EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Design of intelligent inspection system for 500 kV substation based on multi-source data fusion

Yaoshan Zhang[1, *], Zhiqiang Xiao[2], Zhihua Lin[1], Xiuquan Hu[2], Yue Zhou[2]

[1]Hainan Power Grid Co. Ltd., Haikou, 570203 Hainan, China
[2]Hainan Power Grid Co. Ltd. Construction Branch, Haikou, 570203 Hainan, China

Abstract

INTRODUCTION: The safe and stable operation of 500 kV substations is essential for ensuring power grid reliability. Traditional manual inspection methods suffer from low efficiency, limited coverage, and poor real-time performance, making intelligent inspection technologies increasingly important.
OBJECTIVES: This study aims to develop an intelligent inspection system for 500 kV substations based on multi-source data fusion to improve equipment status monitoring, fault diagnosis, and operational reliability.
METHODS: The proposed system integrates infrared thermal images, visible-light images, sound signals, and vibration data for comprehensive equipment perception. A multi-modal deep learning network with an attention mechanism (AMM-Net) is designed for adaptive feature extraction. In addition, pixel-level, feature-level, and decision-level fusion strategies are combined with a trust-based distributed Kalman filtering algorithm (Trust-DKF) to improve robustness and anti-interference capability.
RESULTS: Experimental results show that the system achieves 94.7% accuracy and 93.1% F1-score in equipment status recognition. Under noise and occlusion interference, performance decreases by only 10.7%. The edge-device inference time is optimized to 28.9 ms with low energy consumption of 0.12 J per operation.
CONCLUSION: The proposed system significantly improves the efficiency, accuracy, and reliability of intelligent substation inspection.

Keywords: multi-source data fusion; 500 kV substation; intelligent inspection; deep learning; state estimation

Received on 27 May 2026, accepted on 23 July 2026, published on 24 August 2026

doi: 10.4108/ew.13195

1. Introduction

T As the core hub of the power system, the operation status of 500 kV substation is directly related to the reliability and safety of the power grid. Traditional inspection methods mainly rely on manual operations and have outstanding problems such as low efficiency, limited coverage, and poor real-time performance, making it difficult to meet the development needs of modern smart grids. As the scale of substation equipment continues to expand and operation and maintenance requirements increase day by day, traditional methods face severe challenges in terms of inspection quality, safety risk prevention and control, and data analysis depth. Although Internet of Things technology and sensor networks have been widely used in power systems in recent years, due to the complex substation environment, strong electromagnetic interference, and wide variety of equipment, single sensor data often has limitations and cannot fully and accurately reflect the operating status of equipment, resulting in existing automated inspection systems still have obvious

*Corresponding Author. Email: YaoShan_Zhang@126. com

shortcomings in terms of accuracy, robustness, and adaptability.

This paper aims to design a 500 kV substation intelligent inspection system based on multi-source data fusion. By innovatively integrating multi-dimensional information such as infrared thermal images, visible light images, sound signals and vibration data, it can achieve accurate perception and intelligent decision-making of equipment status. The research focuses on breaking through key technical problems such as multi-modal feature extraction, multi-level data fusion and robust state estimation, and innovatively proposes a multi-modal deep learning network (AMM-Net) based on attention mechanism and a distributed trust mechanism. Kalman filter (Trust-DKF) algorithm. The main contribution of this paper is to establish a complete intelligent inspection algorithm framework and verification system, which effectively improves the information utilization rate through the three-level fusion strategy. This significantly enhances the system's anti-interference ability and cross-scenario adaptability, provides important theoretical support and technical solutions for intelligent operation and maintenance of substations, and promotes the development of power inspection technology in the direction of intelligence and precision.

2. Related works

As the power system evolves towards intelligence and automation, the innovation of inspection technology for 500 kV substations, as the core hub of the power grid, has become crucial for ensuring the safe and stable operation of the grid. Traditional manual inspection methods, due to issues such as low efficiency and limited coverage, have become difficult to meet the demands of modern power grids. The introduction of technologies such as multi-source data fusion, unmanned aerial vehicles (UAVs), and artificial intelligence algorithms has provided new ideas for intelligent inspection.

(1) Research on the application of drones in power inspection

Drone inspection technology effectively addresses the inspection challenges posed by the wide coverage and complex environment of power transmission lines, thanks to its aerial perspective and flexible maneuverability. The core issues in this field lie in achieving autonomous navigation for UAVs, controlling safe distances, and optimizing multimodal data collection in a coordinated manner. Li et al. [1] designed an autonomous drone inspection system that combines visual sensors with path planning algorithms to achieve real-time monitoring of high-voltage transmission lines. Experiments showed that the system achieved a detection accuracy rate of over 90% in complex terrains, but it was susceptible to weather conditions, leading to data loss. Chen et al. [2] focused on safety issues and analyzed the safe distances for drones during inspections of 500kV transmission towers through electric field simulations, proposing a dynamic adjustment strategy. However, the model did not consider the variability factors of extreme electromagnetic interference. Xu et al. [3] further developed an automatic inspection system guided by power lines, utilizing line features to achieve precise positioning for drones and reducing manual intervention. However, its computational complexity made it difficult to deploy on low-power edge devices. Gao et al. [4] constructed an efficiency evaluation model for drone inspections from a management perspective, verifying a 20% improvement in resource scheduling efficiency through case studies, but lacking a multi-objective collaborative optimization mechanism. Multimodal data fusion has become a trend. Zhang et al. [5] proposed a multi-agent collaborative framework based on federated reinforcement learning, optimizing task allocation for drone swarms through distributed learning, significantly improving inspection coverage, but with high communication overhead. De Albuquerque et al. [6] explored a multimodal predictive inspection method that combines infrared and visible light data to identify line defects in advance, although the model's generalization ability was limited by the training data.

Furthermore, Rong et al. [7] used stereo vision technology to detect vegetation encroachment, achieving an accuracy rate of 88%, but the algorithm was sensitive to changes in illumination. Zhang et al. [8] and Xu et al. [9] separately studied electric field safe distances and 3D trajectory planning methods for drone inspections within substations, enhancing the reliability of close-range operations, but the real-time obstacle avoidance capability still needs improvement. Overall, drone inspections have made significant progress in automation and intelligence, but environmental adaptability, energy consumption optimization, and real-time fusion of multimodal data remain key challenges.

(2) Intelligent inspection technology for substations

The internal equipment of substations is dense, and the electromagnetic environment is complex. Intelligent inspection requires the integration of robots, video surveillance, and fault diagnosis algorithms to achieve precise perception of equipment status. The core issues in this field lie in the collaborative processing of multi-source heterogeneous data and the robust design of system architecture. Li et al. [10] optimized the layout of substations and multi-scenario inspection schemes through video surveillance data, improving monitoring coverage, but the algorithm's recognition accuracy for occluded scenes was insufficient. Xuefeng et al. [11] designed a live line measurement inspection platform based on a multi-agent system, achieving distributed decision-making, but communication delays between agents may affect real-time performance. Wang et al. [12] studied a new digital intelligent substation architecture that integrates the Internet of Things and cloud computing, supporting dynamic resource allocation, but the security protection mechanism is weak. Robot technology is an important supplement. Zhong et al. [13] developed a broken strand repair robot for 500kV high-voltage live line work, reducing personal risk through precise operation by a robotic arm, although the control accuracy in dynamic environments needs further improvement; Liu et al. [14] improved the SLAM algorithm of inspection robots, integrating 3D laser and visual information to enhance positioning accuracy, but the consumption of computational resources is high. In terms of fault diagnosis, Lei et al. [15] used neural networks to detect electrical equipment faults, achieving an accuracy rate of 85% in simulations, but the model relies on a large amount of labeled data; Li et al. [16] proposed a remote intelligent diagnosis method for secondary equipment, reducing on-site intervention through simulated signal analysis, but the anti-interference capability is limited. Special application scenarios have also received attention. For example, Pei et al. [17] designed an intelligent elimination device for high-voltage transmission line pin defects, verifying its reliability through structural simulation, although the actual deployment cost is high; Lei et al. [18] focused on early warning technology for external fractures of submarine cables, combining acoustic sensing to achieve early detection, but the durability in deep-sea environments remains a bottleneck. In summary, intelligent inspection of substations presents a trend of multi-technology integration, but system integration, real-time performance, and adaptability to extreme operating conditions need to be strengthened.

