
SPP ELAN = CBS1 Maxpool Maxpool Maxpool C CBS1
RepNCSP = CBS1 RepConvN CBS CBS1 C CBS1
RepNCSP ELAN4 = CBS1 chunk RepNCSP CBS RepNCSP CBS C CBS1

tion techniques, further bolsters feature extraction capabilities. The neck of the network retains the FPN+PAN structure for path aggregation but replaces the E-ELAN module with a GELAN layer. An auxiliary reversible branch is added, incorporating elements from the CBNet composite backbone network to facilitate gradient information flow and aggregation. The prediction end continues to use three different types of prediction boxes for classification, location, and confidence of targets, employing non-maximum suppression to eliminate redundant boxes. The inclusion of the auxiliary reversible branch enhances the information fusion process.

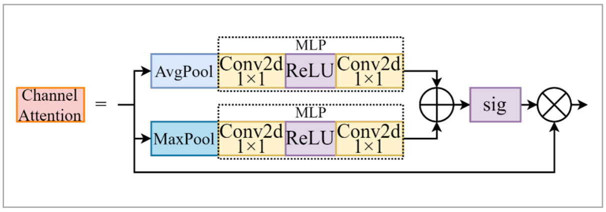
YOLOv9 is designed for target detection tasks in various scenarios. To tailor the network for transfer deployment, modifications have been made for lightweight operation and specific enhancements for bounce lock detection.

2.2. Channel Attention Mechanism

As illustrated in Figure 2, the Convolutional Block Attention Module (CBAM) attention mechanism [17] dynamically adjusts the weights of information, enhancing the neural network's focus on pertinent details while minimizing attention to irrelevant data. This feature significantly improves the detector's ability to allocate attention effectively across different targets, enhancing the perception of useful information.

Figure 2

Channel Attention Structure

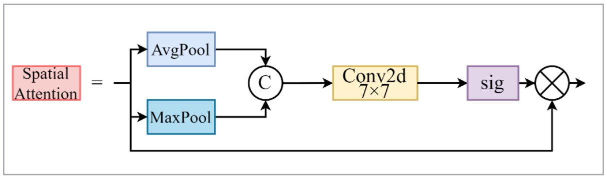


The Channel Attention (CA) module, building on the principles of SENet, introduces a maxpool feature extraction method. This enhancement capitalizes on the inter-channel relationships to generate a channel attention map, focusing the neural network's processing power on the most informative features.

The Spatial Attention (SA) module, as depicted in Figure 3, exploits the spatial relationships between features. It performs parallel average pooling and max pooling operations along the channel axis. The outputs from these pooling operations are concatenated and then processed through a convolutional layer to produce the attention map, as detailed in Figure 3. The computational processes for both the channel attention map and the spatial attention map are as follows:

Figure 3

Spatial Attention Structure



$$M_c(\boldsymbol{F}) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{F})) + \mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{F}))) \quad (1)$$

$$M_s(\boldsymbol{F}) = \sigma(k^{7\times7}([\mathrm{AvgPool}(\boldsymbol{F}); \mathrm{MaxPool}(\boldsymbol{F})])) \quad (2)$$

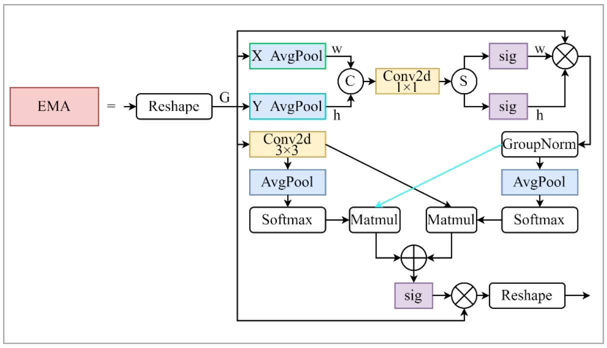
2.3. Efficient Multi-Scale Attention (EMA)

In computer vision networks, attention mechanisms significantly enhance the salience of feature representations. However, modeling inter-channel relationships through channel dimension reduction can sometimes adversely affect feature extraction. To address this, Ouyang et al. [10] proposed the Efficient Multi-Scale Attention (EMA) module, which minimizes computational costs while preserving essential information.

As depicted in Figure 4, the EMA mechanism begins by dividing the input feature maps across the channel dimension into G groups. Each group represents a set of sub-features with distinct semantics. The mechanism employs dual-path 1x1 branches and a single 3x3 branch to derive various attention weights for these feature layers. The outputs from the 1x1 branches are processed through a sigmoid function, facilitating the fusion of two channel attention components. The attention weights are then refined using two-dimensional global average pooling to encode global spatial information and align dimensions. This process culminates in the aggregation of cross-spatial informa-

Figure 4

Efficient Multi-Scale Attention



tion through matrix dot product operations, and the final weighted spatial attentions are mapped through a sigmoid function.

