AI real-time detection of cable corrosion in power inspection videos based on a YOLOv10-transformer hybrid network

Shiquan Sun[1], Meng Xie[2], Yi Feng[3], Jiatian Shangguan[4], Guangjiu Chen[5]
Beijing Huashang Electric Power Industry Development Company, Beijing, 100032, China
[2]Corresponding author
E-mail: [1]*slb3hpv@163.com*, [2]*majhcfho@163.com*, [3]*injerft@163.com*, [4]*gi1753@163.com*, [5]*cbqg88@163.com*

Received 29 January 2026; accepted 22 August 2026; published online 27 September 2026
DOI https://doi.org/10.21595/jme.2026.26065

Abstract. The operation and maintenance of current power infrastructure face critical challenges, such as inefficient cable corrosion detection, strong subjectivity in assessments, and insufficient early-warning capabilities. Traditional manual inspection methods not only demand substantial human resources but are also susceptible to environmental factors and inspector experience, which compromises detection accuracy. Therefore, this study proposes a deep learning model that integrates a multi-scale feature extraction module with an adaptive attention mechanism (MSFEM-AAM). The model employs parallel feature extraction pathways to simultaneously capture global morphological characteristics and localized corrosion details on cable surfaces. A dynamic weight fusion mechanism is incorporated to adaptively integrate multi-scale features, thereby enhancing the model's discriminative capability for corrosion patterns. Experimental results show that MSFEM-AAM achieves 96.8 % accuracy in corrosion grade classification (Task 2) and 94.7 % mAP in corrosion region detection (Task 1). The single-frame inference time reaches 9 ms with GPU acceleration. These quantitative metrics demonstrate the advanced performance and practical applicability of MSFEM-AAM, contributing to improved automation of corrosion detection and enhanced early-warning capabilities, thus supporting the secure and stable operation of power grids.

Keywords: cable corrosion assessment, corrosion level classification, multi-scale feature fusion, adaptive attention mechanism, robustness evaluation, real-time inference.

1. Introduction

In power system inspection, cable corrosion detection relies mainly on manual visual inspection or conventional image processing techniques. Both approaches have clear limitations: low efficiency, strong subjectivity, and poor real-time performance [1]. With the expansion of power grids and increasing environmental complexity, automated solutions are required - ones that combine high precision with efficiency. Although deep learning has advanced object detection, its application to power equipment still faces hurdles [2]. Existing methods struggle with small targets, complex backgrounds, and gradual corrosion features. Achieving a reliable balance between speed and accuracy remains challenging [3].

A YOLOv5-based framework for power equipment defect detection was introduced in Reference [4], which incorporated a lightweight network architecture to facilitate rapid identification; however, its limitations included insufficient sensitivity to subtle textural changes in corroded areas. C.-K. Lee and Y.-J. Shin [5] developed a cable anomaly detection system utilizing Faster R-CNN, yet its primary drawback lied in high computational complexity, rendering it unsuitable for real-time inspection of video streams. A pixel-level corrosion segmentation model based on Mask R-CNN was constructed in Reference [6], featuring an instance segmentation head for precise contour extraction; nevertheless, it demonstrated suboptimal performance in identifying corroded regions under occlusion conditions. A corrosion classification method integrating ResNet with an attention mechanism was proposed by Y. Liu et al. [7], which incorporated a deep feature extraction module; its limitation was confinement to classification tasks without the capability for corrosion localization. Y. Cheng et al. [8] developed a multi-scale feature fusion network employing Inception modules, designed with parallel convolutional structures to capture features at different scales; however, its shortcomings included an excessively large parameter count, leading to high deployment costs. A semantic segmentation scheme integrating U-Net with conditional random fields was presented in Reference [9]; however, its deficiency involved inter-frame inconsistency when processing video sequenced. Based on the Single Shot MultiBox Detector (SSD), a rapid detection system was constructed in Reference [10], which implemented a multi-scale feature map prediction mechanism. However, its main weakness was relatively low detection accuracy for small-target corrosion areas. A hybrid detection methodology fusing traditional image processing with deep learning was proposed by G. Choudhary and D. Sethi [11], which featured a cascaded structure of morphological operations and convolutional neural networks. However, its limitation pertained to a somewhat rigid fusion strategy, failing to fully leverage the respective advantages of each approach.

Addressing the common shortcomings identified in existing research, this study investigates the critical balance between accuracy and speed for real-time detection of cable corrosion in power inspection videos. A novel YOLOv10-Transformer hybrid network architecture is constructed to achieve both accurate identification and rapid localization of corroded regions. The proposed methodology integrates the efficient detection capability of YOLOv10 with the global modeling strengths of the Transformer. The key components include a multi-scale feature extraction module combined with an adaptive attention mechanism (MSFEM-AAM) fusion scheme and a temporal consistency processing module. Experimental results demonstrate that the proposed method achieves a mean average precision (mAP) of 94.7 % on a proprietary power cable corrosion dataset while maintaining a processing speed of 45 frames per second (FPS). This represents a significant improvement in both precision and efficiency compared to current mainstream methods.

The technical innovations of this work are summarized below:

(1) A hierarchical feature interaction mechanism is proposed within MSFEM-AAM. This mechanism enhances multi-scale corrosion feature extraction via dynamic weighted fusion of deep and shallow features.

(2) A lightweight spatio-temporal attention module (STAM) is designed, which effectively leverages inter-frame correlations in video sequences to improve detection consistency with a minimal increase in computational overhead.

(3) A progressive training strategy is established by employing a phased optimization approach that first strengthens fundamental detection capabilities before refining the corrosion feature learning process, thereby ensuring steady performance enhancement.

Collectively, these innovations form MSFEM-AAM. The model supports two use cases: detection (localizing corrosion regions) and classification (grading severity).

2. Related work

2.1. Analysis of the research status of cable corrosion detection

As an important part of power system operation and maintenance, power cable corrosion detection technology has experienced an evolution from manual inspection to intelligent analysis. Current research focuses on key facilities such as overhead lines, underground cables, and substation connections [12]. An analysis of current power patrol inspection technology is shown in Fig. 1.

Fig. 1 summarizes the principal challenges in cable corrosion detection, highlighting the variability of corrosion levels across time and spatial locations, as well as the dependence of detection accuracy on image quality and environmental conditions.