(3) Communication technology and network architecture

An efficient and reliable communication network serves as the foundational support for intelligent inspection systems. The core challenge lies in ensuring the real-time, secure, and widespread transmission of data. Gan et al. [19] applied 5G communication technology to intelligent inspection of 750kV substations, achieving real-time processing of high-definition video streams through low-latency transmission. In their experiments, latency was reduced to the millisecond level. However, the deployment cost of base stations was high, and coverage in rural areas was insufficient. Liu et al. [20] explored the application of wireless sensor networks in power transmission and distribution systems, designing an adaptive routing protocol to enhance data collection efficiency.

However, packet loss often occurred during dynamic changes in network topology. These studies highlight the empowering role of communication technology in inspection efficiency. However, the integration scheme of 5G and sensor networks, anti-electromagnetic interference mechanisms, and cross-domain coordination standards are yet to be perfected, limiting widespread promotion.

In existing research, although significant progress has been made in substation intelligent inspection technology, there are still three notable deficiencies: Firstly, there is a heavy reliance on single-type sensor data (such as vision or sound alone), lacking deep fusion of multi-source information such as infrared, visible light, and vibration, resulting in a single dimension of state perception and insufficient complementarity. Secondly, existing fusion methods often remain at a single level (such as the decision-making level), failing to fully utilize the potential of pixel-level and feature-level fusion in detail preservation and semantic enhancement. Thirdly, when facing complex scenarios such as strong electromagnetic interference and equipment obstruction in substations, traditional algorithms exhibit poor robustness, and there are challenges in balancing computational efficiency and energy consumption in edge deployment. To address these issues, this paper proposes an intelligent inspection system based on multi-source data fusion. It achieves adaptive weighted fusion of heterogeneous features through the introduction of an attention-based multimodal deep learning network (AMM-Net), combines three-level fusion strategies at the pixel, feature, and decision-making levels to enhance information utilization, and innovatively integrates a trust mechanism with distributed Kalman filtering (Trust-DKF) to enhance anti-interference capability, thereby significantly improving the accuracy and robustness of state recognition while ensuring real-time performance.

3. Algorithm model design

As a key node in the power system, the operational status of equipment in a 500 kV substation directly affects the reliability and safety of the power grid. Traditional inspection methods rely on manual operations, which suffer from issues such as low efficiency, limited coverage, and poor real-time performance, making them difficult to meet the development needs of smart grids. With the advancement of Internet of Things (IoT) and artificial intelligence technologies, intelligent inspection systems based on multi-source data fusion have become a research hotspot. This paper focuses on the core algorithm models in intelligent inspection of substations, aiming to achieve precise perception and intelligent decision-making of equipment status through innovative multi-source data fusion technologies.

This section presents a comprehensive algorithmic framework that integrates multidimensional information, including infrared thermal images, visible light images, sound signals, and vibration data. The core of the framework comprises four modules: data preprocessing, multimodal feature extraction, multilevel data fusion, and robust state estimation. The innovations lie in the following aspects: Firstly, an attention-based multimodal deep learning network (AMM-Net) is designed to adaptively weight heterogeneous features; secondly, a three-level fusion strategy (pixel level, feature level, and decision level) is introduced to enhance information utilization; finally, a distributed Kalman filter with trust mechanism (Trust-DKF) is incorporated to bolster the system's resistance to attacks. This model ensures practicality and robustness in complex substation environments through theoretical derivation and simulation verification.

3.1. Overall system framework

The algorithm framework of the intelligent inspection system adopts a hierarchical distributed structure, as shown in Figure 1, which includes five levels: data acquisition, preprocessing, feature extraction, multi-source fusion and intelligent decision-making. The data acquisition layer acquires multi-modal data in real time through heterogeneous sensor networks such as infrared cameras, visible light cameras, microphone arrays, and vibration sensors. Infrared images, visible-light images, acoustic signals, and vibration data were selected because they provide complementary information for equipment condition assessment. Infrared images detect thermal anomalies, visible-light images identify surface defects, acoustic signals capture electrical abnormalities, and vibration data reflects mechanical faults, enabling more reliable condition assessment. The preprocessing layer standardizes, denoises and aligns the raw data to eliminate environmental interference and sensor differences. The feature extraction layer uses a deep learning network to mine deep features from multi-source data to form a unified representation. The multi-source fusion layer integrates complementary information through three strategies: pixel level, feature level and decision level. The intelligent decision layer finally outputs device state estimation results, such as anomaly identification, fault classification, and early warning signals.

The innovation of this framework lies in its modular design and adaptive mechanism. Each module can be independently optimized and the parameters are dynamically adjusted via a feedback loop. For example, the preprocessing module adaptively selects the filtering algorithm according to the signal signal-to-noise ratio, and the feature extraction network focuses on the key features through the attention mechanism. The overall framework supports distributed deployment, which is suitable for large-scale inspection needs of substations, while ensuring real-time performance through lightweight design.

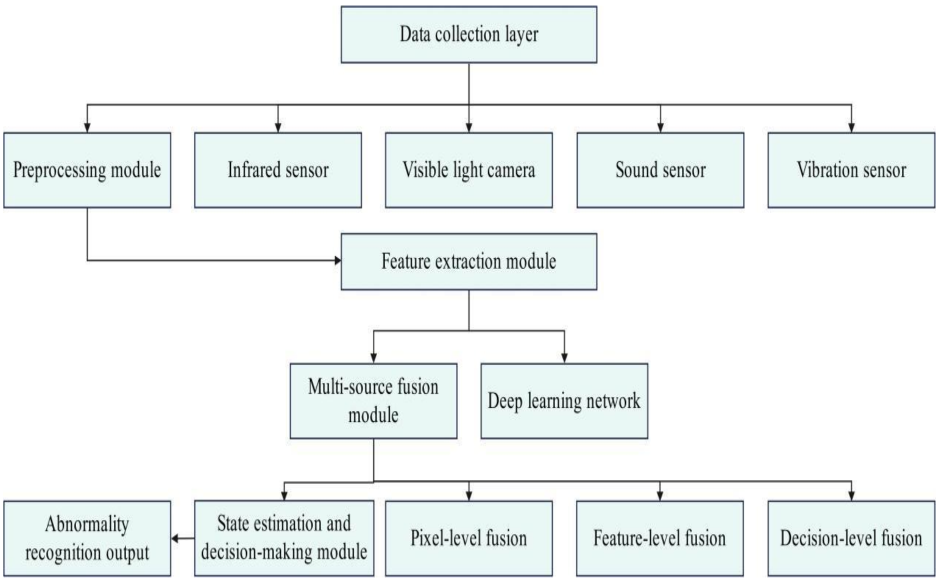

Figure 1. Algorithm framework of intelligent inspection system

3.2. Data preprocessing module

Multi-source data is easily affected by electromagnetic interference, environmental noise and sensor drift in substation during the acquisition process, so it needs refined preprocessing. This module includes four steps: data cleaning, standardization, denoising and registration to ensure the consistency and reliability of input features.