The detailed calculation steps for the EMA attention map are as follows:

$$M_{1h,1w}(\boldsymbol{F}) = \sigma_{h,w}\left(k^{1\times1}\left(\left[\mathrm{AvgPool}_h(\boldsymbol{F}_R);\mathrm{AvgPool}_w(\boldsymbol{F}_R)\right]\right)\right) \quad (3)$$

$$M_1(\boldsymbol{F}) = \mathrm{GN}\left(M_{1h}(\boldsymbol{F}) \times M_{1w}(\boldsymbol{F}) \times \boldsymbol{F}_R\right) \quad (4)$$

$$M_3(\boldsymbol{F}) = k^{3\times3}(\boldsymbol{F}_R) \quad (5)$$

$$\begin{aligned} M_{\mathrm{EMA}}(\boldsymbol{F}) = \sigma\Big(&\mathrm{S}\big(\mathrm{avgpool}\big(M_1(\boldsymbol{F})\big)\big)\times M_3(\boldsymbol{F}) + \\ &\mathrm{S}\big(\mathrm{avgpool}\big(M_3(\boldsymbol{F})\big)\big)\times M_1(\boldsymbol{F})\Big) \end{aligned} \quad (6)$$

3. Proposed Methods

3.1. Improved Efficient Multi-Scale Attention (IEMA)

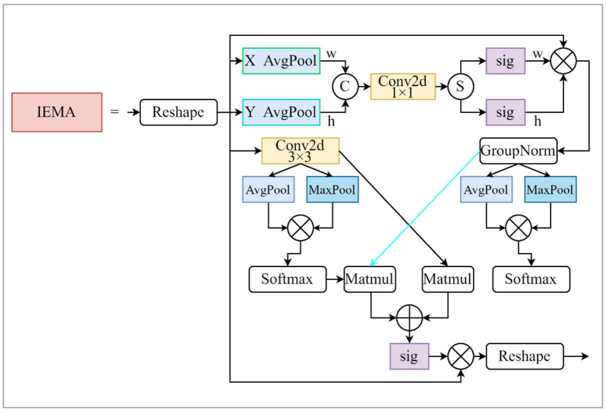
The original Efficient Multi-scale Attention (EMA) mechanism uses global average pooling to effectively encapsulate global spatial information [10]. This method calculates the averages for local areas, thus providing a uniform feature representation across the entire spatial domain. However, this approach might not adequately reflect the prominence of the most significant features within these areas.

To address this limitation and enhance feature representation, the Improved Efficient Multi-scale Attention (IEMA) incorporates maximum pooling during the global spatial information encoding process. By selecting the maximum value from each local area through maximum pooling, it preserves the most salient features. The inspiration for integrating max pooling into IEMA comes from the channel attention mechanism in the Convolutional Block Attention Module (CBAM), where channel attention improves the feature maps generated by convolutional layers by focusing more on informative features, thus enhancing the model's overall discriminative capability.

Adopting a similar strategy, IEMA aims to enhance the capabilities of the original EMA by providing a more detailed and focused spatial feature analysis. The combination of global average pooling and maximum pooling allows for a more comprehensive and detailed understanding of the spatial distribution in the data, which can lead to better performance in tasks requiring high-level spatial awareness and feature specificity. This hybrid pooling method, supported by channel attention principles, marks a significant evolution in the design of neural network attention mechanisms, aimed at improving accuracy and efficiency in processing multi-scale spatial data.

Figure 5

Improved Efficient Multi-Scale Attention



The computational process for IEMA is outlined as follows in Figure 5:

$$M_{\mathrm{IEMA}}(\boldsymbol{F}) = \sigma\left(\begin{array}{l} \mathrm{S}\big(\mathrm{avgpool}\big(M_1(\boldsymbol{F})\big)\times\mathrm{maxpool}\big(M_1(\boldsymbol{F})\big)\big)\times M_3(\boldsymbol{F}) \\ +\mathrm{S}\big(\mathrm{avgpool}\big(M_3(\boldsymbol{F})\big)\times\mathrm{maxpool}\big(M_3(\boldsymbol{F})\big)\big)\times M_1(\boldsymbol{F}) \end{array}\right) \quad (7)$$

3.2. Attention-Guided Feature Extraction and Fusion

The computational process of attention mechanisms indicates that spatial attention prioritizes the significance of each locality within the feature layer, whereas channel attention focuses on the importance of each feature layer itself. In object detection tasks, the primary objective is to identify key targets within an image, where the relevance of different areas varies. Spatial attention thus guides the network to extract features from critical regions. This paper incorporates the IEMA attention mechanism into the YOLOv9 feature extraction network, guiding it to enhance the extraction and flow of key target features after shallow feature extraction, thereby increasing sensitivity to important targets.