ISSN PRINT 2335-2124, ISSN ONLINE 2424-4635

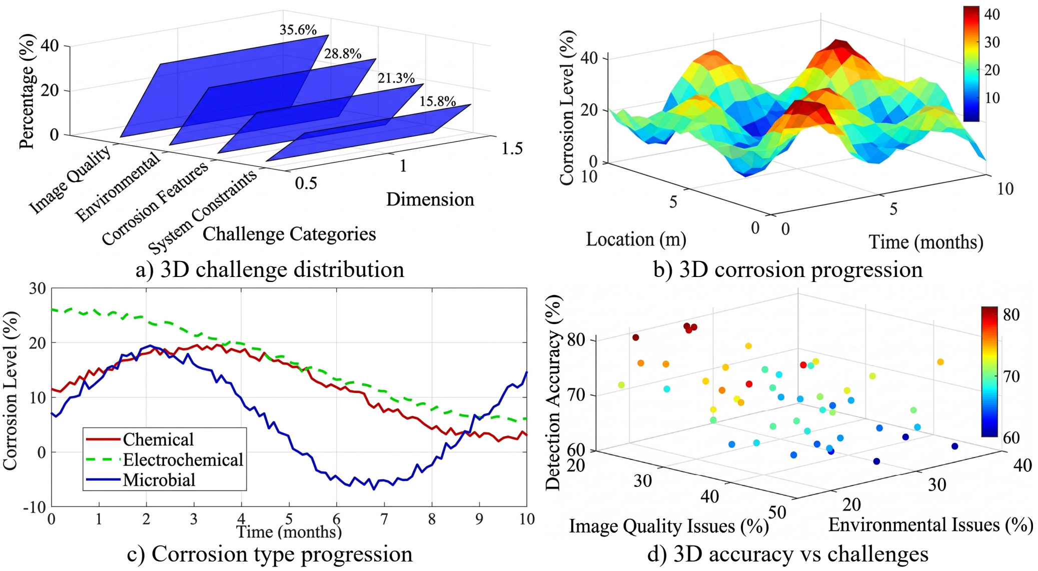


Fig. 1. Research status analysis of cable corrosion detection

Image quality factors account for 35 % of the challenges, encompassing issues such as insufficient resolution, motion-induced blur, and irregular illumination. Environmental interference constitutes 28 %, which involves complex backgrounds, significant occlusions, and adverse weather effects. Corrosion-specific characteristics represent 22 %, including morphological diversity, scale variance, and textural similarity. The remaining 15 % comprises system constraints such as real-time processing requirements and computational resource limitations. Collectively, these issues constitute the major impediments to the practical deployment of corrosion detection technologies [13].

2.2. Technical methods and performance comparison

Based on this analysis, an effective cable corrosion detection approach must balance feature sensitivity with computational efficiency. The predominant methodologies are developed within deep learning frameworks, particularly through optimization of single-stage detectors and semantic segmentation networks. The continuous evolution of CNN-based detection architectures is summarized in Table 1.

Table 1. Comparison of advantages and disadvantages of mainstream detection technologies

Technical methods	Advantage characteristics	Limitations
YOLOv7 [14]	Fast reasoning speed	Limited accuracy of small target detection
YOLOX [15]	Convenient deployment	Limited feature extraction capability
Vision transformer [16]	Strong global modeling ability	High computational complexity
RT-DETR (Real-time detection transformer)	End-to-end detection	Slow convergence

Because of the shortcomings of single-technology applications, it is difficult to solve these issues perfectly through optimization alone. Therefore, greater technology integration is needed for technology complementation, as shown in Table 2.

This study proposes an enhanced YOLOv10-Transformer hybrid network to address the aforementioned technical limitations. The architecture retains the efficient feature extraction and real-time inference capabilities of YOLOv10 while incorporating a Transformer encoder to strengthen global perception of corrosion features.

Table 2. Analysis of complementary characteristics of technology integration

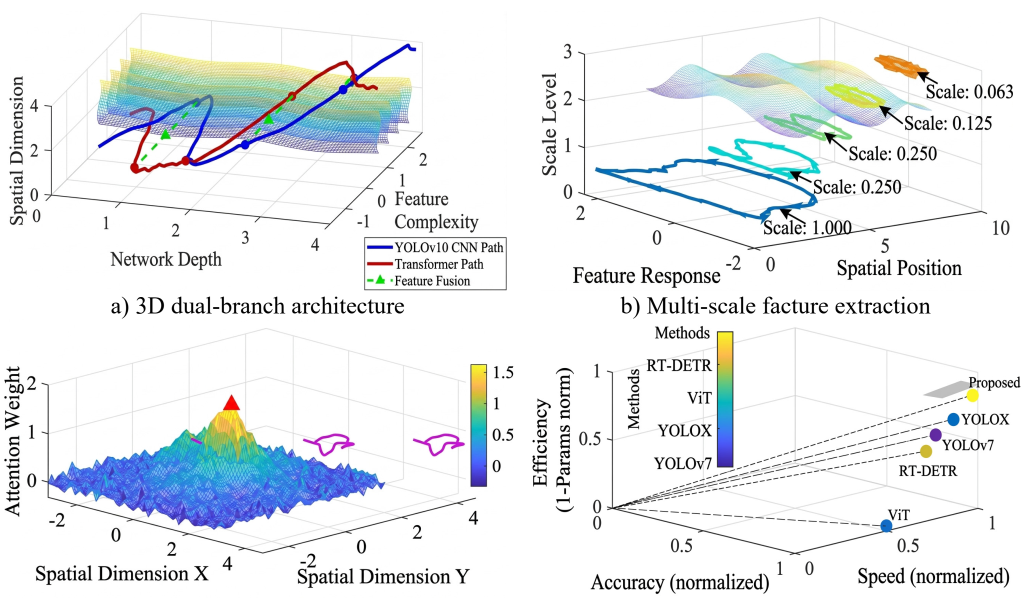
Combination modes	Expected effects	Difficulties in implementation
CNN-Transformer [17]	Local and global feature complementation	The structural design is complex.
Multi-scale feature fusion [18]	Improve the ability to detect objects of different sizes	Feature alignment difficulty
Attention mechanism enhancement [19]	Highlight key areas and suppress background interference	Increased computational overhead

3. Implementation of MSFEM-AAM for cable corrosion detection

The proposed YOLOv10-Transformer hybrid network adopts a dual-branch parallel structure. The backbone network, based on the CNN framework of YOLOv10, is responsible for extracting local features and spatial information. The Transformer branch serves as a complementary pathway, focusing on global dependency modeling. The outputs from both branches are fused through a feature alignment module to form complementary feature representations.

