ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on Intelligent Blocking and Disposal Technology of Power Metering Message Abnormalities Based on CLIP and Multimodal Fusion

Jun Chen[1], *Zhi Xu*[1,*], *Zhiyong Zhang*[1], *Zhenglei Zhou*[1] *and Litao Tang*[1]

[1] Guangxi Power Grid Co., Ltd, Nanning, Guangxi, 530024, China

SUMMARY: *Aiming at the problem of high false alarm rate and disposal lag caused by insufficient semantic alignment of multimodal features in power metering messages, a cross-modal fusion anomaly blocking method based on CLIP is proposed. The CLIP model jointly encodes text, timing signals and device images to construct a 512-dimensional unified semantic space, which is combined with GMM modeling for dynamic threshold determination to achieve accurate matching of anomalies. DQN is introduced to optimize the blocking strategy, and real-time disposal decision is generated by integrating grid topology and historical data. Experiments show that the method has a false alarm rate of ≤2.3%, a leakage rate of ≤1.8%, an average response time of 0.78 seconds, and improves the blocking efficiency by 8.8%, providing an intelligent guarantee for the security of the power system.*

KEYWORDS: *CLIP; GMM; DQN; power metering message*

1 Introduction

The power metering system is an important part of the electric power system, which undertakes key tasks such as electric energy measurement, electricity billing and power market transactions [1]. With the development of the power market economy, the advancement of new power systems, and the establishment of the national power company, the performance of the power metering system puts forward high requirements, and the real-time and accuracy of each metering point becomes particularly important [2-4]. However, there are obvious shortcomings in the power metering system message transmission link. The traditional manual meter reading method is inefficient, error-prone, difficult to avoid environmental interference, insufficiently efficient for the detection of abnormal telegrams, and the detection of real-time is missing [5, 6]. This single-modal approach, which is based on manual identification and monitoring of telegrams, makes it difficult to detect anomalies and solve them in a timely manner in the face of abnormalities such as equipment failures, power theft attacks, and communication attacks [7-9]. And these anomalies not only affect the normal operation of the power system and the accuracy of the data, leading to errors in power scheduling and load management, and increasing the risk and instability of power grid operation; they also disrupt the normal order of the power market, affecting the transparency and credibility of the power transaction, and causing direct economic losses [10-14].

A single modality usually cannot contain all the effective information needed to produce accurate results, and the multimodal fusion process combines information from two or more modalities to achieve information complementation and broaden the coverage of information contained in the input data to obtain more comprehensive, accurate, and valuable information

*xxuzhi@yeah.net
https://doi.org/10.65102/is2026029 Published: 30 April, 2026

[15-18]. In addition, the CLIP (Contrastive Language-Image Pre-training) model is a pioneering cross-modal model with a cleverly designed architecture that integrates visual and linguistic modalities, aiming to achieve efficient alignment and understanding between images and text [19]. CLIP is capable of aligning text and images efficiently in the feature space, enabling the model to accurately understand the relationship between text descriptions and image content, with strong cross-modal capabilities and efficient zero-sample learning, and can also be extended to a variety of application scenarios such as image classification, text generation, multimodal Q&A, etc., which is broadly applicable [20-23]. The development of CLIP model and multimodal fusion technology has become an important part of the power metering new direction of message anomaly blocking and processing technology.

In the context of smart technology empowerment, power metering anomaly detection and processing techniques have evolved from traditional rule-based approaches to machine learning and deep learning, and are now being explored in multimodal techniques. Literature [24] proposes a two-tier detection strategy for data tampering attacks in decentralized microgrid advanced metering infrastructures, where the first tier detects a threshold value that sets the residuals of the harmonic to arithmetic mean ratio safety range of the daily electricity consumption data, and the second tier observes the residuals exceeding the threshold continuously to determine the attack, which is real-time in nature. The rule-driven based model warns by setting thresholds, which is simple and feasible, but poorly adaptable [25]. Therefore, machine learning and deep learning algorithms are applied in the field of power metering anomaly detection in the continuous maturation of artificial intelligence technology.