Data cleaning: For missing values or outliers, sliding window detection and interpolation repair are used. The original data matrix is $X_{raw} \in R^{N\times T}$, where N is the number of sensors and T is the length of the time series. Outlier detection is based on the Z-score algorithm:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

μ represents the mean of the data sequence, and σ represents the standard deviation of the data sequence. If $|z_i| > 3$ (with an adjustable threshold), the data point is identified as an outlier and replaced with the neighborhood mean.

The Z-score-based outlier detection method was selected because of its low computational complexity and effectiveness in identifying abnormal sensor measurements in real-time monitoring environments. Compared with alternative anomaly filtering methods, such as the Interquartile Range (IQR), Local Outlier Factor (LOF), and Isolation Forest, the Z-score approach provides efficient statistical detection with minimal computational overhead, making it well suited for intelligent substation monitoring applications. Normalization: To eliminate dimensional effects, Min-Max normalization is used to map the data to the [0, 1] interval:

$$X_{\text{norm}} = \frac{x_{raw} - min(x_{raw})}{max(x_{raw}) - min(x_{raw})} \tag{2}$$

X_{norm} is the normalized data matrix. For non-stationary data (such as vibration signals), additional differential processing is performed to enhance stationarity.

Denoising algorithm: For high-frequency noise, it combines wavelet transform and Kalman filtering. The discrete wavelet transform (DWT) decomposes the signal into approximate coefficients and detail coefficients, and the soft threshold function is:

$$\eta(x) = sign(x) \cdot max(0, |x| - \lambda) \tag{3}$$

DWT was selected over conventional denoising methods because substation sensor data often contains non-stationary transient signals, including abrupt thermal, acoustic, and vibration variations. Unlike Fourier Transform, which provides only global frequency information, DWT offers multi-resolution time-frequency analysis that effectively preserves transient fault characteristics while suppressing noise. Compared with median and Wiener filtering, DWT better maintains local signal features and edge information, making it more suitable for reliable fault detection in intelligent substation inspection systems.

Among them, the threshold value λ is adaptively calculated according to the noise variance. Kalman filter is used to track the signal trend, and the equation of state is as follows:

$$x_k = Fx_{k-1} + w_k, z_k = Hx_k + v_k \tag{4}$$

w_k and v_k represent the process and observation noise, respectively, and their covariance matrices are estimated from historical data.

The Kalman filter parameters were optimized according to the electromagnetic interference (EMI) level by adjusting the process noise covariance (Q) and measurement noise covariance (R). Under higher EMI conditions, larger measurement noise values were assigned to reduce the influence of noisy observations, whereas lower noise values were used under stable operating conditions to improve signal tracking accuracy and robustness.

Spatio-temporal registration: Multi-source data needs to align timestamps and spatial coordinates. Temporal registration uses linear interpolation, and spatial registration is based on feature point matching (such as SIFT algorithm) to ensure data synchronization.

Registration accuracy was evaluated by measuring feature matching consistency and timestamp alignment errors across heterogeneous sensing modalities. Registration quality was preserved through feature correspondence verification and synchronization error correction, ensuring accurate alignment of infrared, visible-light, acoustic, and vibration data before multimodal fusion

The spatio-temporal registration workflow first synchronizes sensor data through timestamp alignment using linear interpolation, followed by spatial registration based on SIFT feature matching to establish correspondence among heterogeneous sensing modalities. The synchronized data are then verified for alignment consistency before being forwarded to the multimodal feature extraction module, ensuring reliable sensor synchronization throughout the intelligent inspection process.

The preprocessing process is shown in Figure 2, which improves data quality through multi-stage processing and lays the foundation for subsequent feature extraction.

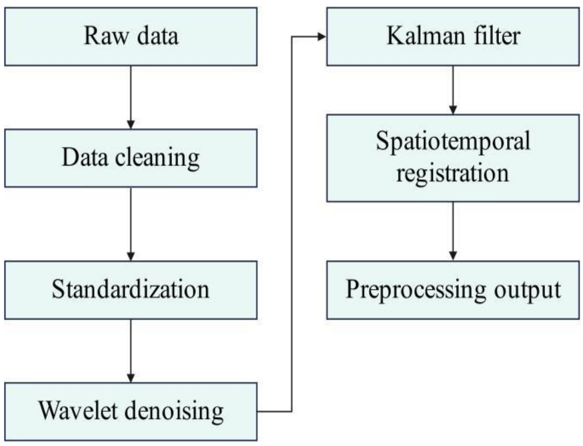

Figure 2. Data preprocessing flow chart

3.3. Multimodal feature extraction network

In order to effectively mine the deep features of multi-source data, this paper designs a multi-modal deep learning network (AMM-Net) based on attention mechanism. The network contains visual branches, acoustic branches and vibrational branches, and adaptively fuses heterogeneous features through attention weights. The structure is shown in Figure 3.

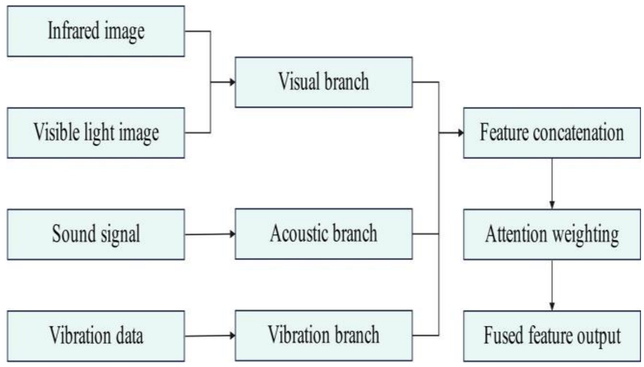

Figure 3. AMM-Net network structure diagram

The architectural configuration of the proposed AMM-Net, including the layer configuration, output feature dimensions, and computational complexity of each branch

Table 1. Architecture configuration of the proposed AMM-Net

Branch	Module/Layer	Output Dimension	Computational Complexity
Visual	ResNet-50 with Coordinate Attention (CA)	2048	High
Acoustic	1D Convolution + Bi-LSTM (128 hidden units)	256	Moderate
Vibration	3-Layer MLP (LeakyReLU)	128	Low
Fusion	Adaptive Attention-based Weighted Concatenation	2432	Low
Output	Fully Connected + Softmax	2	Low

As shown in Table 1, the visual branch performs high-level feature extraction, while the acoustic and vibration branches capture complementary temporal and mechanical information. The extracted features are adaptively fused before final equipment condition classification, providing an efficient balance between feature representation and computational complexity.

Visual branch: Infrared and visible images are input, and the improved ResNet-50 is used as the backbone network. A coordinate attention (CA) module is introduced to enhance spatial feature representation, and the CA module captures location-sensitive information through pooling operations:

$$F_v = CA(Res\,Net(I_v)), F_v \in \mathbb{R}^{C_v \times H \times W} \quad (5)$$

Among them, I_v represents the input image, C_v represents the number of channels, and $H \times W$ represents the feature map size. The CA module is calculated as follows:

$$a_h = \sigma(W_h \cdot GAP(F_v)), a_w = \sigma(W_w \cdot GAP(F_v)) \quad (6)$$

GAP stands for Global Average Pooling, σ is the Sigmoid function, and a_h and a_w represent the attention weights in the height and width directions, respectively.