3.1. Hybrid network architecture design

The hybrid network accepts three-channel cable images of size 640×640 as input. The YOLOv10 branch performs five downsampling operations to generate multi-scale feature maps, as illustrated in Fig. 2. Simultaneously, the Transformer branch partitions the input image into sequential patches of dimension 16×16, which are linearly projected into embedding vectors. Spatial information is preserved through positional encoding, while global context is learned via multi-layer encoders. The complete processing pipeline of the detection architecture is visualized in Fig. 2.



a) 3D dual-branch architecture
b) Multi-scale facture extraction
c) Adaptive attention mechanism
d) Performance comparison

Fig. 2. Improved YOLOv10-transformer hybrid network architecture

Fig. 2 shows the parallel structure of the YOLOv10 CNN path and the Transformer path, including the spatial and channel attention weight distribution at feature fusion points and the performance comparison.

The network output contains information on the location, confidence, and classification of the corrosion area. The three detection scales respectively correspond to corrosion areas of different sizes to ensure that both large and small targets can be accurately identified. In the post-processing stage, weighted non-maximum suppression is used to effectively reduce overlapping frames while retaining the real target.

3.2. Multi-scale feature fusion mechanism

Within the YOLOv10 branch of MSFEM-AAM, the backbone network adopts a CSPNet structure to enhance gradient flow. Five downsampling operations - each with a stride of two, except the final layer - generate a multi-scale feature pyramid, preserving both fine-grained spatial details and high-level semantic information. Concurrently, the Transformer branch processes the input image by partitioning it into patches. These patches are linearly projected into 768-dimensional embedding vectors. A learnable positional encoding is then added to retain spatial structure. This sequence is subsequently passed through six stacked Transformer encoder layers. Each layer integrates a multi-head self-attention module (configured with 12 attention heads) and a multi-layer perceptron (with an expansion factor of four, resulting in a 3072-dimensional hidden layer). This configuration enables the modeling of long-range dependencies across the image - for instance, capturing the spatial relationship between a small corroded pit and a larger area of flaking rust.

The fusion of features from these two branches is not a simple concatenation but is instead mediated by an adaptive weighting module. First, the global features from the Transformer branch are upsampled via bilinear interpolation to match the spatial resolution of a corresponding feature map from the YOLOv10 branch. Subsequently, a lightweight subnetwork - composed of sequential channel-wise and spatial attention blocks – generates a set of dynamic fusion coefficients. Formally, the fused feature F_{fused} is computed as:

$$F_{fused} = Conv\left(\sigma\left(Attn_{cs}\left(Concat(FYOLO, F_{Trans})\right)\right) \odot F_{YOLO} + \left(1 - \sigma(\cdot)\right) \odot F_{Trans}\right), \quad (1)$$

where, $Attn_{cs}$ denotes the channel-spatial attention mechanism, σ is the sigmoid activation function, and $\odot$ represents element-wise multiplication. This mechanism ensures that the fusion strategy is context-dependent, thereby maximizing the complementary strengths of both branches.

Channel attention screens useful feature channels. The two types of attention are calculated in parallel, and the results are fused via matrix multiplication:

$$A_{final} = \sigma(W_s \cdot F) \otimes \sigma(W_c \cdot F), \quad (2)$$

where, A_{final} is the final attention map, W_s represents the spatial attention weight, W_c is the channel attention weight, denotes the input feature, σ is the sigmoid activation function, and $\otimes$ denotes element-wise multiplication.

The sigmoid function maps the weight to the [0, 1] interval, indicating the degree of interest. Element-wise multiplication achieves the combination of the two types of attention.

3.3. Bounding box optimization function

The loss function optimizes different aspects of performance separately:

$$L_{bbox} = \lambda_{ciou}L_{ciou} + \lambda_{angle}L_{angle} + \lambda_{shape}L_{shape}, \quad (3)$$

where, L_{bbox} is the total bounding box loss, L_{ciou} is the complete IoU loss, L_{angle} represents the directional alignment loss, L_{shape} represents the shape consistency loss, and λ is the weight coefficient of each component.

This loss function is particularly suitable for cable corrosion detection tasks. The corrosion region often presents an irregular shape, and traditional loss functions struggle to describe it accurately. Directional alignment ensures correct positioning of sloping corrosion areas, and shape consistency avoids distortion of proportions.

Pseudo-code:

Algorithm 1: YOLOv10-Transformer Hybrid Network Inference.

Input: Cable image I, Model parameters θ.

Output: Detection results D.

1: $F_{cnn} \leftarrow CNN_Backbone(I)$

2: $F_{trans} \leftarrow Transformer_Encoder(I)$

3: for $i \leftarrow 1$ to N do

4: $F_{fusion} \leftarrow Feature_Fusion(F_{cnn}, F_{trans})$

5: $A \leftarrow Attention_Module(F_{fusion})$

6: $F_{enhanced} \leftarrow F_{fusion} \otimes A$

7: end for

8: $P_{cls}, P_{box} \leftarrow Detection_Head(F_{enhanced})$

9: $B \leftarrow Decode_Boxes(P_{box})$

10: $D \leftarrow NMS(B, P_{cls})$

11: return D

12: End

The stochastic gradient descent (SGD) optimizer with momentum is employed for model training. An initial learning rate of 0.01 is adopted, coupled with a cosine annealing scheduling strategy to dynamically adjust the learning rate during training. A mini-batch size of 16 is configured to ensure both training stability and computational efficiency. To mitigate overfitting, weight decay regularization is incorporated, and dropout layers are integrated to enhance model robustness.