Literature [26] pointed out that the support vector machine algorithm aids in the improvement of power metering anomaly detection accuracy by screening data, and introduces the SSD (Single Shot MultiBox Detector) algorithm, which can also diagnose anomalies and assess the condition of metering equipment. Literature [27] integrated Random Forest, Extra Tree Classifier, K-mean Clustering, and Predictive Maintenance modeling algorithms to identify and monitor the abnormal consumption of power theft anomalies and equipment anomalies in the power metering system, and achieved power theft detection and billing bias minimization. Literature [28] proposed a multi-model fusion anomaly detection method using machine learning algorithms for detecting anomalies in power metering collection data, and also shared a set of correction schemes for power metering anomalies. Literature [29] combined graph convolutional network and bi-directional long and short term memory network to detect anomalies in smart metering data with battery storage system and electric vehicle, identified outliers and missing values of smart metering and combined it with multi-objective optimization for power management. Literature [30] constructed the anomaly detection of distributed power metering system based on variational autoencoder and long and short-term memory network, which improved the effectiveness of massive and complex data, more accurately obtained the temporal dependence of the data, and provided a path for smart metering monitoring. Literature [31] used Transformer to create fault-tolerant anomaly detection algorithms, which can detect anomalies in power metering in complex environments and data loss during data transmission through three strategies: topological feature encoding, parallel sensing of spatio-temporal features, and autoregression. Although machine learning and deep learning can efficiently capture data features and improve the accuracy and real-time performance of anomaly detection, the machine learning model requires a large amount of data annotation, and the generalization ability of deep learning is poor and does not have interpretability, which makes it difficult to guarantee the accuracy and efficiency of the hidden anomaly detection and processing of metering messages in complex environments.

The cross-industry fusion innovation of power data has made multimodal technology shine in the power field. Literature [32] utilized three types of data, namely, equipment operation,

environment, and historical records, fused with the help of multimodal data fusion technology, and evaluated the abnormal handling of power metering equipment with the support of hierarchical analysis method and Transformer network. Literature [33] obtained 97.6% accuracy in power information anomaly detection by fusing two modalities of power data in time and frequency domains and extracting the time and frequency domain features of the data, which effectively maintains the power system stability. Literature [34] effectively recognized household appliances based on power consumption data from smart meters using visual converters and multimodal data fusion, and the recognition performance remained good across households and datasets, providing a reference for power metering anomaly detection. Literature [35] proposed robust CLIP models for targeted data tampering and backdoor attacks, which are able to reduce the success rate of data tampering and backdoor attacks to 12.5% and 0%, and can be used to address cyber-attacks on power grids and tampering in power metering message transmission. Literature [36] developed a sample less anomaly detection and localization framework integrating CLIP-based discriminative and self-supervised architecture, feature adapter, lightweight anomaly feature generator, binary anomaly discriminator, and CLIP-based self-supervised learning module, which successfully achieves 94.58% precision and recall. Literature [37] analyzed the effectiveness of the CLIP model in the application of automatic classification of aerial images of power line infrastructures, which achieved more than 96% accuracy in detecting power lines and was able to use the classification results for anomaly detection. These studies verified the application value of multimodal technology in the field of electric power, but did not explore for power metering message blocking and processing.

In order to ensure the safe and stable operation of the power system and improve the intelligent level of metering management. The study first realizes the feature fusion of multimodal data based on CLIP. The timing signal in the power metering message is converted into a two-dimensional image, and the feature fusion of text message and image message is realized based on visual encoder and text encoder. Then anomalous data is detected based on GMM model and the detection threshold is solved using EM algorithm. Finally, the grid topology and historical data are used to form an empirical pool, and a blocking strategy for anomalous information is designed using DQN.

2 CLIP-based cross-modal fusion of power metering messages

2.1 Power metering message characteristics

Based on the principles of electromagnetic induction and electronic measurement, power metering machinery converts electric parameters such as current and voltage in the circuit into measurable physical quantities to realize the accurate measurement of electric energy. In the AC circuit, the voltage and current are proportionally transformed by voltage transformers and current transformers, which are then converted into mechanical rotation or digital signals by the measuring mechanism to visually display the value of electric energy. Automation control technology in terms of data acquisition utilizes sensors to acquire electrical parameter data in real time. Data transmission relies on communication modules to transmit the collected data to the processing terminal through wired or wireless communication. Data processing uses microprocessors or computer systems to analyze, calculate and store data using specific algorithms. The control side is based on the results of the processing, the working state of the power metering machinery to adjust, to achieve intelligent operation.

In the traditional power grid, there are limitations in the application of automation control technology for power metering machinery. The massive message data generated by the power metering system has multimodal characteristics, including text-based messages, timing signals, equipment images and so on. Traditional anomaly detection methods usually deal with these modalities independently, or the semantic alignment of the fused multimodal features is insufficient, resulting in low transmission rates, poor stability, and unable to meet the needs of real-time monitoring and rapid control. In this regard, this paper proposes a multimodal fusion based on CLIP to pave the way for anomaly blocking and disposal techniques for power metering messages.