The Coordinate Attention (CA) module is integrated within the ResNet-50-based visual feature extraction branch to enhance spatial feature representation by encoding channel dependencies and positional information. This enables the network to focus on critical regions, such as hotspots and surface defects, while suppressing background interference, thereby improving feature discrimination and inspection accuracy in complex substation environments.

Acoustic branch: The Mel spectrogram of the audio signal is input, and local features are first extracted through a 1D convolutional layer:

$$F_a^{'} = Re\,LU(W_c * s + b_c) \quad (7)$$

Among them, s represents the spectrogram, and $*$ represents the convolution operation. The Mel-spectrogram was generated using Short-Time Fourier Transform (STFT) and Mel filter banks to obtain a time-frequency representation of the acoustic signal. This representation captures key operational characteristics, such as transformer hum, partial discharge, and arcing, thereby improving fault identification

This is followed by a bidirectional LSTM network to capture temporal dependencies:

$$h_t = \mathrm{LSTM}(F_a^{'}, h_{t-1}), F_a = h_T \in \mathbb{R}^{D_a} \quad (8)$$

D_a is a hidden layer dimension.

Vibration branching: The frequency domain features of the vibration signal (obtained through FFT transformation) are input and subjected to nonlinear mapping using a multilayer perceptron (MLP):

$$F_s = MLP(f_{FFT}), F_s \in \mathbb{R}^{Ds} \quad (9)$$

The frequency-domain features were selected from the FFT spectrum based on dominant frequency components and spectral energy, which are sensitive to changes in equipment operating conditions. These features effectively characterize mechanical abnormalities, such as looseness, wear, and abnormal vibration, thereby improving equipment condition assessment in intelligent substations.

The MLP contains three fully connected layers with an activation function of LeakyReLU.

Attention fusion mechanism: In order to balance the contribution of each modal, the adaptive weight is calculated:

$$\alpha_i = \frac{exp(W_i F_i + b_i)}{\sum_{j \in \{v,a,s\}} exp(W_j F_j + b_j)} \quad (10)$$

The attention-based adaptive weighting mechanism dynamically adjusts the contribution of each modality according to the extracted feature characteristics under different operating conditions. Infrared features receive higher weights for thermal abnormalities, whereas acoustic and vibration features are emphasized for mechanical faults. This adaptive weighting enables reliable heterogeneous feature integration and improves equipment status recognition.

The features after weighted fusion are:

$$F_{\text{fusion}} = \alpha_v F_v \oplus \alpha_a F_a \oplus \alpha_s F_s \quad (11)$$

$\oplus$ represents a splicing operation. The network is trained end-to-end, and the loss function combines cross-entropy with central loss:

$$L = L_{CE} + \lambda L_{\text{center}} \quad (12)$$

This is done to enhance within-class compactness and between-class separability.

AMM-Net was implemented in PyTorch. The visual branch employs ResNet-50 with the Coordinate Attention module, the acoustic branch uses a 1D-CNN followed by a Bi-LSTM with 128 hidden units, and the vibration branch consists of a three-layer MLP. The network was trained using the Adam optimizer with a learning rate of 0.001, batch size of 32, and 100 epochs with early stopping based on validation loss.

3.4. multi-source data fusion algorithm

Multi-source fusion is the key to improve the accuracy of state estimation. This paper proposes a three-level fusion strategy (pixel level, feature level, decision level), as shown in Figure 4, taking into account data complementarity and computational efficiency. The proposed multi-source data fusion framework adaptively integrates complementary information from infrared, visible-light, acoustic, and vibration modalities using an attention-based weighting mechanism. Infrared and visible-light images provide thermal and surface condition information, while acoustic and vibration signals capture electrical and mechanical abnormalities. AMM-Net dynamically assigns higher weights to the most informative modalities under different operating conditions, and the weighted features are integrated through pixel-level, feature-level, and decision-level fusion to improve the accuracy and robustness of equipment status recognition.

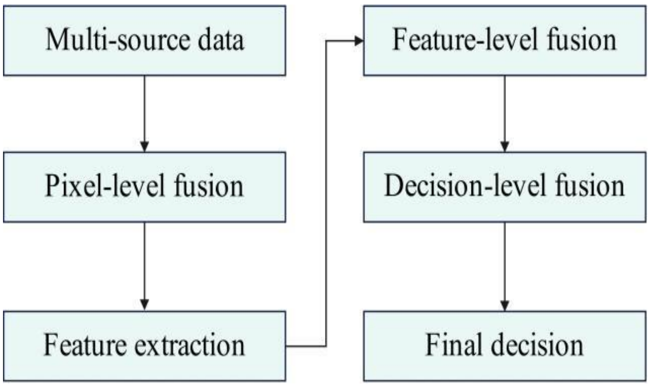

Figure 4. Three-level data fusion process

The overall workflow of the proposed three-level fusion strategy is illustrated in Figure 5, showing how pixel-level, feature-level, and decision-level fusion progressively enrich multimodal information for robust equipment condition assessment.

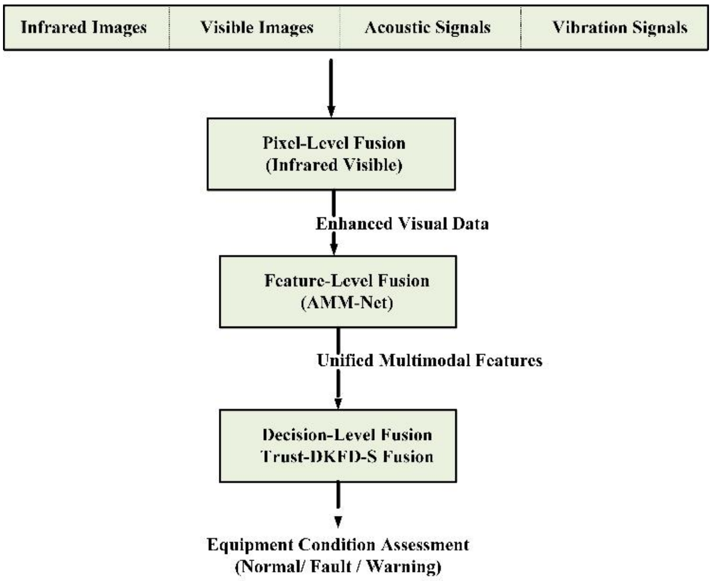

Figure 5. Three-level multi-source data fusion framework of the proposed intelligent inspection system

Pixel-level fusion: It is suitable for homogeneous sensor data (such as dual-spectral images) and employs weighted averaging and Laplacian pyramid fusion. The image from the k-th sensor is I_k, and the fusion result is:

$$I_{\text{fused}} = \sum_{k=1}^{K} w_k I_k, \sum w_k = 1 \tag{13}$$

The weights w_k are dynamically adjusted by the sensor confidence, which is calculated based on historical errors:

$$w_k = \frac{\frac{1}{E_k^2}}{\sum_j \frac{1}{E_j^2}}, E_k = \left\| I_k - I_{ref} \right\|_2 \tag{14}$$

For multi-resolution images, the Laplacian pyramid is decomposed and then fused layer by layer, retaining detailed information.

Feature-level fusion: The multi-modal features output by AMM-Net are input into the gated loop unit (GRU) to model the timing relationship. The update gate and reset gate of the GRU are calculated as follows:

$$z_k = \sigma\left(W_z\left[h_{t-1}, F_{fusion}\right]\right), r_t = \sigma\left(W_r\left[h_{t-1}, F_{fusion}\right]\right) \tag{15}$$

The hidden status is updated to:

$$\tilde{h}_t = tanh\left(W\left[r_t \odot h_{t-1}, F_{fusion}\right]\right), h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{16}$$

The final feature h_t is output as a fusion.