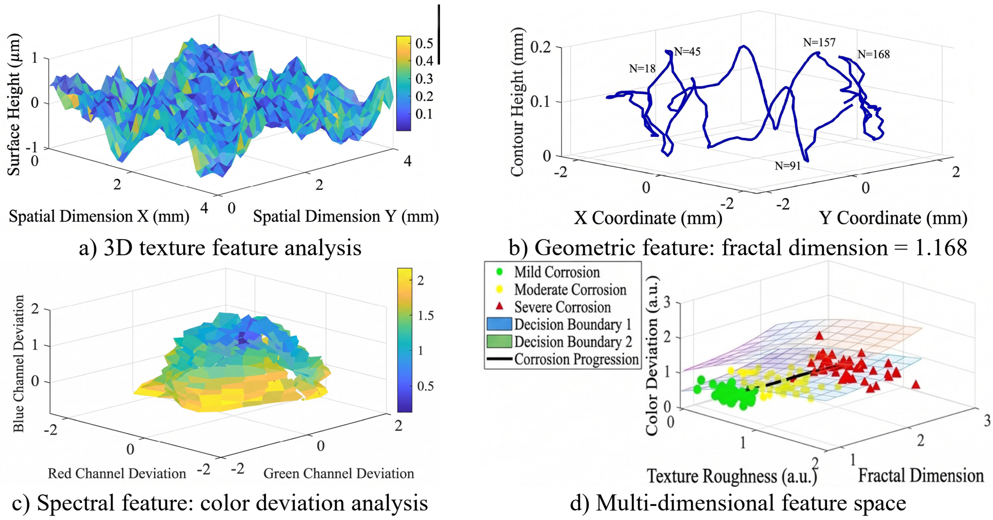
In practical deployment, complexity scores are computed in real time by a lightweight neural network. This introduces negligible additional computational overhead while yielding significant performance improvements. Such a design reflects thorough consideration of real-world application requirements.

4. Evaluation model for cable corrosion state

The corrosion characterization framework establishes quantitative descriptors across three dimensions: textural, geometric, and spectral features. Textural attributes capture surface degradation patterns, geometric properties describe morphological alterations, and spectral signatures reflect compositional changes in materials. Task 1 – Corrosion Region Detection: The model locates corrosion areas in an image and classifies their severity. This is an object detection task. Its primary metric is mean average precision (mAP). Results are reported in Section 5.2.3. Task 2 – Corrosion Grade Classification: The model receives a cropped region (or an image containing a single corrosion area) and outputs a severity level: mild, moderate, or severe. This is a multi-class classification task. Its metrics are accuracy and macro F1-score. Results appear in Section 5.2.1. The two tasks share features extracted by the same backbone. Their outputs differ: Task 1 produces bounding boxes plus class labels; Task 2 produces only class probabilities.

Collectively, these multi-dimensional features provide a comprehensive representation of corrosion states, as schematically illustrated in Fig. 3.

Fig. 3 presents the analysis of textural characteristics, demonstrating surface roughness and local binary pattern (LBP) features. Fractal dimension computation and corrosion profile complexity are evaluated, while spectral feature analysis reveals color deviation and variations in material composition. The distribution within the multidimensional feature space and the corresponding decision boundaries are also illustrated.



a) 3D texture feature analysis b) Geometric feature: fractal dimension = 1.168

c) Spectral feature: color deviation analysis d) Multi-dimensional feature space

Fig. 3. Corrosion feature quantification system: multi-dimensional analysis

The preprocessing phase involves noise reduction and contrast enhancement, followed by postprocessing procedures that include feature fusion and dimensionality reduction:

$$F_{texture} = \frac{1}{N}\sum_{i=1}^{N} |\nabla I(x_i, y_i)|, \tag{4}$$

where, $F_{texture}$ is the texture roughness, N is the total number of pixels, and $\nabla I(x_i, y_i)$ is the gradient magnitude at position (x_i, y_i).

This indicator is particularly sensitive to subtle texture changes caused by early corrosion:

$$D_f = \frac{\log(N(\epsilon))}{\log(1/\epsilon)}, \tag{5}$$

where, D_f is the fractal dimension, $N(\epsilon)$ represents the contour length at scale ϵ, and ϵ is the measurement scale.

This index effectively quantifies the roughness and self-similarity of the corrosion edge:

$$C_{dev} = \sqrt{\frac{1}{M}\sum\nolimits_{j=1}^{M}(H_j - H_{ref})^2}, \tag{6}$$

where, C_{dev} is the color deviation, M represents the number of sampling points, H_j is the tone value of the j-th point, and H_{ref} represents the reference tone value. The degree of color deviation measures the color difference between the corroded area and the normal area.

The model deployment considers the balance between computational efficiency and accuracy. In the feature extraction stage, the optimization algorithm is implemented, and fuzzy reasoning is accelerated using a lookup table method. These measures ensure that the system can meet real-time requirements.

5. Simulation experiment analysis

5.1. Experimental environment setup

Given that real-world cable corrosion detection data is often influenced by complex factors such as environmental interference and equipment variability, this study employs simulated data for experimental validation. This approach enables precise control over corrosion severity to establish a reliable evaluation benchmark. Validation is performed using the Cable Equipment Defect Visible Light Detection Dataset (CSTR: 16666.11. nbsdc. gsn8dm78): this dataset contains 528 pieces of data collected in a cable tunnel simulation environment using a point tunnel inspection robot's internal camera and annotated accordingly. The dataset includes five types of visual defects: cable burning, skin damage, bracket corrosion, grounding box corrosion, and blurry handwriting on the grounding box [20].

The experimental platform configuration is detailed in Tables 3 through 5.

Table 3. Hardware configuration parameters

Items	Parameters	Effects
Processor	Intel Xeon Gold 6348	Provide high-performance parallel computing capabilities
Memory	256GB DDR4	Support large-scale data processing
Graphics processing unit	NVIDIA RTX A6000	Accelerate image feature extraction operation
Storage system	2TB NVMe SSD	Ensure data read and write speed

Table 4. Software environment configuration

Items	Parameters	Effects
Operating system	Ubuntu 20.04 LTS	Provide a stable operating platform
Computational framework	TensorFlow 2.9	Realize deep learning algorithm
Programming language	Python 3.8	Develop experimental procedures and scripts
Programming language	NumPy 1.21	Support scientific computing function

Table 5. Training parameter settings

Items	Parameters	Functions
Learning rate	0.001	Control model parameter update step
Batch size	32	Balance memory usage and training stability
Number of iterations	1000	Ensure that the model is sufficiently converged
Optimizer	Adam	Adaptively adjust learning process
Loss function	Mean squared error (MSE)	Quantify the difference between the predicted and true values
Note: The parameters in this table (lr = 0.001, Adam, MSE) are used only for training baseline models in the comparative experiments (Section 5.2). These settings follow references [21]-[29] to ensure fair benchmarking. They differ from the main model training parameters described in Section 3.3 (lr = 0.01, SGD, cosine annealing)		

A five-fold cross-validation strategy is employed to rigorously evaluate model performance. The dataset is randomly partitioned into five mutually exclusive subsets. In each iteration, four subsets are utilized for training, while the remaining single subset serves as the test set. This procedure is repeated five times to ensure that each subset is used exactly once for testing. By averaging the results obtained from all five test folds, the final performance metrics are computed to enhance the robustness of the evaluation.