2.2 CLIP-based multimodal data fusion

2.2.1 CLIP model

The CLIP model, as a powerful multimodal macromodel, has a core structure consisting of a visual encoder and a text encoder. These 2 encoders are responsible for mapping image and text descriptions into a shared embedding space, respectively, thus realizing semantic alignment between image and text. During the training process, the CLIP model relies on a huge amount of image-text pairing data, which covers the exact correspondence between images and text descriptions [38]. The principle of the CLIP model is shown in Fig. 1. It mainly consists of 2 core components: text encoder and image encoder. The text encoder is responsible for converting the input text into a low-dimensional vector representation, and the image encoder is responsible for converting the input image into a low-dimensional vector representation in the same space as the text embedding. During the training process, the model tries to maximize the similarity of the positive samples while minimizing the similarity of the negative samples. This learning approach allows the model to learn a cross-modal generic feature representation that enables joint understanding of text and images.

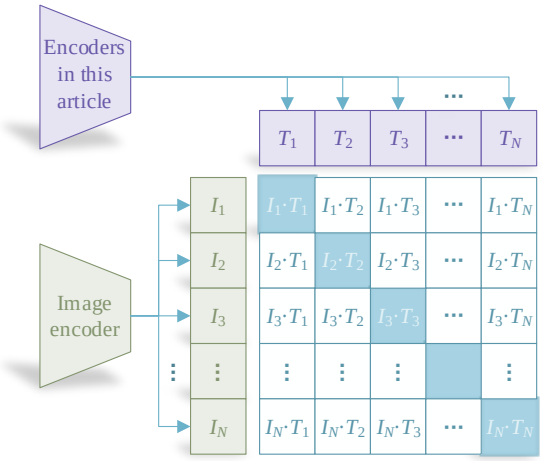

Figure 1: The principle of the CLIP model

2.2.2 2D image conversion method for time-series signals

Based on the previous analysis, the CLIP model is responsible for mapping image and text descriptions into a shared embedding space. Therefore for the time-series signals generated by power metering system, this paper proposes to use Short Time Fourier Transform (STFT) to convert one-dimensional signal data to two-dimensional images.

(1) Data set preprocessing and noise reduction

Firstly, all samples in the timing signal dataset generated by the power metering system are

subjected to a normalization operation. In this paper, $\mathbf{X}=\left(x_1,x_2,\ldots,x_n\right)$ is used to denote a certain sample in the power timing signal dataset, and $x_{\min}$ and $x_{\max}$ are used to denote the minima and maxima in the sample. Using $\mathbf{X}_{ts}=\left(x_{ts}^1,x_{ts}^2,x_{ts}^n\right)$ to denote the vector of samples after normalization to $[-1,1]$, then the normalization process can be expressed as equation (1):

$$x_{ts}^i=\left(\frac{x_i-x_{\min}}{x_{\max}-x_{\min}}\right) \tag{1}$$

After normalizing the raw power timing signal samples, they need to be fed into a high-pass filtering operation for processing. If $Y=\left(y_1,y_2,\ldots,y_n\right)$ is used to denote the output after high-pass filtering, it is calculated as shown below:

$$y_n=\alpha y_{n-1}+\alpha\left(x_{ts}^n-x_{ts}^{n-1}\right) \tag{2}$$

where $\alpha=\frac{1}{2\pi\Delta Tf_c},\Delta T$ is the time interval between two neighboring sampling points, and f_c denotes the cutoff frequency.

In order to remove the high-frequency background noise while still retaining some valuable peaks, this paper uses the discrete wavelet transform to perform noise reduction on the sample data after normalization and high-pass filtering. The main process of noise elimination by discrete wavelet transform is as follows:

1) According to the characteristics of the time series signal to be processed, select the appropriate mother and father wavelets and the number of decomposition layers to carry out wavelet decomposition of the discrete time series signal to be processed.

2) Calculate the approximation value coefficients and detail value coefficients corresponding to each layer based on the signal obtained from step 1 consisting of a low-frequency approximation portion and a high-frequency detail portion inversely.

3) Threshold the wavelet coefficients. There are three thresholding methods: hard thresholding, soft thresholding and soft-hard thresholding compromise.

4) Use the processed wavelet coefficients and combine with the signal obtained in step 1 to reconstruct a signal that is the signal after noise reduction.

The wavelet decomposition process in step 1 of the denoising process using discrete wavelet transform is shown in Fig. 2. Where $x[n]$ is used to denote the discrete power time-series signal to be processed of length N, $h[n]$ denotes the low-pass filter, $g[n]$ denotes the high-pass filter, and $\downarrow 2$ denotes the 2-fold downsampling.