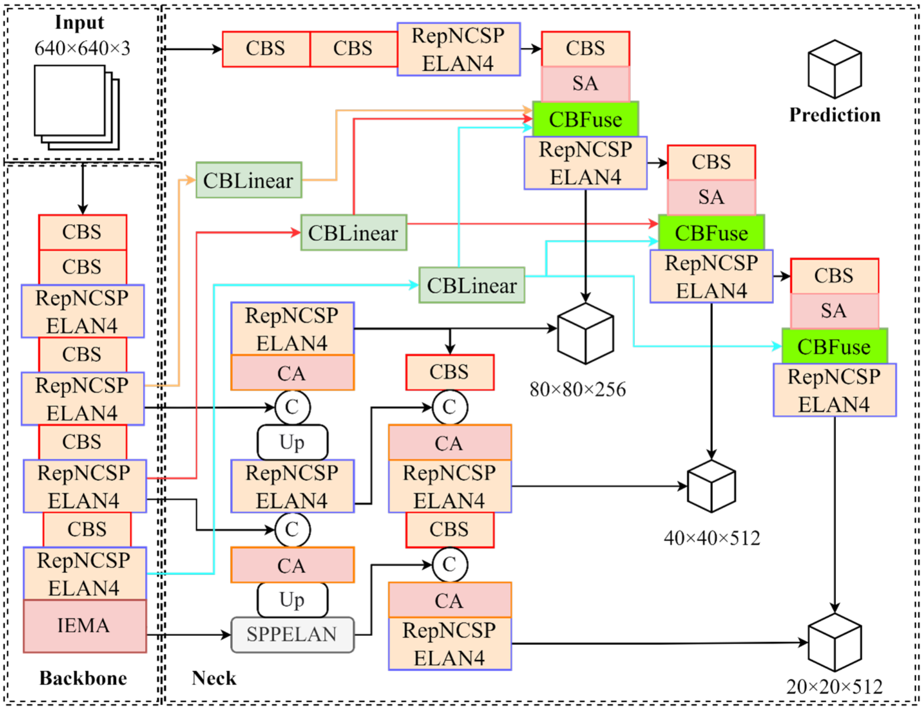
Meanwhile, the backbone network's extracted feature layers contain rich semantic information, which varies across different layers in their contribution to target detection. Channel attention mechanisms are employed to guide feature fusion. To enhance network portability, the neck component has been modified to be more lightweight, focusing more effectively on feature fusion.

3.3. AL-YOLOv9s Network Model

The standard version of the YOLOv9 model is large, with a file size of 122.4 MB, which is not ideal for portability or real-time object detection. By adjusting the scaling ratios of the YOLO series and setting the network depth to 0.33 and width to 0.25, the YOLOv9s model is created with a significantly reduced size of 9.1 MB. This reduction greatly enhances detection speed and efficiency.

In the YOLOv9s network, the integration of improved efficient multi-scale attention, spatial attention, and channel attention, along with lightweight modifications to the neck component, results in the AL-YOLOv9s network proposed in this paper. The network structure is illustrated in Figure 6.

Figure 6
AL-YOLOv9s Network Model



4. Experimental Results

4.1. Experimental Environment and Data

The experimental dataset was sourced from video footage recorded in substation environments, from which 7,756 images with a resolution of 1920x1080 dpi were extracted. These images feature four types of toggle locks, encompassing a total of eight distinct categories. To ensure the diversity of the dataset, we selected video clips under various conditions to capture different environmental changes. Each category has enough samples in the dataset to avoid class imbalance issues. We performed standard processing on each image, including cropping, scaling, and data augmentation techniques such as random rotation, translation, flipping, and color adjustment, to increase data diversity and prevent model overfitting. Before dividing the dataset, each image was manually annotated and then randomly split into training, validation, and test sets in an 8:1:1 ratio to ensure the effectiveness of model training and evaluation. The training set is used for model training, the validation set for tuning hyperparameters and model selection, and the test set for evaluating the final performance of the model.

4.2. Evaluation Metrics

In object detection, model performance is commonly assessed using Average Precision (AP) and the Mean Average Precision (mAP). AP is defined as the area under the curve that is formed by plotting Recall against Precision. The formula for calculation is expressed as follows:

$$Pr = \frac{TP}{TP + FP} \qquad (8)$$

$$Re = \frac{TP}{TP + FN} \qquad (9)$$

$$AP = \frac{1}{101}\sum_{i=0}^{100} Pr(Re = \frac{i}{100}) \qquad (10)$$

$$mAP = \sum_{i=1}^{N} \frac{AP_i}{N} \qquad (11)$$

$$F1 - Score = 2 \bullet \frac{Pr \bullet Re}{Pr + Re} \qquad (12)$$

In this formula, Pr denotes precision, Re represents recall, TP stands for true positives, FP for false positives, FN for false negatives, and N is the number of categories detected. The F1-Score, also known as the balanced F score, is defined as the harmonic mean of precision and recall.

4.3. Experimental Setup

The experiments were conducted using an Ubuntu 20.04 operating system, powered by an Intel i9-10920X CPU with 32GB of RAM and an NVIDIA GeForce RTX 3070 graphics card. The programming was executed in Python 3.8 using the Pytorch 1.8.0 deep learning framework. The hyperparameters were set as follows: an initial learning rate of 0.0031, decay factor of 0.12, momentum at 0.937, a batch size of 24, with training spanning 400 epochs.