To comprehensively assess the model's performance from multiple perspectives, the following metrics are adopted: MAE, R^2, and average inference time.

5.2. Analysis of simulation experiment

The experimental setup uses the parameters listed in Table 5 (lr = 0.001, Adam, MSE) to retrain the baseline models compared in Sections 5.2.1 to 5.2.3. This ensures consistency with the original implementation. However, the proposed MSFEM-AAM is trained with its own parameters (lr = 0.01, SGD, cosine annealing). Therefore, the goal is to compare the model architectures under the optimal settings of each model.

5.2.1. Corrosion grade classification accuracy comparison test

A comparative analysis against several contemporary state-of-the-art methods is performed to clearly delineate the quantitative performance positioning of our model in terms of classification capability. The comparative methods include Vision Transformer (ViT) [21], Convolutional Block Attention Module (CBAM) [22], and EfficientNet [23].

The calculation formula is as follows:

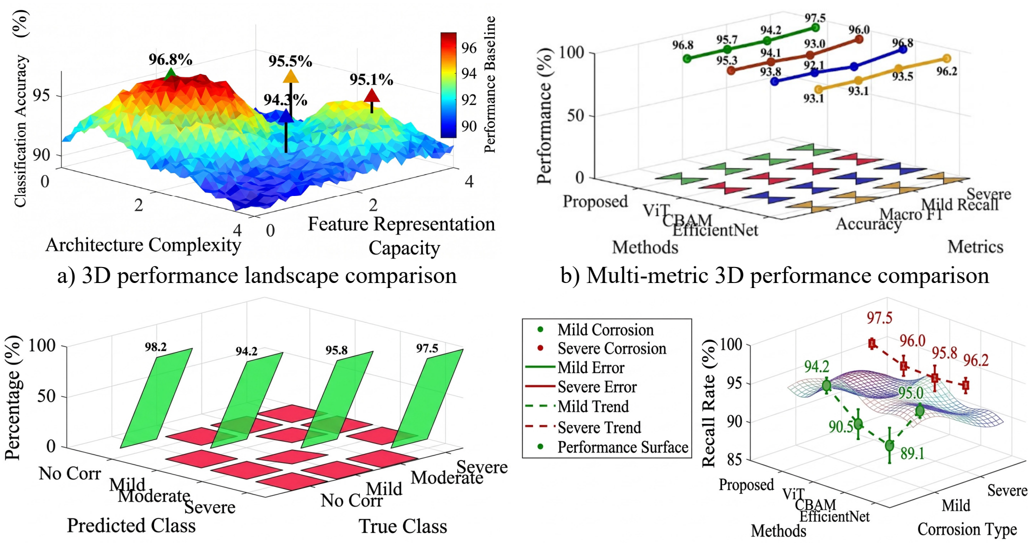
$$Accuracy = \frac{N_{correct}}{N_{total}}, \tag{7}$$

where, $Accuracy$ is the classification accuracy, $N_{correct}$ is the number of samples correctly classified by the model, and N_{total} is the total number of samples in the test set.

Macro F1-score is adopted to deal with possible category imbalance in the dataset. The experimental results are shown in Table 6 and Fig. 4.

Table 6. Test results of corrosion class classification accuracy

Test items	The proposed model	ViT	CBAM	Efficient Net
Overall classification accuracy	96.8 %	95.1 %	94.3 %	95.5 %
Macro average F1 score	95.7 %	93.5 %	92.8 %	94.1 %
Mild corrosion recall	94.2 %	90.5 %	89.1 %	95.0 %
Heavy corrosion recall	97.5 %	96.0 %	95.8 %	96.2 %
Note: This table reports Task 2 (classification) performance. Detection results (mAP) are reported separately in Section 5.2.4				



a) 3D performance landscape comparison
b) Multi-metric 3D performance comparison
c) 3D confusion matrix- proposcd model
d) 3D recall performance with error bars

Fig. 4. Comparative analysis of corrosion grade classification accuracy

Fig. 4 illustrates the performance distribution of various methods within the space of architectural complexity and feature representation capability, including multi-metric performance profiles, confusion matrices, and recall performance.

The experimental results demonstrate that the confusion matrix of the proposed model exhibits the darkest hue along its main diagonal, indicating the highest consistency between its predictions and the ground-truth labels. The proposed model yields the optimal values in terms of both overall classification accuracy and macro F1-score. This suggests that the feature fusion mechanism incorporated in the model design effectively aggregates global and local information, thereby enhancing the confidence level of classification decisions. In practical applications, cable images are often subject to various degradation factors, such as noise, illumination variations, and partial occlusions. Tests are conducted to evaluate the robustness of the model under these non-ideal conditions.

5.2.2. Anti-interference robustness test

Cable images in practical applications are often disturbed by noise, illumination changes, and partial occlusion. Tests are designed to assess the robustness of the model under non-ideal conditions.

The test comparison methods include Denoising Convolutional Neural Network (DnCNN) [24], Adversarial Training [25], and ResNet [26].

The calculation formula is as follows:

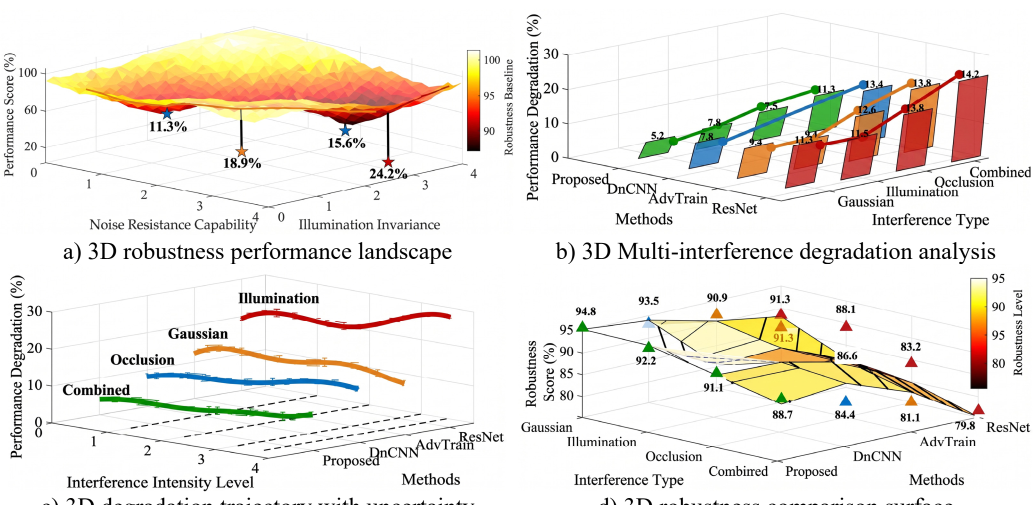
$$PerformanceDrop = \frac{P_{clean} - P_{noisy}}{P_{clean}} \times 100\ \%, \tag{8}$$

where, $PerformanceDrop$ represents the performance degradation rate, P_{clean} is the macro F1-score of the model on the clean test set, and P_{noisy} represents the macro F1-score of the model on the test set after adding interference.