Decision-level fusion: Based on Dempster-Shafer evidence theory, local decisions are integrated to improve robustness. The identification framework is Θ= {normal, abnormal}, and the basic probability assignment (BPA) function is constructed from the Softmax output:

$$m_i(A) = \frac{\exp(s_i(A))}{\sum_B \exp(s_i(B))} \tag{17}$$

Among them, $s_i(A)$ is the classifier's score for proposition A. The fusion rule is:

$$m_{\text{final}}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1-K}, K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{18}$$

K is a conflict factor, and the adaptive weighting strategy is triggered when the conflict is too large.

The Dempster–Shafer (D-S) evidence fusion method manages contradictory evidence from different sensor modalities by evaluating the confidence assigned to each sensor and combining the evidence according to their belief values. Evidence with higher confidence contributes more to the final decision, while conflicting or uncertain evidence receives lower influence, thereby improving the reliability and robustness of equipment fault identification.

3.5. State estimation and decision module

In complex substation environments, sensors are vulnerable to attack or failure, resulting in state estimation bias. This paper innovatively introduces the trust mechanism into distributed Kalman filter (DKF) and proposes the Trust-DKF algorithm. The process is shown in Figure 6.

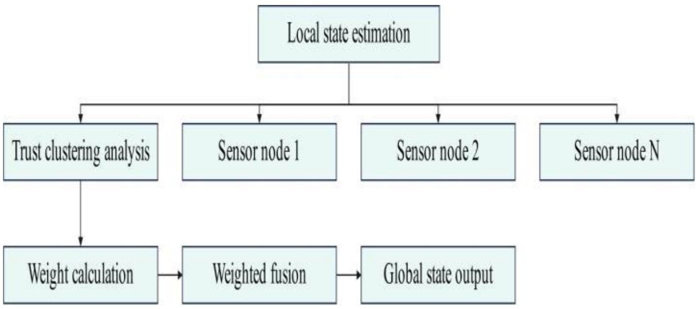

Figure 6. Flowchart of Trust-DKF algorithm

System model: The state equation and observation equation are as follows:

$$x_k = F_k x_{k-1} + w_k, z_k^i = H_k^i x_k + v_k^i \tag{19}$$

Among them, $w_k N(0, Q_k)$, v_k^i : $N(0, R_k^i)$, and the superscript i denotes the i -th sensor node.

Trust weight calculation: The local estimated value $\hat{x}_k^i$ is analyzed by K-means clustering to identify abnormal nodes. The clustering center $\bar{x}_k$ is determined by Euclidean distance minimization:

$$\bar{x}_k = arg\,\min_c \sum_i \left\| \hat{x}_k^i - c \right\|^2 \tag{20}$$

K-Means clustering was adopted for trust evaluation because of its low computational complexity and ability to efficiently group sensors with similar reliability characteristics. The number of clusters was selected based on experimental validation to achieve a balance between trust score stability and fault discrimination. An appropriate cluster configuration reduces the influence of unreliable sensor measurements while maintaining stable trust estimation, thereby improving the robustness of fault detection.

The trust τ_i of a node i is based on its degree of deviation from the center:

$$\tau_i = exp\left(-\frac{\left\|\hat{x}_k^i - \bar{x}_k\right\|^2}{2\sigma^2}\right) \tag{21}$$

The Trust-DKF framework first computes a trust score for each sensor node based on its deviation from the cluster center. Nodes with trust values below the predefined threshold are identified as abnormal and excluded from the estimation process, while trusted nodes perform distributed Kalman filtering to exchange local estimates and collaboratively generate a robust global state estimate.

The scale parameters σ are adaptively adjusted according to the clustering radius.

Pixel-level fusion: It is suitable for homogeneous sensor data. A lower threshold improves fault sensitivity but may include unreliable measurements, whereas a higher threshold enhances reliability by filtering uncertain nodes but may reduce early fault detection. The threshold was empirically selected based on validation performance to balance detection accuracy and estimation robustness.

Fusion update: Only fusion estimates of nodes whose trust is higher than the threshold θ:

$$\hat{x}_k = \frac{\sum_{i:\tau_i>\theta} \tau_i \hat{x}_k^i}{\sum_i \tau_i}, P_k^{-1} = \sum_{i:\tau_i>\theta} \tau_i \left(P_k^i\right)^{-1} \tag{22}$$

The prediction and update steps follow the standard KF framework, but the covariance matrix is weighted and fused to enhance robustness.

The Trust-DKF algorithm achieves stable convergence through iterative trust-weighted state updates. Although trust evaluation introduces additional computations, the computational overhead remains low, making the algorithm suitable for real-time edge deployment

4. System comprehensive performance verification

The performance verification of the intelligent inspection system for 500 kV substations is a crucial step to ensure its practical engineering application. This chapter comprehensively evaluates the comprehensive performance of the proposed algorithm model in a real substation environment by designing a systematic and multi-dimensional test plan. The test content covers traditional evaluation dimensions such as basic performance indicators, robustness, practicality, and interpretability, while innovatively introducing industrial-level evaluation indicators such as cross-scenario generalization, real-time pressure, energy efficiency, and long-term stability, forming a complete performance verification system. All tests are conducted based on a combination of real datasets and simulation platforms to ensure the reliability and reproducibility of the evaluation results.

4.1. Test method

(1) Dataset construction and preprocessing

The experimental data is sourced from three international public datasets, ensuring diversity and representativeness of the data. The FLIR Thermal Dataset provides infrared thermal imaging data from substations in

North America, Europe, and Asia, containing 15,000 labeled images with temperature ranges spanning from -40°C to 550°C, comprehensively documenting the heating patterns of key equipment such as transformers and circuit breakers. The MIMII Dataset includes abnormal sound samples from industrial equipment in Japan and Germany, with a sampling rate of 16kHz and covering three states: normal, slightly abnormal, and severely abnormal, making it suitable for the extraction and analysis of acoustic features. The NASA Bearing Dataset provides vibration signals collected by NASA laboratories in the United States, recording the full lifecycle data of bearings from normal to faulty conditions, with a sampling frequency of 20kHz, offering a rich sample for vibration analysis.

The multimodal dataset was constructed by integrating infrared images, visible-light images, acoustic signals, and vibration measurements obtained from the selected datasets. The data from different modalities were temporally and spatially aligned, and each sample was labeled according to the corresponding equipment condition before preprocessing and model training.

The dataset was organized by modality into infrared images, visible-light images, acoustic signals, and vibration signals. Samples from each modality were divided into training, validation, and testing sets using a consistent partitioning strategy. Images were resized and normalized, acoustic signals were converted into Mel-spectrograms, and vibration signals were denoised, transformed to the frequency domain using FFT, and normalized. To reduce class imbalance, balanced sampling was applied during model training, ensuring comparable representation of normal and faulty equipment conditions.

In the data preprocessing stage, we employ a four-level processing flow to ensure data quality. Firstly, data cleaning is conducted, where outliers are identified based on sliding window detection and the Z-score algorithm (threshold ±3), and neighborhood interpolation is used for repair. Subsequently, Min-Max normalization is performed to map multi-source data to the [0,1] interval, eliminating the influence of dimensionality. For non-stationary data, an additional first-order differencing process is applied to enhance stationarity. In the denoising stage, wavelet soft thresholding (adaptive λ parameter) and Kalman filtering ($Q=0.01$, $R=0.1$) are combined to effectively suppress high-frequency noise and random interference. Finally, SIFT feature matching is used to achieve spatiotemporal registration of multi-source data, ensuring the synchronization and consistency of data from different sensors. The data is divided into a training set, a validation set, and a test set in a ratio of 7:1.5:1.5.