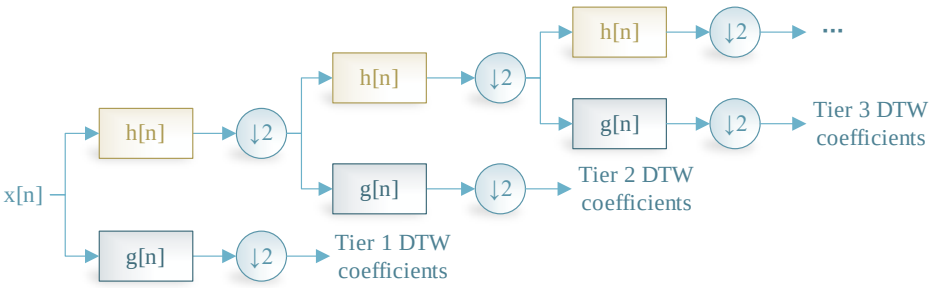

Figure 2: Wavelet transform decomposition

If $\varphi(t)$ is used to denote the parent wavelet, $\psi(t)$ is used to denote the mother wavelet, and c and d are used to denote the approximation value coefficients and detail value coefficients computed in step 2 of the denoising process of the discrete wavelet transform, respectively, then the discrete signals in the space can be written in the form as in equation (3):

$$x[n]=\sum_{k} c_{j_0,k}\varphi_{j_0,k}(n)+\sum_{j>j_0}\sum_{k} d_{jk}\psi j,k(n) \tag{3}$$

where k is used to control the degree of translation of the wavelet basis in the time dimension. The j denotes the scaling factor, which changes the frequency magnitude of the wavelet basis. The j_0 then denotes the basic scaling factor.

In this paper, the generalized inter-value rule is used for transform selection and combined with the inter-value processing function of soft and hard inter-value compromise for noise reduction. Where the transform selection formula is shown in equation (4):

$$threshold=\frac{median(|cD1|)}{0.6745}\sqrt{2\ln N} \tag{4}$$

where $cD1$ is the detail coefficient of the first layer decomposition and N is the length of the detail value vector of the first layer decomposition. The time-series electrical signal after denoising is performed has a high resolution. In order to reduce the dimensionality of the power time-series signal, downsampling can be performed using maximum pooling, which preserves those peaks that can demonstrate the presence of partial discharge phenomena.

(2) One-dimensional time series data to two-dimensional images

After the preprocessing in the previous section, the localized discharge signal may appear on any subsequence after the preprocessing. Therefore, information can be considered to be extracted from each subsequence of the time series signal after preprocessing. The short-time Fourier transform (STFT) [39] has been widely used in both speech recognition and enhancement, and is a versatile tool for speech signal processing. The discrete short-time Fourier transform first divides the entire power time-series signal into multiple overlapping time-series subsequences. The discrete Fourier transform is then performed using the product of the window function and the signal function on the subsequences. Here in this section, N_w is used to denote the window size of the time series subsequence, which is also the length of the subsequence: N_o is used to denote the length of the overlapping portion between the neighboring subsequences: $N_h=N_w-N_o$ is used to denote the number of sample points that are staggered between two neighboring windows. For the i th power time series signal subsequence, its short time Fourier transform $STFT(i,k)$ is expressed as equation (5):

$$STFT(i,k)=\sum_{m}^{N_w-1} x\left[iN_h+m\right]\gamma[m]e^{\frac{-j2\pi km}{N}},0\le k\le N-1 \tag{5}$$

where k denotes the frequency and γ denotes the window function with window size N_w.

After this, the logarithmic spectrogram of the processed time-series signal is computed, $spectrogram(i,k)=\log\left(|STFT(i,k)|^2\right)$, and the conversion of the one-dimensional time-series signal data into a two-dimensional image can be achieved.

2.2.3 Multimodal feature assimilation for semantic variability

For the same power target, after feature assimilation of multimodal features using the CLIP model, their angular difference in the joint representation space should be small because they have similar semantics. Based on this, in this paper, the cosine distance is used to guide the search of the joint representation space from the angular difference.