The experimental results are shown in Table 7 and Fig. 5.

Table 7. Anti-interference robustness test results

Test items	The proposed model	DnCNN	Confrontation training	ResNet
Gaussian noise	5.2 %	7.8 %	9.1 %	12.5 %
Light mutation	6.5 %	–	8.7 %	11.9 %
Partial occlusion	8.9 %	–	13.4 %	16.8 %
Integrated jamming	11.3 %	15.6 %	18.9 %	24.2 %



a) 3D robustness performance landscape

b) 3D Multi-interference degradation analysis

c) 3D degradation trajectory with uncertainty

d) 3D robustness comparison surface

Fig. 5. Comparative analysis of anti-interference robustness test

Fig. 5 shows the distribution of robustness of different methods in the space of noise resistance and illumination invariance, including multi-interference degradation analysis, degradation trajectory and uncertainty, and a robustness comparison matrix.

The polygon corresponding to the model proposed in this study has the smallest area and is the closest to the coordinate origin, indicating that its performance degradation under various disturbances is the slightest. This demonstrates that the proposed model exhibits the minimum performance degradation rate and good robustness, making it more suitable for deployment in actual power scenarios with variable environmental conditions.

5.2.3. Computational complexity and inference efficiency test

The purpose of the test is to quantify the difference between the proposed model and the comparison methods in terms of computational complexity and actual inference time, and to assess the feasibility of their engineering applications. The test comparison methods include MobileNet [27], ShuffleNet [28], and Deep Separable Convolution [29].

The calculation formula is as follows:

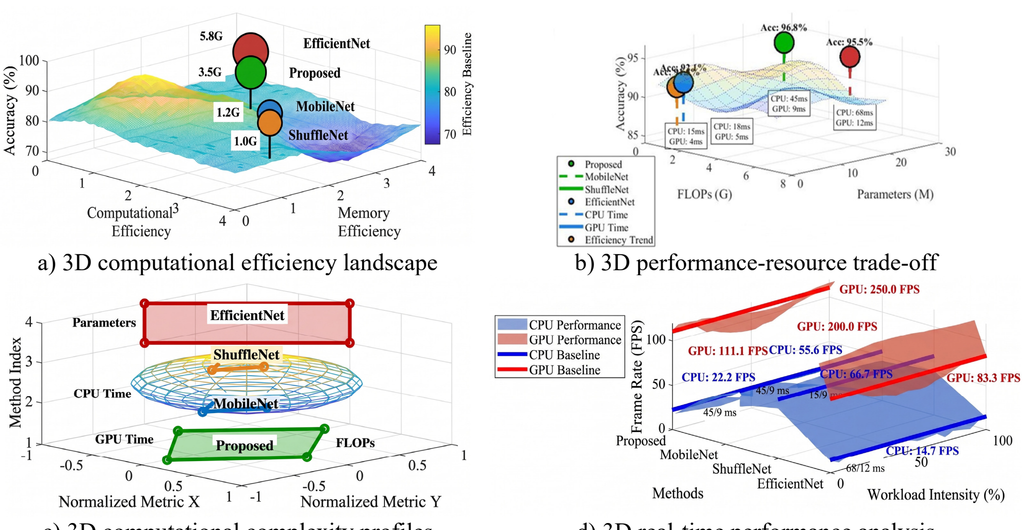
$$FLOPs = 2 \times H \times W \times C_{in} \times C_{out} \times K_h \times K_w, \tag{9}$$

where, $FLOPs$ represents the number of floating-point operations of the convolutional layer; H and W are the height and width of the output feature map, respectively; C_{in} is the number of input channels; C_{out} is the number of output channels; and K_h and K_w are the height and width of the convolution kernel, respectively.

The experimental results are shown in Table 8 and Fig. 6.

Table 8. Test results of computational complexity and inference efficiency

Test items	The proposed model	MobileNet	ShuffleNet	EfficientNet
Floating-point operations (FLOPs)	3.5G	1.2G	1.0G	5.8G
Number of parameters	15.3M	3.4M	2.8M	20.1M
CPU average inference time (ms)	45	18	15	68
GPU average inference time (ms)	9	5	4	12



a) 3D computational efficiency landscape b) 3D performance-resource trade-off

c) 3D computational complexity profiles d) 3D real-time performance analysis

Fig. 6. Comprehensive analysis on computational complexity and inference efficiency

Fig. 6 displays the performance distribution in computational and memory efficiency space, including performance-resource trade-offs, computational complexity profiles, and real-time performance analysis.

The proposed model has higher computational complexity and parameter count than MobileNet and ShuffleNet, which are specifically designed for mobile platforms, yet it remains substantially lower than the bulky EfficientNet. Under GPU acceleration, it achieves an extremely short inference time of merely 9 milliseconds, fully meeting real-time processing requirements.

The inference latency of 9 ms per frame under GPU acceleration (NVIDIA RTX A6000) translates to a theoretical throughput of over 100 FPS, significantly surpassing the 25-30 FPS requirement typical for real-time video analysis. This performance margin provides a crucial buffer for practical deployment, accommodating potential overhead from preprocessing (e.g., video decoding, image stabilization) and postprocessing (e.g., alert generation, data logging) without compromising the system's real-time capabilities. When considering deployment on resource-constrained edge devices, however, the model's 15.3M parameters and 3.5G FLOPs present a challenge. Two primary strategies are envisaged for adaptation: first, the application of channel pruning to the convolutional layers, which could reduce the parameter count by an estimated 30-40 % with minimal accuracy trade-off; second, the quantization of weights from FP32 to INT8, a technique that demonstrably accelerates inference on edge TPUs. Such optimizations are essential for embedding the system onto inspection drones or portable handheld units.