4.4. Model Comparison

From Table 1, it is evident that the integration of attention mechanisms significantly altered the number of layers in the network. A lightweight modification to the neck component notably reduced the number of network parameters, model size, and floating-point operations. The weight size of the AL-YOLOv9s model was reduced to just 4.7MB, enhancing its portability significantly. The AL-YOLOv9s model incorporates

Table 1

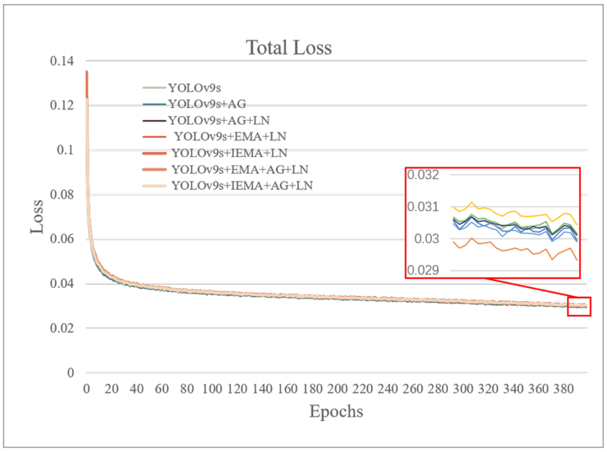
Comparison for Algorithmic Model Complexity

Baseline	EMA	IEMA	AG	Lightweight Neck(LN)	Layers	Parameter Quantity	Mode Size/MB	FLOPs/G
YOLOv9	×	×	×	×	580	60512080	122.4	264.0
YOLOv9s	×	×	×	×	580	4218832	9.1	18.2
	×	×	√	×	620	4277396	9.3	18.3
	×	×	√	√	620	2024316	4.7	8.3
	√	×	×	√	588	1980664	4.6	8.2
	×	√	×	√	589	1980664	4.6	8.2
	√	×	√	√	628	2026940	4.7	8.3
	×	√	√	√	629	2026940	4.7	8.3

lightweight neck components and attention mechanisms, aimed at enhancing efficiency and accuracy without significantly increasing computational complexity. By utilizing attention mechanisms, the model can focus on the most relevant parts of the image, improving feature representation and reducing false positives and negatives. This leads to better precision, recall, and overall detection performance, especially in complex and high-resolution images.

Figure 7

Training Loss Comparison



During the training, loss function curves for the YOLOv9s and its improved versions were monitored. As illustrated in Figure 7, the models initially showed a rapid decrease in loss, with minor fluctuations thereafter. By the 40th epoch, the average loss values stabilized at approximately 0.04, eventually converging to around 0.03, indicating effective convergence across all models.

The training results revealed that the highest accuracy achieved by AL-YOLOv9s was 0.99636, with a peak recall rate of 0.99886. The maximum mean Average Precision (mAP) at a confidence threshold of 0.5 was 0.99500. As shown in Table 2 and Figure 8, AL-YOLOv9s demonstrated superior accuracy and performed best across various metrics compared to other models of similar scale.

As depicted in Figure 9, the AL-YOLOv9s model maintained comparable detection capabilities to the YOLOv9s. The comparison of the network parameters is illustrated in the table below.

As demonstrated in Table 3, the AL-YOLOv9s model shows a marked improvement in accuracy over the YOLOv9s model, despite an increase in network layers and a decrease in model weight, contributing to a faster detection performance. After reparameterization, the model weight was further reduced to just 4.7MB,