The search process of joint representation space is the search process of mapping function $f(s)$ and $g(i)$. Since the process of power target feature extraction with convolutional neural networks is a nonlinearized function fitting process, this paper combines feature extraction and spatial mapping to optimize the two types of features by constructing the cosine difference as shown in Eq. (6), so as to directly carry out feature assimilation in the process of feature extraction:

$$l_a = 1 - \cos\left(s', i'\right) \tag{6}$$

where: l_a is the feature assimilation loss. $\cos\left(s', i'\right)$ characterizes the cosine angle of the data pair s' and i'. s' is a 512-dimensional sensor data feature and i' is a 512-dimensional image feature. In the range of $[0, \pi]$, the larger the angular difference between the two, the smaller the cosine value, so in order to minimize the loss, the constructed loss function is shown in equation (6). After the optimization of the two types of features in Eq. (6), the angular difference between s' and i' is smaller, and it can be considered that after their feature assimilation, the two types of features are mapped to the joint representation space respectively and have the same feature description.

3 Anomaly detection and blocking technology

3.1 GMM-based dynamic threshold determination

3.1.1 GMM model

After the attacker injects anomalous information, the distribution of the observation vectors changes, especially when the different classes of observation samples in the first session have been projected to different regions of the new space with a high degree of differentiation, making this change even more obvious. Therefore, in this section, a probability distribution P is obtained by fitting a GMM model [40] for the observation samples $\mathbf{Z}_{nor}$ that have not been injected into the attack. For real-time observation samples, an observation is recognized as having been attacked if its probability worth output does not reach a predefined threshold.

The expression of GMM is shown in equation (7) below:

$$P(\mathbf{Z} \mid \Theta) = \sum_{k=1}^{K} \alpha_k p_k \left(\mathbf{Z} \mid \theta_k\right) \tag{7}$$

where $K > 0$ represents the number of Gaussian density functions. θ represents the parameters of the corresponding Gaussian function, including the mean vector and covariance matrix. The $\alpha \geq 0$ is a mixture of weights that satisfy the constraints $\sum_{k=1}^{K} \alpha_k = 1; \Theta = \left(\alpha_1, \alpha_K, \theta_1, \theta_K\right)$ is the set of parameters of the GMM. $p_k\left(\mathbf{Z} \mid \theta_k\right)$ is the Gaussian function in equation (8) below:

$$p_k\left(\mathbf{Z}\mid\theta_k\right)=\frac{1}{(2\pi)^{h/2}\left|\mathbf{\Sigma}_k\right|^{0.5}}\exp\left(-\frac{1}{2}\left(\mathbf{Z}-\boldsymbol{\mu}_k\right)^T\mathbf{\Sigma}_k^{-1}\left(\mathbf{Z}-\boldsymbol{\mu}_k\right)\right) \tag{8}$$

where $\boldsymbol{\mu}_k$ and $\mathbf{\Sigma}_k$ are the mean vector and covariance matrix of the k th Gaussian function, respectively. The h is the dimension of the observed sample after going through one stage.

The optimization problem of fitting the GMM using N samples $\mathbf{Z}_{nor}(q), q=1,2,N$ is shown in equation (9) below:

$$\Theta^*=\arg\max_{\Theta}\prod_{q=1}^{N}\sum_{k=1}^{K}\alpha_k p_k\left(\mathbf{Z}_{nor}(q)\mid\theta_k\right) \tag{9}$$

where N is the number of unattacked observation samples used to fit the GMM.

3.1.2 EM algorithm with semi-supervised threshold search

Since the optimization problem of Eq. (9) cannot be shown to be solved, in this paper, we use the EM algorithm to solve the optimal parameters Θ^* of the GMM.

Define the set of hidden variables Y, where $y(q)\in\{1,2,,K\}$. Here the hidden variable $y(q)$ represents the class of Gaussian functions to which the observation $\mathbf{Z}_{nor}(q)$ belongs. Define $Q(\cdot)$ as the conditional probability expectation function of $\left(\mathbf{Z}_{nor},Y\right)$, whose expression is shown in equation (10) below:

$$Q\left(\Theta,\Theta^{(g)}\right)=\int P\left(Y\mid\mathbf{Z}_{nor},\Theta^{(g)}\right)\ln P\left(\mathbf{Z}_{nor},Y\mid\Theta\right)dY \tag{10}$$

It can be shown that with the knowledge of the result of the g st iteration, if the result $\Theta^{(g+1)}$ of the $g+1$ nd iteration is equal to Θ of the maximized conditional probability expectation function, then the likelihood function of $\Theta^{(g+1)}$ is greater than the likelihood function of $\Theta^{(g)}$ at this point. It can be expressed as equation (11):

$$\Theta^{(g+1)}=\arg\max_{\Theta}Q\left(\Theta,\Theta^{(g)}\right)\Rightarrow\ln P\left(\mathbf{Z}_{nor}\mid\Theta^{(g+1)}\right)\geq\ln P\left(\mathbf{Z}_{nor}\mid\Theta^{(g)}\right) \tag{11}$$