The experimental results have demonstrated that the proposed model delivers comprehensive advantages in corrosion grade classification accuracy, robustness against disturbances, and computational efficiency. In terms of classification accuracy, the model outperforms contemporary state-of-the-art methods by approximately 1.3 % to 2.5 %, with particularly notable performance in identifying severe corrosion. During interference resistance tests, the model exhibits the lowest performance degradation rate under composite disturbances, reducing degradation by more than half compared to baseline models. Although computational efficiency is marginally lower than that of lightweight networks, the model still achieves real-time inference at 9 ms with GPU acceleration, effectively balancing accuracy and speed. These findings validate the efficacy of the model design.

6. Conclusions

In conclusion, we introduce MSFEM-AAM – a deep learning model for cable corrosion detection – which employs an innovative parallel architecture to enable simultaneous feature extraction and classification of corrosion images, thereby enhancing the operational efficiency of corrosion assessment in power cable inspection. Experiments conducted in a simulated environment indicate that MSFEM-AAM achieves a classification accuracy of 96.8 % with a noise immunity degradation rate of 11.3 %. Under GPU acceleration, the inference time is only 9 ms, which meets the real-time requirements of future power inspection tasks. These quantitative improvements indicate substantial breakthroughs in the model’s accuracy, robustness, and efficiency.

Beyond raw performance metrics, the system’s scalability for real-world deployment has been considered. The architecture’s modular design – where the YOLOv10 and Transformer branches function as semi-independent modules – facilitates future upgrades. For instance, the feature extractor could be swapped for a more efficient variant without necessitating a complete model redesign. Furthermore, the framework is inherently scalable to multi-sensor fusion. The current RGB-based detection could be extended by integrating a parallel processing stream for infrared thermal data using a similar dual-branch logic. This would allow the system to not only detect visible surface corrosion but also identify subsurface anomalies indicated by thermal differentials, thereby providing a more comprehensive assessment of cable health. Future work will focus on developing a lightweight, quantized version of this model specifically optimized for deployment on embedded platforms, followed by field trials to validate its robustness and long-term reliability

ISSN PRINT 2335-2124, ISSN ONLINE 2424-4635

in uncontrolled environmental conditions.

Currently, the model remains in the design phase and has not been deployed for practical validation. Therefore, future work will focus on model lightweighting for deployment on embedded devices, as well as extending the framework to multimodal data fusion by integrating infrared thermal imaging and ultrasonic testing, thereby enhancing the practical applicability of this innovative approach. Further research will investigate the model's online learning capability in dynamic environments to achieve real-time corrosion monitoring and early warning. Additionally, we plan to explore the application of transfer learning in cross-domain corrosion detection, promoting the development of more versatile and intelligent assessment methodologies.

Acknowledgements

The study was supported by "Intelligent AI Application Research Service for Industrial Development Company in 2025 (Grant No. SGBJHW00BGJS2500120)".

Data availability

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

Author contributions

Shiquan Sun: conceptualization. Meng Xie: formal analysis. Yi Feng: investigation. Jiatian Shangguan: methodology; Guangjiu Chen: writing-original draft preparation and writing-review and editing.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Y. He et al., "UAV based sensing and imaging technologies for power system detection, monitoring and inspection: a review," *Nondestructive Testing and Evaluation*, Vol. 40, No. 12, pp. 5681–5748, Dec. 2025, https://doi.org/10.1080/10589759.2024.2421938

[2] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: a comprehensive review," *International Journal of Energy Research*, Vol. 44, No. 9, pp. 7136–7157, Mar. 2020, https://doi.org/10.1002/er.5331

[3] A. Gasparin, S. Lukovic, and C. Alippi, "Deep learning for time series forecasting: the electric load case," *CAAI Transactions on Intelligence Technology*, Vol. 7, No. 1, pp. 1–25, Sep. 2021, https://doi.org/10.1049/cit2.12060

[4] Y. Wu et al., "A visual fault detection algorithm of substation equipment based on improved YOLOv5," *Applied Sciences*, Vol. 13, No. 21, p. 11785, Oct. 2023, https://doi.org/10.3390/app132111785

[5] C.-K. Lee and Y.-J. Shin, "Detection and assessment of I&C cable faults using time-frequency R-CNN-based reflectometry," *IEEE Transactions on Industrial Electronics*, Vol. 68, No. 2, pp. 1581–1590, Feb. 2020, https://doi.org/10.1109/tie.2020.2970677

[6] I. Katsamenis, E. Protopapadakis, A. Doulamis, N. Doulamis, and A. Voulodimos, "Pixel-level corrosion detection on metal constructions by fusion of deep learning semantic and contour segmentation," in *Lecture notes in computer science*, Vol. 12509, Cham: Springer International Publishing, 2020, pp. 160–169, https://doi.org/10.1007/978-3-030-64556-4_13

[7] Y. Liu, Z. Zhang, X. Liu, W. Lei, and X. Xia, "Deep learning based mineral image classification combined with visual attention mechanism," *IEEE Access*, Vol. 9, pp. 98091–98109, Jul. 2021, https://doi.org/10.1109/access.2021.3095368

[8] Y. Cheng et al., "A multi-feature fusion and attention network for multi-scale object detection in remote sensing images," *Remote Sensing*, Vol. 15, No. 8, p. 2096, Apr. 2023, https://doi.org/10.3390/rs15082096

[9] H. Li et al., "Sematic segmentation of loess landslides with STAPLE mask and fully connected conditional random field," *Landslides*, Vol. 20, No. 2, pp. 367–380, Nov. 2022, https://doi.org/10.1007/s10346-022-01983-8

[10] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *EURASIP Journal on Wireless Communications and Networking*, Vol. 2020, No. 1, p. 204, 2020, https://doi.org/10.1186/s13638-020-01826-x

[11] G. Choudhary and D. Sethi, "From conventional approach to machine learning and deep learning approach: an experimental and comprehensive review of image fusion techniques," *Archives of Computational Methods in Engineering*, Vol. 30, No. 2, pp. 1267–1304, 2022, https://doi.org/10.1007/s11831-022-09833-5

[12] L. Bian et al., "Current status of bridge cable climbing robot technology and design of cable corrosion detection device," in *5th International Conference on Mechatronics Technology and Intelligent Manufacturing (ICMTIM)*, pp. 542–547, Apr. 2024, https://doi.org/10.1109/icmtim62047.2024.10629445