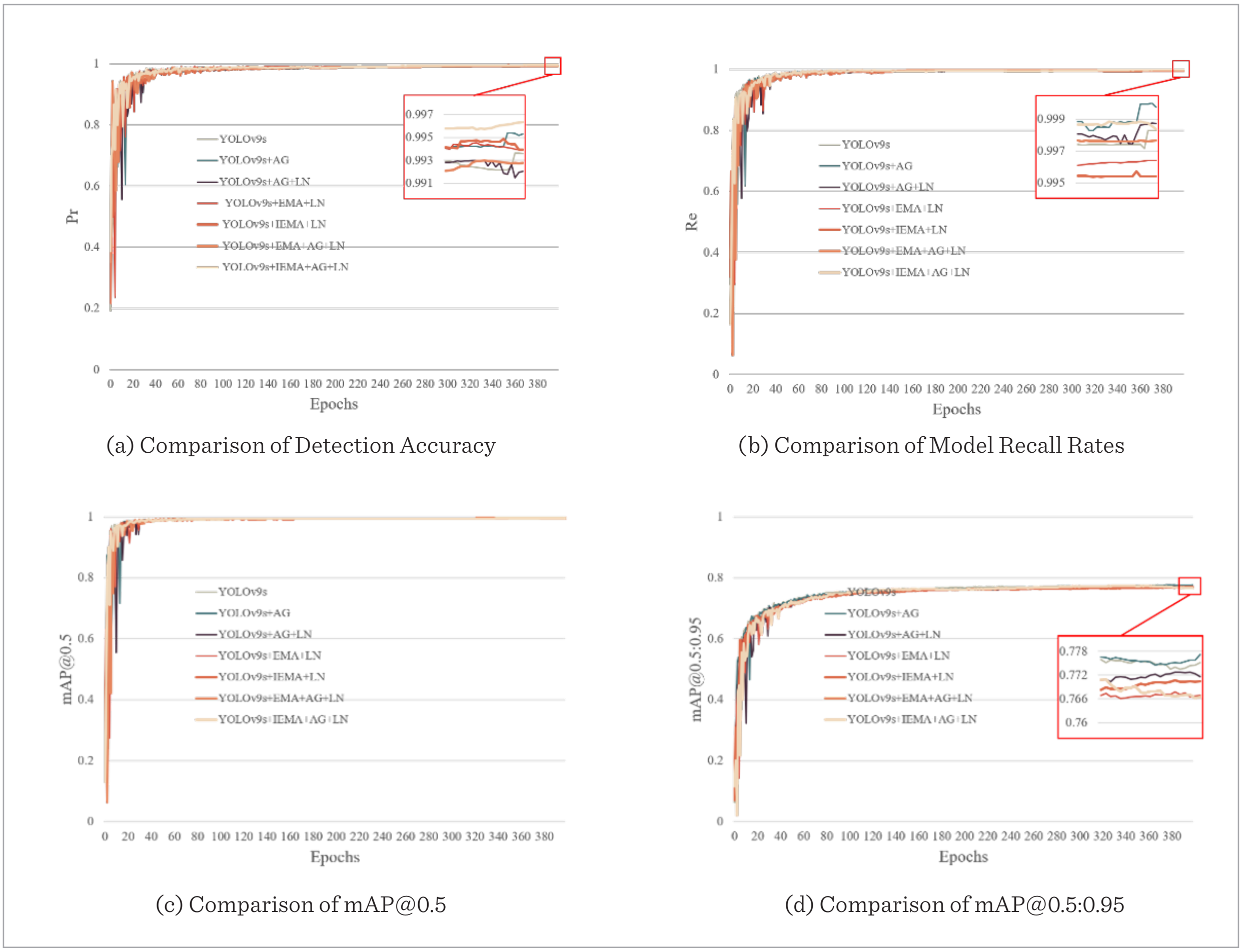
Table 2

Comparison of Model Evaluation Metrics

Baseline	EMA	IEMA	AG	LightweightNeck	Precision (%)	Recall (%)	F1-Score	mAP@0.5(%)	mAP@0.5:0.95(%)
YOLOv9	×	×	×	×	-	-		-	-
YOLOv9s	×	×	×	×	99.363	99.830	99.596	99.487	77.626
	×	×	√	×	99.539	99.993	99.766	99.500	77.728
	×	×	√	√	99.299	99.874	99.586	99.443	77.270
	√	×	×	√	99.454	99.640	99.547	99.500	77.073
	×	√	×	√	99.498	99.686	99.592	99.431	77.066
	√	×	√	√	99.297	99.884	99.590	99.441	77.412
	×	√	√	√	99.636	99.886	99.761	99.500	77.412

Figure 8

Training Metrics Comparison Curve



(a) Comparison of Detection Accuracy

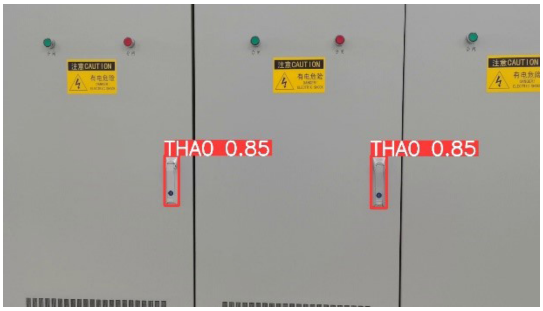
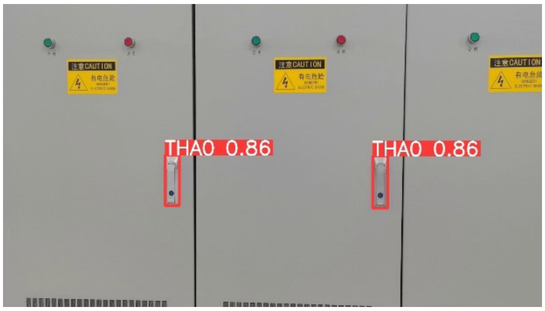
(b) Comparison of Model Recall Rates