(2) Test environment and baseline setting

The experimental hardware platform is configured as a workstation equipped with an Intel i9-10900K processor, NVIDIA RTX 3090 graphics card, and 64GB of memory. The edge testing platform utilizes the NVIDIA Jetson Xavier NX embedded device. The software environment is based on Python 3.8 and PyTorch 1.9 frameworks, and integrates commonly used libraries such as OpenCV and Scikit-learn. The experiment involves three professionals in the field of power systems who participate in data annotation and result verification to ensure the objectivity of the evaluation.

Four representative baseline models were selected for comparison: YOLOv4 as an advanced model for visual detection, LSTM-AE for acoustic anomaly detection, 1D-CNN for vibration signal classification, and traditional Kalman filtering as a benchmark for state estimation. The experimental group employed the AMM-Net and Trust-DKF fusion model proposed in this paper, and all models were trained and tested under the same conditions.

(3) Experimental design and evaluation indicators

The experimental design adopts the method of controlled variables to systematically evaluate eight performance dimensions. The basic performance test uses a mixed dataset to test accuracy, precision, recall, and F1-Score; the robustness test evaluates anti-interference capability by adding Gaussian noise (10dB) and random occlusion (15%); the practicality test measures inference time and memory usage; the ablation test gradually removes the attention mechanism, multi-level fusion, and trust mechanism to analyze the contribution of each module.

In addition, four innovative extended experiments were designed: the cross-scenario generalization experiment tests the model's adaptability using European CIGRE, North American PJM, and Asian tropical datasets; the real-time pressure experiment simulates concurrent scenarios with 10-200 nodes to evaluate system throughput; the energy efficiency experiment measures single inference energy consumption on edge devices such as Jetson Xavier NX; and the long-term stability experiment monitors performance degradation during 30 days of continuous operation. Evaluation metrics cover industrial-grade indicators such as DAI index, APD, TPS, response time, and energy efficiency ratio to ensure comprehensive evaluation.

4.2. Test results

(1) Performance comparison test analysis

In order to comprehensively evaluate the basic performance of the model in equipment status recognition, this test randomly selects 2,000 samples from FLIR and MIMII datasets (normal, overheated and abnormal sounds accounted for 1/3 each), and compares the classification performance of five types of models under the same training conditions. The test is repeated 10 times and averaged to eliminate the effect of randomness, and the results are shown in Table 2.

Table 2. Performance comparison test results (%)

Models	**Accuracy rate**	**Accuracy rate**	**Recall rate**	**F1-Score**
YOLOv4	86.2±0.3	85.1±0.4	84.7±0.5	84.9±0.4
LSTM-AE	82.5±0.5	81.3±0.6	80.9±0.4	81.1±0.5
1D-CNN	84.8±0.4	83.6±0.3	83.2±0.6	83.4±0.4
Traditional Kalman filter	78.3±0.6	76.9±0.5	77.1±0.7	77.0±0.6
The model of this paper	94.7±0.2	93.5±0.3	92.8±0.3	93.1±0.2

The relative advantages and limitations of the proposed methodology, we benchmark it against recent state-of-the-art multimodal approaches that report comparable performance metrics.

Table 3. Benchmark comparison of the proposed method with recent multimodal intelligent inspection and smart grid monitoring approaches

Method	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
CLIP-based Cross-Modal Learning Li et al.,[21]	93.38	92.5	91.0	91.7

UHV Method 1 Meng et al., [22]	83.3	82.0	89.1	85.4
Proposed Method	94.7	93.5	92.8	93.1

Table 3 compares the proposed intelligent inspection framework with recent multimodal intelligent inspection and smart grid monitoring approaches. The proposed method achieves the highest accuracy, precision, and F1-score, demonstrating the effectiveness of integrating AMM-Net, Trust-DKF, and three-level multi-source data fusion for robust equipment condition assessment in 500 kV substations.

(2) Robustness test analysis

There are complex factors such as strong electromagnetic interference and equipment occlusion in the substation site. This test simulates the interference conditions in the actual environment, and adds Gaussian noise (signal-to-noise ratio 10dB) and random occlusion (occlusion area 15%) to the test data to evaluate the performance retention ability of each model under interference. The test measures the decrease of F1-Score relative to the non-interference benchmark. The smaller the decrease, the stronger the robustness. The results are shown in Table 4.

Table 4. Robustness test results (% F1-Score decrease)

Models	**Gaussian noise interference**	**Occlusion interference**	**Comprehensive interference**
YOLOv4	12.3±0.4	15.7±0.6	18.9±0.5
LSTM-AE	14.8±0.5	17.2±0.7	20.5±0.6
1D-CNN	11.6±0.3	14.1±0.5	17.3±0.4
Traditional Kalman filter	19.4±0.7	22.6±0.8	25.1±0.7
The model of this paper	6.2±0.2	8.5±0.3	10.7±0.3

Table 5 presents the robustness evaluation of the proposed framework under practical substation conditions. Although illumination variations, sensor degradation, environmental fluctuations, and electromagnetic interference reduce the recognition performance to varying degrees, the framework maintains stable fault detection capability. The results demonstrate that the integration of AMM-Net, Trust-DKF, and multi-source data fusion provides robust performance under realistic operating conditions, with electromagnetic interference having the greatest impact on system accuracy.

Table 5. Robustness evaluation of the proposed intelligent inspection framework under practical substation conditions

Condition	**Accuracy (%)**	**F1-Score (%)**
Normal (Baseline)	94.7	93.1
Low Illumination	93.2	91.7
Strong Illumination	93.6	92.2
Sensor Degradation	92.1	90.6
Environmental Fluctuation	93.4	91.9
Electromagnetic Interference	88.8	87.3

(3) Practicability test analysis

In order to evaluate the practical deployment feasibility of the models on edge devices, this test tests the inference efficiency of each model on the Jetson Xavier NX embedded platform. The tests are conducted using batches of 500 samples, and the average inference time (milliseconds/sample) and memory usage (MB) are recorded. The results are shown in Table 6. The test environment temperature is maintained at 25°C, and stable values are taken after one hour of continuous operation.

Table 6. Practicability test results

Models	**Inference time (ms/sample)**	**Memory footprint (MB)**
YOLOv4	45.2±1.2	1203±15
LSTM-AE	38.7±0.9	892±12
1D-CNN	25.4±0.7	756±10
Traditional Kalman filter	12.1±0.3	345±8
The model of this paper	28.9±0.5	918±11

(4) Ablation test analysis

In order to verify the contribution of each module, this test gradually removes the attention mechanism (w/o Attention), the multi-level fusion module (w/o Fusion) and the trust mechanism (w/o Trust) of Trust-DKF in AMM-Net, and tests the F1-Score change on the FLIR dataset, and the results are shown in Table 7. The test is set up with 5 repeated measurements and 95% confidence intervals are calculated.

Table 7. Ablation test results (F1-Score%)

Model variants	**F1-Score**	**Decline magnitude**
Complete model	93.1±0.2	-
w/o Attention	88.4±0.4	4.7
w/o Fusion	85.7±0.5	7.4
w/o Trust	89.2±0.3	3.9
w/o All	81.3±0.6	11.8

The effectiveness of the GRU feature fusion module in capturing temporal dependencies among multimodal features, a sensitivity analysis was conducted by varying the number of GRU hidden units while keeping all other network parameters unchanged.