Based on this, it is then possible to use an iterative approach until Θ converges. Based on the definition of the hidden variables above, two of the integral symbols in Eq. (10) can be written in the form of Eq. (12) and Eq. (13):

$$P(Y\mid\mathbf{Z},\Theta)=\prod_{q=1}^{Q}\frac{\alpha_{y(q)}p_{y(q)}\left(\mathbf{Z}_{nor}(q)\mid\theta_{y(q)}\right)}{\sum_{s=1}^{K}a_s p_s\left(\mathbf{Z}_{nor}(q)\mid\theta_s\right)} \tag{12}$$

$$P(\mathbf{Z},Y\mid\Theta)=\prod_{q=1}^{Q}\alpha_{y(q)}p_{y(q)}\left(\mathbf{Z}_{nor}(q)\mid\theta_{y(q)}\right) \tag{13}$$

The EM algorithm contains two main steps: the E-step and the M-step. By continuously looping these two steps, thus the parameters Θ of the GMM are finally obtained. In the E-

step, the expectation of the likelihood of ($\mathbf{Z}_{nor}, Y$) is estimated based on the parameter $\Theta^{(g)}$ of the current iteration. In the M-step, this is solved for $\Theta^{(g+1)}$ by maximizing the expectation in the E-step. The specific EM steps are shown in Eqs. (14) through (18) below:

E step:

$$Q\left(\Theta,\Theta^{(g)}\right)=\sum_{l=1}^{K}\sum_{q=1}^{N}\ln\left[\alpha_l p_l\left(\mathbf{Z}_{nor}(q)\,|\,\theta_l\right)\right]P\left(l\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right) \tag{14}$$

$$P\left(l=k\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right)=\frac{\alpha_k p_k\left(\mathbf{Z}_{nor}(q)\,|\,\theta_k\right)}{\sum_{s=1}^{K}\alpha_s p_s\left(\mathbf{Z}_{nor}(q)\,|\,\theta_s\right)} \tag{15}$$

M Step:

$$\alpha_l^{(g+1)}=\frac{1}{N}\sum_{q=1}^{N}P\left(l\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right) \tag{16}$$

$$\mu_l^{(g+1)}=\frac{\sum_{q=1}^{N}\mathbf{Z}_{nor}(q)P\left(l\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right)}{\sum_{q=1}^{N}P\left(l\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right)} \tag{17}$$

$$\sum_{l}^{(g+1)}=\frac{\sum_{q=1}^{N}\left[\mathbf{Z}_{nor}(q)-\mu_l^{(g+1)}\right]\left[\mathbf{Z}_{nor}(q)-\mu_l^{(g+1)}\right]^T P\left(l\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right)}{\sum_{q=1}^{N}P\left(l\,|\,\mathbf{Z}_{nor}(q),\Theta^{(g)}\right)} \tag{18}$$

After that, the detection threshold δ of the GMM detector needs to be determined. In this paper, F_1 is used as the evaluation index, and the optimal value of is searched using a semi-supervised algorithm. The expression of F_1 is shown in equation (19) below:

$$\begin{aligned} F_1 &= 2\,\mathrm{Pr}\,\mathrm{Re}/(\mathrm{Pr}+\mathrm{Re}) \\ \mathrm{Pr} &= tp/(tp+fp) \\ \mathrm{Re} &= tp/(tp+fn) \end{aligned} \tag{19}$$

where Re and Pr represent recall and precision, and tp, fp, fn represent true positives, false negatives, and false positives, respectively.

This semi-supervised threshold search method is divided into two steps. First, the observations are substituted into the GMM model derived earlier: $P\left(\mathbf{Z}_{nor}(n)\right)$. After that, each $P\left(\mathbf{Z}_{nor}(n)\right)$ is selected as a detection threshold in turn, i.e., $\delta=P\left(\mathbf{Z}_{nor}(n)\right)$. For each threshold, the F_1 value is computed according to equation (19) and compared with the current optimal F_1 value F_{1best}, and if $F_1>F_{1best}$, the current threshold is tentatively considered as the optimal threshold and the current F_1 value is tentatively considered as the optimal one, i.e.: $\delta_{best}=\delta, F_{1best}=F_1$. After the output $P\left(\mathbf{Z}_{nor}(n)\right)$ of all the observed samples in the GMM has acted as an overthreshold, δ_{best} is the detection threshold for the solution.