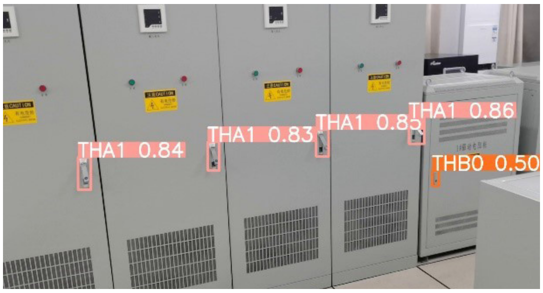
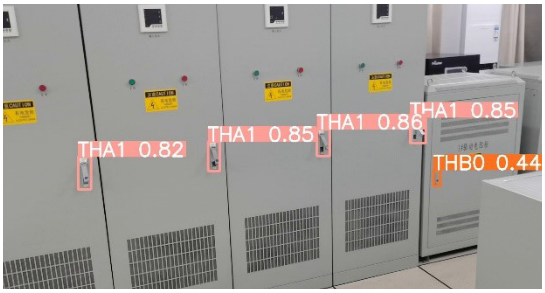
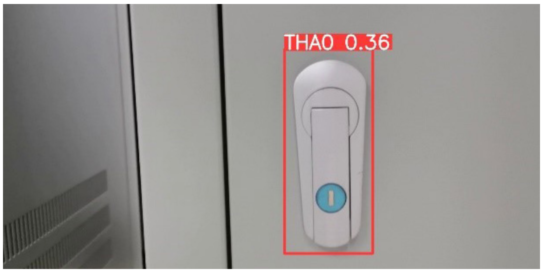
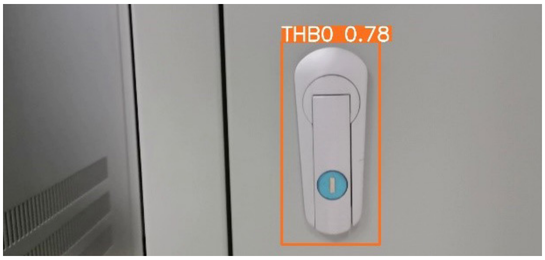
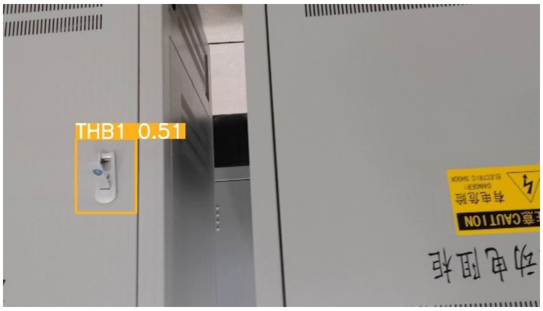
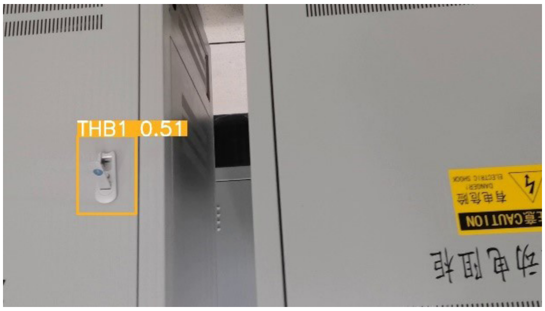
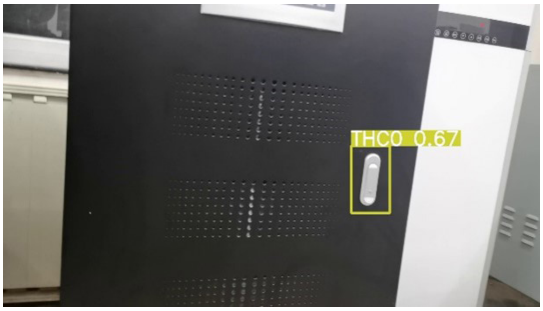
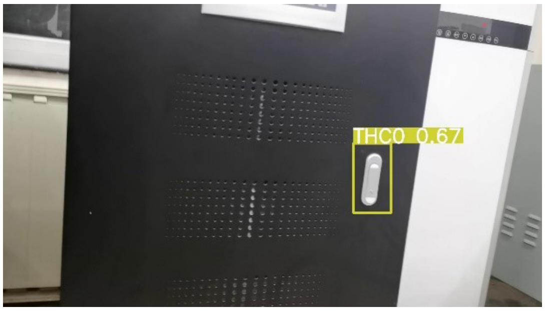
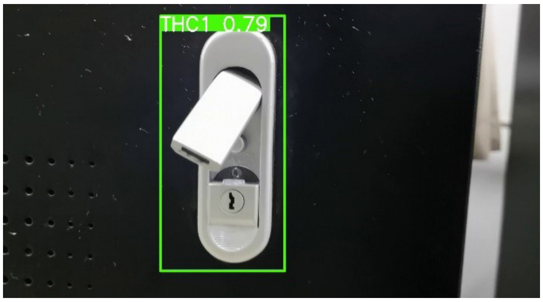
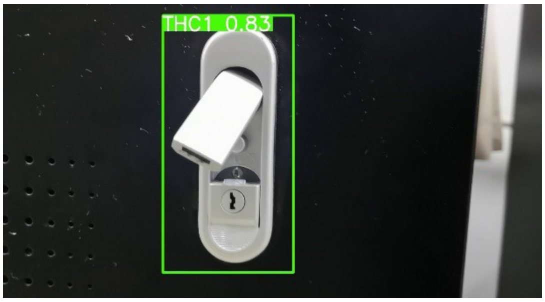
(c) Comparison of mAP@0.5

(d) Comparison of mAP@0.5:0.95

Figure 9

Detection Results for Toggle Locks using YOLOv9s and AL-YOLOv9s

Class	YOLOv9s	AL-YOLOv9s
THA0		

Class	YOLOv9s	AL-YOLOv9s
THA1	 THA1 0.84 THA1 0.83 THA1 0.85 THA1 0.86 THB0 0.50	 THA1 0.82 THA1 0.85 THA1 0.86 THA1 0.85 THB0 0.44
THB0	 	
THB1	 	
THC0	 	
THC1	 	