[13] S. Liu, Y. Sun, X. Jiang, and Y. Kang, "A review of wire rope detection methods, sensors and signal processing techniques," *Journal of Nondestructive Evaluation*, Vol. 39, No. 4, p. 85, 2020, https://doi.org/10.1007/s10921-020-00732-y

[14] J. Hu, W. Wan, P. Qiao, Y. Zhou, and A. Ouyang, "Power insulator defect detection method based on enhanced YOLOV7 for aerial inspection," *Electronics*, Vol. 14, No. 3, p. 408, Jan. 2025, https://doi.org/10.3390/electronics14030408

[15] K. Yang, Y. Zhang, X. Zhang, and L. Zheng, "YOLOX with CBAM for insulator detection in transmission lines," *Multimedia Tools and Applications*, Vol. 83, No. 14, pp. 43419–43437, 2023, https://doi.org/10.1007/s11042-023-17245-1

[16] M. Ayub and E.-S. M. El-Alfy, "Household appliance identification using vision transformers and multimodal data fusion," *IEEE Transactions on Consumer Electronics*, Vol. 71, No. 2, pp. 2774–2792, Apr. 2025, https://doi.org/10.1109/tce.2025.3565850

[17] Y. Bai, H. Sun, L. Zhang, and H. Wu, "Hybrid CNN-transformer network for electricity theft detection in smart grids," *Sensors*, Vol. 23, No. 20, p. 8405, Oct. 2023, https://doi.org/10.3390/s23208405

[18] X. Shao, C.-S. Kim, and P. Sontakke, "Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN-LSTM," *Energies*, Vol. 13, No. 8, p. 1881, Apr. 2020, https://doi.org/10.3390/en13081881

[19] J. Bian, L. Wang, R. Scherer, M. Wozniak, P. Zhang, and W. Wei, "Abnormal detection of electricity consumption of user based on particle swarm optimization and long short term memory with the attention mechanism," *IEEE Access*, Vol. 9, pp. 47252–47265, Mar. 2021, https://doi.org/10.1109/access.2021.3062675

[20] M. Wenhao, T. Yuanpeng, H. Fangzhou, and L. Haiying. "Visible light detection dataset for cable equipment defects (V1)." China Electric Power Research Institute Co., Ltd., https://cstr.cn/16666.11.nbsdc.gsn8dm78 (accessed May 2023).

[21] T. Yin et al., "Corrosion characteristics of anchor cables in electrolytic corrosion test and the applicability of the test method in study of anchor cable corrosion," *Advances in Civil Engineering*, Vol. 2021, No. 1, p. 6695288, Jan. 2021, https://doi.org/10.1155/2021/6695288

[22] M. Bounabi et al., "Smart PV monitoring and maintenance: a vision transformer approach within urban 4.0," *Technologies*, Vol. 12, No. 10, p. 192, Oct. 2024, https://doi.org/10.3390/technologies12100192

[23] X. Yang, Q. Zhang, S. Wang, and Y. Zhao, "Detection of solar panel defects based on separable convolution and convolutional block attention module," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Vol. 45, No. 3, pp. 7136–7149, Jun. 2023, https://doi.org/10.1080/15567036.2023.2218301

[24] M. D. Fitri Mat Zabidi, S. Shahbudin, S. I. Sulaiman, F. Y. Abdul Rahman, and H. Saad, "Power quality disturbances classification analysis using EfficientNet architecture," in *2024 IEEE 15th Control and System Graduate Research Colloquium (ICSGRC)*, pp. 64–69, Aug. 2024, https://doi.org/10.1109/icsgrc62081.2024.10691145

[25] A. E. Ilesanmi and T. O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review," *Complex and Intelligent Systems*, Vol. 7, No. 5, pp. 2179–2198, 2021, https://doi.org/10.1007/s40747-021-00428-4

[26] W. Li, D. Deka, R. Wang, and M. R. A. Paternina, "Physics-constrained adversarial training for neural networks in stochastic power grids," *IEEE Transactions on Artificial Intelligence*, Vol. 5, No. 3, pp. 1121–1131, Jan. 2023, https://doi.org/10.1109/tai.2023.3236377

[27] X. Chen, W. Chen, V. Dinavahi, Y. Liu, and J. Feng, "Short-term load forecasting and associated weather variables prediction using ResNet-LSTM based deep learning," *IEEE Access*, Vol. 11, pp. 5393–5405, Jan. 2023, https://doi.org/10.1109/access.2023.3236663

[28] M. P. Singh and N. Sinha, "A novel image classification approach via MobileNet model for security assessment of power system," *Electric Power Components and Systems*, Vol. 53, No. 12, pp. 1381–1394, Jul. 2025, https://doi.org/10.1080/15325008.2024.2329336

[29] Y. Huang, Y. Lu, L. Fan, K. Xiang, and H. Ma, "Fault arc detection method based on improved ShuffleNet V2 network," *Processes*, Vol. 13, No. 1, p. 135, Jan. 2025, https://doi.org/10.3390/pr13010135

Shiquan Sun earned a master's degree in architecture and civil engineering from Beijing University of Civil Engineering and Architecture in 2019. Currently, he works as an engineer in the Office of Beijing Huashang Electric Power Industry Development Co., Ltd. His main research areas include electrical equipment safety inspection, equipment safety protection, and artificial intelligence applications.

Meng Xie earned a master's degree in Systems Engineering and Engineering Management from The Chinese University of Hong Kong (CUHK) in 2016. Currently, he works as an engineer in the Office of Beijing Huashang Electric Power Industry Development Co., Ltd. His main research areas include: power engineering management, planning, design, and evaluation of management information systems (MIS).

Yi Feng earned a bachelor's degree in Internet of Things (IoT) Engineering from Beijing University of Technology (BJUT) in 2018. Currently, he works as an engineer in the Safety Supervision Department of Beijing Huashang Electric Power Industry Development Co., Ltd. His main research areas include: transformer equipment maintenance, production safety management, and artificial intelligence (AI) applications.

Jiatian Shangguan earned a master's degree in Power System and Automation from Wuhan University in 2008. Currently, he works as a Senior Engineer at Beijing Huashang Electric Power Industry Development Co., Ltd. His main research areas include: power system automation, relay protection, and artificial intelligence (AI) applications.

Guangjiu Chen earned a bachelor's degree in electrical engineering and Automation from Sichuan University in 2004. Currently, he works as a Senior Engineer at Beijing Huashang Electric Power Industry Development Co., Ltd. His main research areas include: power system automation, artificial intelligence (AI) applications, and related fields.