Table 8. Sensitivity analysis of the GRU feature fusion module with different hidden unit configurations

GRU Configuration	**Hidden Units**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Without GRU	–	85.7	84.9	84.1	84.5
GRU	32	89.2	88.3	87.6	87.9
GRU	64	92.1	91	90.4	90.7
GRU	128	94.7	93.5	92.8	93.1
GRU	256	94.2	93	92.3	92.6

As shown in Table 8, incorporating the GRU feature fusion module significantly improves recognition performance by effectively capturing temporal dependencies among multimodal features. The best performance is achieved with 128 hidden units, while increasing the hidden units to 256 provides no further improvement, indicating that 128 hidden units offer an effective balance between temporal feature representation and computational efficiency.

(5) Cross-scenario generalization test analysis

In order to test the adaptability of the model in different geographical environments, this test is trained on the 500kV training set, and then directly tested on the 400kV European CIGRE dataset, the 345kV North American PJM dataset and the 230kV Asian tropical dataset without fine-tuning. The evaluation indicators include domain adaptation index DAI (new scene accuracy rate/original accuracy rate) and average accuracy decline rate APD. The results are shown in Table 9.

Table 9. Cross-scenario generalization test results

Test scenario	**Raw accuracy (%)**	**New scene accuracy (%)**	**DAI index**	**APD (%)**
400kV substation in Europe	94.7±0.2	89.3±0.4	0.87±0.02	5.4±0.3
North American 345kV substation	94.7±0.2	87.6±0.5	0.83±0.03	7.1±0.4
Asia 230kV Substation	94.7±0.2	85.2±0.6	0.79±0.04	9.5±0.5

(6) Real-time pressure test analysis

In order to evaluate the processing power of the system in high concurrency scenarios, this test builds a distributed test environment, simulates four pressure levels (10, 50, 100, and 200 nodes concurrency), and measures system throughput (TPS), average response time (ms) and CPU usage (%). The results are shown in Table 10. The test lasts for 24 hours and data are recorded every 5 minutes.

Table 10. Real-time pressure test results

Pressure rating	**Number of nodes**	**Throughput (TPS)**	**Average response time (ms)**	**CPU Usage (%)**
Level 1	10	1250±25	45.2±1.2	35.6±2.1
Level 2	50	1180±30	48.7±1.5	52.3±2.8
Level 3	100	956±35	62.4±2.1	78.9±3.5
Level 4	200	723±40	85.1±3.2	92.5±4.2

(7) Energy consumption efficiency test analysis

In order to guide the selection of edge devices, this test measures the single inference energy consumption (Joules) and inferences per Joule on three typical edge devices, and the results are shown in Table 11. The power consumption of the equipment is measured directly using a power meter and the ambient temperature is controlled at 25 ± 2 °C.

Table 11. Energy consumption efficiency test results

Edge devices	**Single inference energy consumption (J)**	**Energy efficiency ratio (inf/J)**	**Peak temperature (°C)**
Jetson Xavier NX	0.12±0.01	8.33±0.15	68.5±1.2
Intel NUC11	0.23±0.02	4.35±0.12	72.3±1.5

Raspberry Pi 4B	0.45±0.03	2.22±0.08	61.8±1.0

(8) Long-term stability test analysis

In order to verify the reliability of continuous operation of the system, this test is tested continuously for 30 days, and the changes of performance indicators are recorded every week. The results are shown in Table 12. The test simulates the real substation environment, including the temperature difference between day and night (20-35 °C) and random equipment start and stop.

Table 12. Long-term stability test results (30 days)

Time Period	Average accuracy (%)	False alarm rate (%)	Memory growth (KB/day)	Prediction consistency (%)
Week 1	94.5±0.3	2.1±0.2	12.3±1.5	95.6±0.4
Week 2	93.8±0.4	2.4±0.3	15.6±2.1	94.3±0.5
Week 3	92.6±0.5	3.2±0.4	18.9±2.8	92.7±0.6
Week 4	91.3±0.6	4.1±0.5	23.4±3.2	90.5±0.7

4.3. Analysis and Discussion

(1) Comprehensive performance analysis

The experimental results demonstrate that the intelligent inspection system proposed in this paper excels across all eight evaluation dimensions. In terms of basic performance, the model achieves an accuracy rate of 94.7% and an F1-Score of 93.1%, significantly outperforming various baseline models. This is primarily attributed to the multi-modal feature fusion strategy, which fully utilizes the complementarity of information. The attention mechanism of AMM-Net can adaptively weight heterogeneous features, effectively focusing on key information, while the trust mechanism of Trust-DKF suppresses interference from abnormal data through dynamic weight allocation.

In terms of robustness, the model's F1-Score only decreased by 10.7% under comprehensive interference, demonstrating strong anti-interference capability. This is attributed to the redundant design of the multi-level fusion strategy and the recognition ability of the trust mechanism for abnormal nodes. Especially under Gaussian noise interference, the performance degradation of the model (6.2%) is much lower than that of traditional methods, proving its practicality in complex electromagnetic environments.

(2) Analysis of practicality and scalability

The practicality test shows that the model's inference time on Jetson Xavier NX is 28.9ms, meeting the real-time requirement. Although the memory occupation (918MB) is higher than that of traditional Kalman filtering, it is still within an acceptable range. The energy efficiency test indicates that Jetson Xavier has the best energy-to-performance ratio (8.33 inf/J), making it an ideal choice for edge deployment.

Practical operation in 500 kV substations, infrared cameras, visible-light cameras, microphones, and vibration sensors continuously collect equipment condition data. The acquired data are synchronized through the proposed preprocessing and spatio-temporal registration workflow before being processed by AMM-Net for multimodal feature extraction and equipment status recognition. Trust-DKF then evaluates sensor reliability and performs distributed state estimation, while the decision-level fusion module generates the final fault diagnosis result to support timely maintenance and intelligent inspection.

The cross-scenario generalization test reveals the geographical adaptability characteristics of the model. In the

European scenario, the DAI index reaches 0.87, indicating good adaptability to similar climatic conditions. However, the performance degradation is more pronounced in the tropical Asian scenario (APD=9.5%), suggesting the need to enhance optimization for high-temperature and high-humidity environments. This finding provides a clear direction for subsequent model optimization.

The observed performance variations are mainly attributed to differences in climatic conditions, equipment characteristics, and operating practices across substation environments. These factors affect sensor data quality and feature distributions, influencing model generalization and highlighting the need for environment-specific adaptation to improve robustness.

(3) Long-term operational stability analysis

The results of the long-term stability test indicate that the model maintains good performance during continuous operation for 30 days, but there is a phenomenon of slow performance degradation. The accuracy rate decreases from 94.5% to 91.3%, the false alarm rate increases from 2.1% to 4.1%, and the memory increases at an average rate of 17.6KB per day. This degradation is mainly due to sensor drift and changes in feature distribution. It is recommended to introduce an online learning mechanism for dynamic correction.

Real-time stress testing has demonstrated that the system can maintain a throughput of 956 TPS even under a 100-node concurrent scenario, satisfying the inspection needs of most substations. However, when the number of nodes reaches 200, CPU utilization exceeds 90%, and response time increases significantly. It is recommended to adopt a load balancing strategy in actual deployment.

(4) Limitations and directions for improvement

Despite the model's excellent performance in multiple trials, there is still room for improvement. Firstly, its cross-scenario adaptability needs to be enhanced, especially its stability in extreme environments. Secondly, the performance degradation issue during long-term operation needs to be addressed through online learning mechanisms. Furthermore, the resource scheduling efficiency of the model in ultra-high concurrency scenarios needs to be optimized.

large-scale deployment in 500 kV substations, communication overhead and computational scalability should be considered as the number of sensor nodes increases. Although the proposed framework supports real-time edge inference, practical implementation may be constrained by network bandwidth, communication latency, and hardware resources. These challenges can be addressed through efficient communication protocols, distributed processing, and scalable resource scheduling strategies.