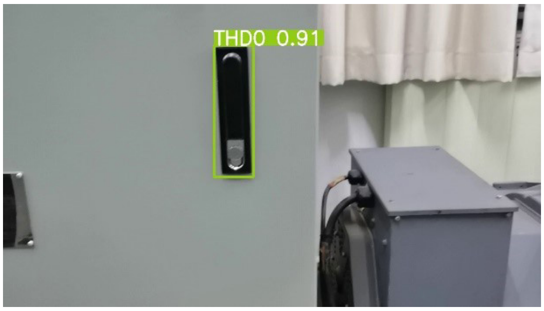
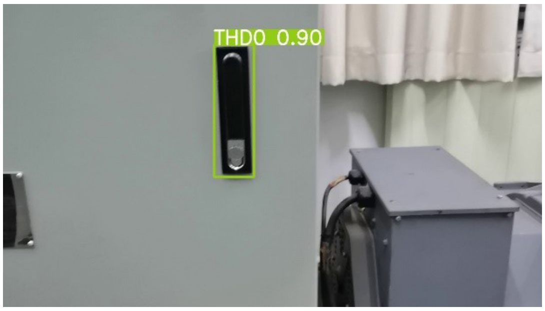
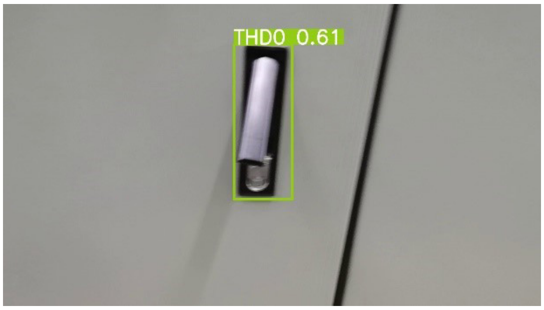
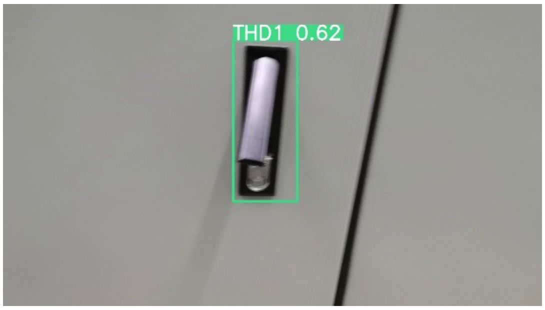
Class	YOLOv9s	AL-YOLOv9s
THD0		
THD1		

Table 3
Comparative Evaluation Metrics for Network Models

Baseline	EMA	IEMA	AG	Lightweight Neck	Misi dentification	Accuracy (%)	FPS	Inference Time (ms)
YOLOv9	×	×	×	×	-	-	-	-
YOLOv9s	×	×	×	×	17	97.81	15.7	49306
	×	×	√	×	15	98.06	14.2	54271
	×	×	√	√	31	96.00	16.6	46592
	√	×	×	√	27	96.52	18.3	42339
	×	√	×	√	21	97.29	18.1	42872
	√	×	√	√	23	97.03	16.6	46666
	×	√	√	√	12	98.45	16.5	46877

enhancing the model's adaptability for deployment on various platforms and meeting the requirements for real-time detection in high-definition images.

5. Conclusions and Future Work

This study addresses the challenge of detecting pop-up locks in switchgear cabinets by building upon the latest YOLOv9s model to develop an enhanced version, named AL-YOLOv9s. This advanced detection model integrates an efficient multi-scale attention mechanism that significantly improves feature extraction capabilities. It incorporates both channel and spatial attention modules to refine feature fusion, along with a lightweight modification to the feature fusion network. These enhancements have achieved a notable increase in detection accuracy while maintaining the model size at just 4.7MB. The experimental results confirm that the AL-YOLOv9s model successfully meets the requirements for both detection accuracy and portability, effectively addressing the task of pop-up lock detection.

Future work will focus on improving the execution efficiency and real-time responsiveness of the AL-YOLOv9s model, and studying improved attention mechanisms to enhance the adaptability and recognition accuracy of the model in various real-world environments, thereby providing more powerful and reliable detection capabilities.

Data Sharing Agreement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, author-ship, and publication of this article.

Funding

The authors received no financial support for the research.

Acknowledgment

Not applicable.