3.2 DQN-based blocking strategies

3.2.1 DQN environment design

The core of the reinforcement learning environment is the physics engine, the STEP( ) function, whose input is the action a, the output is the next moment state s′, and also includes the reward r for the current action, whether to terminate the training DONE, and the debug item INFO, which describes all the information about the interaction of the intelligent body with the environment. In this system, this function utilizes the kinematic and kinetic models of the intelligent body to calculate the state and immediate reward of the next step and determine whether the termination state is reached or not [41].

The step( ) function consists of 3 main parts:

(1) Obtaining the state of the system at the next moment

The state of the system mainly consists of the location of the anomaly information, which is modeled in the power metering system, starting from the boundary of the guard region.

(2) Obtaining rewards for actions

In the experiment, the reward obtained when the intelligent body blocks the abnormal information is set as follows: when the abnormal information is within the blocking range, if blocking is performed, the reward is 1, otherwise -1. Similarly, when the abnormal information is outside the blocking range, if blocking is performed, the reward is -1, otherwise 1.

(3) Obtain the termination signal of training

The system ends the training with done=true when the anomaly information is successfully blocked. Otherwise, done=false, indicating that the system continues training until the termination condition is reached.

3.2.2 DQN decision-making process

The inputs to the learning( ) function of DQN are (s_t, a_t, r_t, s_{t+1}), which are the current state, the current action, the current reward and the next moment state, respectively. In most cases, the future state obtained by the intelligent body is closely related to the reward of the current state, which needs to be taken into account in the learning process. However, in the process of blocking anomalous information, they are not so closely connected instead, and the quaternion makes the learning space expand nearly 320 times at maximum, which slows down the convergence speed, and at the same time increases the difficulty of solving and reduces the quality of the solution. In addition, the Q-value updating has been changed by the process of:

$$Q_e(s_t, a_t) = Q_e(s_t, a_t) + \alpha\left[r + \gamma \max_{a_{t+1}} Q_t(s_t, a_t) - Q_e(s_t, a_t)\right] \tag{20}$$

The update formula is basically the same as DQN, except that the state of the next moment is not considered, i.e., $Q_t(s_{t+1}, a_{t+1})$ is changed to $Q_t(s_t, a_t)$, and the Q value of the state of the next moment is not considered.

In this regard, this paper synthesizes the grid topology and historical data to generate real-time disposition decisions in the following steps:

(1) The initialization of the improved DQN algorithm is divided into three parts: initialization of the Q network $Q_e(s,a)$ parameter ϕ and the target network $Q_t(s,a)$ parameter ϕ', which are the same as the parameter ϕ and the parameter ϕ' in the first training. In this paper, the grid topology and historical data are used to form an empirical pool (D), which is set as a 3-tuple model, and the capacity n needs to be sized according to the

specific problem.

(2) Complete the training and optimization of the intelligences.

(3) Initialize the state s and select the starting point for training, the state s is randomly selected from the boundary coordinates of the guard zone.

(4) Select the action a and the intelligent body starts interacting with the environment.

(5) The environment makes a change based on the action a, the state s changes, the done value changes, and the environment gives the intelligent body a reward r.

(6) Place (s,a,r) into the experience pool D.

(7) Check if the experience pool D is stored, if it is then learn, otherwise continue to store. After deciding to learn from the experience pool D, test whether the termination condition is reached or not, if not, the Q value is updated according to equation (20).

(8) The following equation is used as a loss function to update the network:

$$\left(r+\gamma \max_{a_{t+1}} Q_t\left(s_{t+1},a_{t+1}\right)-Q_e\left(s_t,a_t\right)\right)^2 \tag{21}$$

(9) An update of the state s is performed, followed by an update ϕ' every 100 steps.

(10) When the termination condition is reached, the neural network optimized by the algorithm at this point is the network $Q_e(s,a)$ that needs to be output.

4 Analysis of results

4.1 Simulation experiment design

In order to verify the effectiveness of the proposed method, this paper generates the network topology based on NS2 network simulation software, and selects the National Grid metering message database as the dataset to generate the background traffic, while simulating the generation of offensive anomalous traffic. Considering that the detection algorithm is expected to obtain a high accuracy rate, low false alarm rate and leakage rate, and because the algorithm will be deployed on the network backbone routers, in order to reduce the additional overhead caused by detection, after repeated comparison experiments, the final anomaly detection cycle length is determined to be 25s, and the threshold of anomaly discriminant value is 1.4.

In this paper, we will define three parameters of accuracy rate, false alarm rate and missed alarm rate to evaluate the detection algorithm.

The accuracy rate is calculated as follows: the number of anomaly detection cycles for all detected anomalies due to aggressive behavior ÷ the number of anomaly detection cycles for all detected anomalies with aggressive anomalies.