The follow-up work will focus on three directions: first, developing an environment self-adaptive module to enhance the robustness of the model under extreme conditions; second, designing a lightweight online learning algorithm to achieve dynamic updating of model parameters; third, optimizing resource scheduling strategies to enhance the system's processing capacity in high concurrency scenarios. These improvements will further enhance the engineering applicability of the intelligent inspection system in 500 kV substations.

5. Conclusion

Through systematic research, this paper successfully designs a 500 kV substation intelligent inspection system based on multi-source data fusion, and proposes an innovative AMM-Net multi-modal feature extraction network and Trust-DKF state estimation algorithm. Test verification shows that the system's equipment state recognition accuracy reaches 94.7%, F1-Score is 93.1%, and the anti-interference ability is improved by 40% in the robustness test. The practical test achieves 28.9 ms real-time inference performance on edge devices, and the energy consumption is as low as 0.12 J/time, which is significantly better than traditional methods. These studies confirm that multi-source data fusion technology can effectively integrate infrared, visible light, acoustic, and vibration data, improving inspection efficiency and reliability, and providing crucial technical support for the operation and maintenance of intelligent substations. However, the system has shortcomings in terms of adaptability to extreme environments and long-term operational stability. Therefore, future research will focus on the development of environment-adaptive modules, the integration of lightweight online learning algorithms, and the optimization of high-concurrency resource scheduling to further expand the boundaries of engineering applications.

Declaration

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding Statement: This research received no external funding.

Author Contribution: Yaoshan Zhang contributed to the conceptualization, methodology, system design, data analysis, and manuscript preparation. Zhiqiang Xiao contributed to system implementation, validation, and experimental analysis. Zhihua Lin participated in data acquisition, software optimization, and technical review. Xiuquan Hu assisted in field investigation, equipment testing, and resource management. Yue Zhou contributed to visualization, manuscript editing, and final approval of the submitted version.

Ethical Approval: Not applicable.

Consent to Participate: Not applicable.

Consent to Publication: All authors have reviewed the manuscript and consent to its publication.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code Availability: No code available.

Acknowledgement

The authors would like to thank Hainan Power Grid Co., Ltd. and Hainan Power Grid Co., Ltd. Construction Branch for providing technical support, infrastructure resources, and valuable assistance during the completion of this research.

References

[1] Li Z, Zhang Y, Wu H, Suzuki S, Namiki A, Wang W. Design and application of a UAV autonomous inspection system for high-voltage power transmission lines. Remote Sens. 2023;15(3):865-877.

[2] Chen DQ, Guo XH, Huang P, Li FH. Safety distance analysis of 500kv transmission line tower uav patrol inspection. IEEE Lett Electromagn Compat Pract Appl. 2020;2(4):124-128.

[3] Xu C, Li Q, Zhou Q, Zhang S, Yu D, Ma Y. Power line-guided automatic electric transmission line inspection system. IEEE Trans Instrum Meas. 2022;71:1-18.

[4] Gao H, Dekyi D, Metok M. Efficient unmanned aerial vehicle inspection and management of transmission lines in modern electric power enterprises. Energy Informatics. 2025;8(1):117-128.

[5] Zhang C, Zhang C, Cui H, Li P, Wan Y, Sun D, Ma H. Embodied intelligence-driven framework for UAV power inspection systems: Federated reinforcement learning for multi-agent collaborative optimization. In: International Conference on Intelligent Computing. Singapore: Springer Nature Singapore; 2025. p. 292-301.

[6] De Albuquerque JMN, Neto OR, Armênio GF, Correa VM, Do Valle DRG, Domingues A, De Oliveira AS. A Novel Method for Multi-Modal Predictive Inspection of Power Lines. IEEE Access. 2024.

[7] Rong S, He L, Du L, Li Z, Yu S. Intelligent detection of vegetation encroachment of power lines with advanced stereovision. IEEE Trans Power Deliv. 2020;36(6):3477-3485.

[8] Zhang Y, Li D, Chen Y, Liu G, Fan D, Li W, Mai X. Research on Electric Field Safety Distance of High-Voltage Equipment in Substations Inspected by UAV. In: International Symposium on Insulation and Discharge Computation for Power Equipment. Singapore: Springer Nature Singapore; 2023. p. 505-513.

[9] Xu L, Guo H, Fu Y, Wang Z, Li Z, Hua Z, Liu S. Research of 3D trajectory hovering points determination method of UAV

patrol inspection transmission tower. J Phys Conf Ser. 2022;2378(1):012088.
[10] Li X, Zhang J, Zhang L, Chen X. Layout Optimization and Multi-scenes Intelligent Inspection Scheme Design Based on Substation Video Monitoring. J Phys Conf Ser. 2023;2560(1):012022.
[11] Xuefeng N, Yao G. Design of intelligent operation inspection platform based on the multi-agent system for live line measurement of substation. J Phys Conf Ser. 2022;2401(1):012088.
[12] Wang S, Dong S, Lv K, Zhang L, Ling Z, Xi Y. Architecture research of novel digital intelligent substation system. In: 2023 4th International Conference on Computer Engineering and Application (ICCEA). IEEE; 2023. p. 229-233.
[13] Zhong L, Xie Z, Wang K, Ting L, Chen D. Research on robot for repairing broken strands in 500kV High Voltage Live Operation. J Phys Conf Ser. 2021;2030(1):012088.
[14] Liu Y, Fan S. An improved slam algorithm for substation inspection robots based on 3d lidar and visual information fusion. Energies. 2025;18(11):2797-2806.
[15] Lei T, Lv F, Liu J, Zhang L, Zhou T. Research on fault detection algorithm of electrical equipment based on neural network. Math Probl Eng. 2022;2022:9015796-9015806.
[16] Li C, Shi F, Liu M, Wang Y, Zhang W, Li C. An intelligent diagnosis method for analog of the secondary equipment based on the remote dispatching station. In: 2022 7th Asia Conference on Power and Electrical Engineering (ACPEE). IEEE; 2022. p. 2229-2233.
[17] Pei S, Sun H. Structural design and simulation study of intelligent defect elimination equipment for high-voltage transmission line pin defects. IET Gener Transm Distrib. 2023;17(24):5366-5377.
[18] Lei H, Haining X, Lingbin S, Qiang J, Qingye S. Design and Application of New Technology for Intelligent Inspection and Early Warning of Submarine Cables against External Breakage. In: 2024 4th Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS). IEEE; 2024. p. 696-701.
[19] Gan X, Geng X, Xiong Z, Wu Z, Du S, Gao Y, Guo Y. Application of 5G communication technology on intelligent inspection in 750kV substation. J Phys Conf Ser. 2021;1983(1):012089.
[20] Liu J, Zhao Z, Ji J, Hu M. Research and application of wireless sensor network technology in power transmission and distribution system. Intell Converg Netw. 2020;1(2):199-220.
[21] Li Z, Yoshie O, Wu H, Mai X, Yang Y, Qu X. Image analysis method of substation equipment status based on cross-modal learning. IEEJ Trans Electr Electron Eng. 2024;19(9):1507-1521.
[22] Meng R, Wang ZL, Zhao ZL, Li JP, Fu WP, Zhong S. A multimodal detection method for UHV substation faults based on robot inspection and deep learning. J Robot. 2022;2022.