References

1. Amudhan, A., Sudheer, A. J. I., Computing, V. Lightweight and Computationally Faster Hypermetropic Convolutional Neural Network for Small Size Object Detection. Image and Vision Computing, 2022, 119, 104396. https://doi.org/10.1016/j.imavis.2022.104396
2. Attaran, M. The Impact of 5G on the Evolution of Intelligent Automation and Industry Digitization. Journal of Ambient Intelligence and Humanized Computing, 2023, 14, 5977-5993. https://doi.org/10.1007/s12652-020-02521-x
3. Chen, Y.-C., Chen, R.-S., Sun, H.-M., Wu, S. F. Using RFID Technology to Develop an Intelligent Equipment Lock Management System. International Journal of Computer Science and Engineering, 2019, 20(2), 157-165. https://doi.org/10.1504/IJCSE.2019.103810
4. Fang, S., Moreno Brenes, A., Brusoni, S. Technology Intelligence and Digitalization in the Manufacturing Industry. Research-Technology Management, 2023, 66(5), 22-33. https://doi.org/10.1080/08956308.2023.2234758
5. Hao, W., Ren, C., Han, M., Zhang, L., Li, F., Liu, Z. Cattle Body Detection Based on YOLOv5-EMA for Precision Livestock Farming. Animals, 2023, 13(22), 3535. https://doi.org/10.3390/ani13223535
6. Li, C., Naimeng, C., Hao, J., Shuang, W. Improved Faster R-CNN for Dense Small Objects. Paper presented at the 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), 2022. https://doi.org/10.1109/ICFTIC57696.2022.10075150
7. Maity, M., Banerjee, S., Chaudhuri, S. S. Faster R-CNN and YOLO Based Vehicle Detection: A Survey. Paper Presented at the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021. https://doi.org/10.1109/ICCMC51019.2021.9418274
8. Motta, L. L., Ferreira, L. C., Cabral, T. W., Lemes, D. A., Cardoso, G. d. S., Borchardt, A., Neto, F. B. General Overview and Proof of Concept of a Smart Home Energy Management System Architecture. Electronics, 2023, 12(21), 4453. https://doi.org/10.3390/electronics12214453
9. Muñiz, R., Nuño, F., Díaz, J., González, M., Prieto, M. J., Menéndez, Ó. Real-Time Monitoring Solution with Vibration Analysis for Industry 4.0 Ventilation Systems. The Journal of Supercomputing, 2023, 79(6), 6203-6227. https://doi.org/10.1007/s11227-022-04897-3
10. Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., Huang, Z. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. Paper Presented at the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023. https://doi.org/10.1109/ICASSP49357.2023.10096516
11. Pan, W., Chen, J., Lv, B., Peng, L. Optimization and Application of Improved YOLOv9s-UI for Underwater Object Detection. Applied Sciences, 2024, 14, 7162. https://doi.org/10.3390/app14167162
12. Redmon, J., Farhadi, A. Yolov3: An Incremental Improvement. International Joint Conference on Neural Networks (IJCNN), 2018. https://doi.org/10.1109/IJCNN48605.2020.9206848
13. Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. Paper Presented at the

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023. https://doi.org/10.1109/CVPR52729.2023.00721

14. Wang, Y., Zhang, X., Li, L., Wang, L., Zhou, Z., Zhang, P. An Improved YOLOv7 Model Based on Visual Attention Fusion: Application to the Recognition of Bouncing Locks in Substation Power Cabinets. Applied Sciences, 2023, 13(11), 6817. https://doi.org/10.3390/app13116817

15. Wang, Y., Zhao, S., Liu, Y., Li, J., Zhang, L. CAMixerSR: Only Details Need More "Attention". Image and Video Processing, 2024. https://doi.org/10.1109/CVPR52733.2024.02441

16. Wich, C., Ungsumalee, S. Embedded Histogram of Oriented Gradients for Glaucoma Classification of Fundus Images. Paper Presented at the Proceedings of the 2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC), 2023. https://doi.org/10.1109/ITC-CSCC58803.2023.10212574

17. Woo, S., Park, J., Lee, J.-Y., Kweon, I. S. CBAM: Convolutional Block Attention Module. Paper Presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018. https://doi.org/10.1007/978-3-030-01234-2_1

18. Xiao, D., Meng, Q., Li, S., Yuan, X. Improving Transformers with Dynamically Composable Multi-Head Attention. 41st International Conference on Machine Learning, 2024. DOI: arxiv-2405.08553.

19. Yan, K., Li, Q., Li, H., Wang, H., Fang, Y., Xing, H. Deep Learning-Based Substation Remote Construction Management and AI Automatic Violation Detection System. IET Generation, Transmission & Distribution, 2022, 16(9), 1714-1726. https://doi.org/10.1049/gtd2.12387

20. Yan, X., Jia, L., Cao, H., Yu, Y., Wang, T., Zhang, F., Guan, Q. Multitargets Joint Training Lightweight Model for Object Detection of Substation. IEEE Transactions on Neural Networks and Learning Systems, 2022. https://doi.org/10.1109/TNNLS.2022.3190139

21. Yang, F., Duan, X., Wang, J., Wang, Y., Zhang, H. Research and Application of Target Detection Algorithm for Live Operation in Substation. Paper Presented at the Journal of Physics: Conference Series, 2024. https://doi.org/10.1088/1742-6596/2703/1/012038

22. Zhao, Z., Feng, S., Zhai, Y., Zhao, W., Li, G. Infrared Thermal Image Instance Segmentation Method for Power Substation Equipment Based on Visual Feature Reasoning. IEEE Transactions on Instrumentation and Measurement, 2023. https://doi.org/10.1109/TIM.2023.3322998

23. Zheng, H., Ping, Y., Cui, Y., Li, J. Intelligent Diagnosis Method of Power Equipment Faults Based on Single-Stage Infrared Image Target Detection. IET Electrical Systems in Transportation, 2022, 17(12), 1706-1716. https://doi.org/10.1002/tee.23681

24. Zhu, J., Zhou, D., Lu, R., Liu, X., Wan, D. C2DEM-YOLO: Improved YOLOv8 for Defect Detection of Photovoltaic Cell Modules in Electroluminescence Image. Nondestructive Testing and Evaluation, 2024, 109, 104159. https://doi.org/10.1080/10589759.2024.2319263