False alarm rate is calculated as follows: number of anomaly detection cycles for all detected anomalies that are not due to aggressive behavior ÷ number of anomaly detection cycles for all detected anomalies that have aggressive anomalies.

The underreporting rate is calculated as follows: the number of anomaly detection cycles in which an offensive behavior exists but an offensive anomaly is not detected ÷ the number of anomaly detection cycles in which all offensive behaviors exist.

4.2 Adaptation of STFT algorithm in multimodal networks

In order to verify the superiority of the STFT algorithm, it is compared with three common transformation algorithms, namely, Gram's Angle Field (GAF), Markov Transfer Field (MTF), and Recurrence Plot (RP), in the comparison experiment. The training accuracy variations are

shown in Fig. 3. The RP algorithm converges faster than GAF and MTF, and there is a large jump in accuracy for GAF within 8-30 rounds and for MTF within 10-20 rounds. Both algorithms need to increase the number of training rounds to improve the performance of the model RP after 20 rounds the accuracy has stabilized, indicating that the RP algorithm has a stronger performance than GAF and MTF, while STFT converges faster than RP, and the accuracy change amplitude in the middle and late stages of training is more stable than that of the original algorithm, which suggests that STFT algorithm is more suitable for multimodal networks.

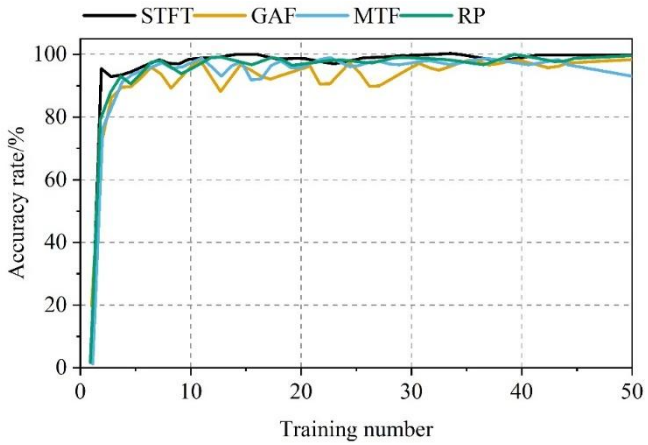

Figure 3: Training accuracy of different algorithms

4.3 Performance analysis of DQN algorithm

Fixing the learning rate to 0.005, the convergence performance of the DQN algorithm and the DQN algorithm with the addition of grid topology and historical data is shown in Fig. 4. The horizontal axis of the figure indicates the number of rounds, and the vertical axis indicates the average extrinsic reward accumulated in each round. The intrinsic rewards are not included in the comparison of the algorithms since they are only used to improve the intelligence's own exploration ability. From the figure, it can be seen that the proposed algorithm reaches convergence around 25 rounds, while the DQN algorithm reaches convergence only around 97 rounds, and the volatility of the DQN algorithm is significantly higher than the volatility of the proposed algorithm.The overall average rewards of the DQN algorithm are lower than the proposed algorithm reward values.

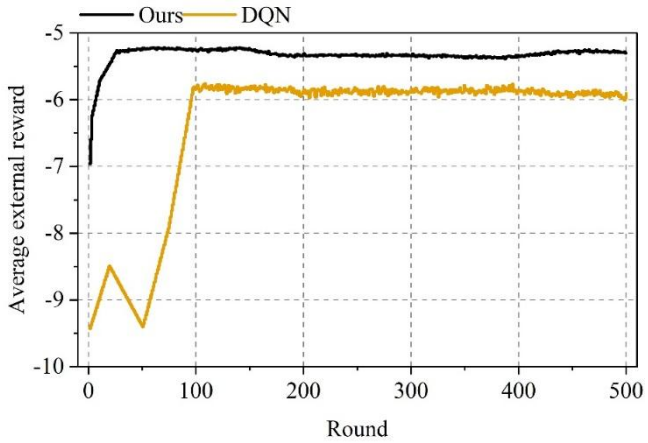

Figure 4: The algorithm is compared with the DQN algorithm

4.4 Analysis of anomaly detection and blocking results

Taking the National Grid Metering Message Database dataset as the background traffic without any additional offensive anomaly traffic, the number of text, timing signal and device image message segments are counted in the anomaly detection cycle of 8 phases, and the obtained results are shown in Fig. 5. Where the horizontal axis represents the time in seconds and the vertical axis represents the number of message segments. From the figure, it can be seen that among the grid metering messages, the text message has the most data, followed by the timing signal message, and the device image message has the least data.

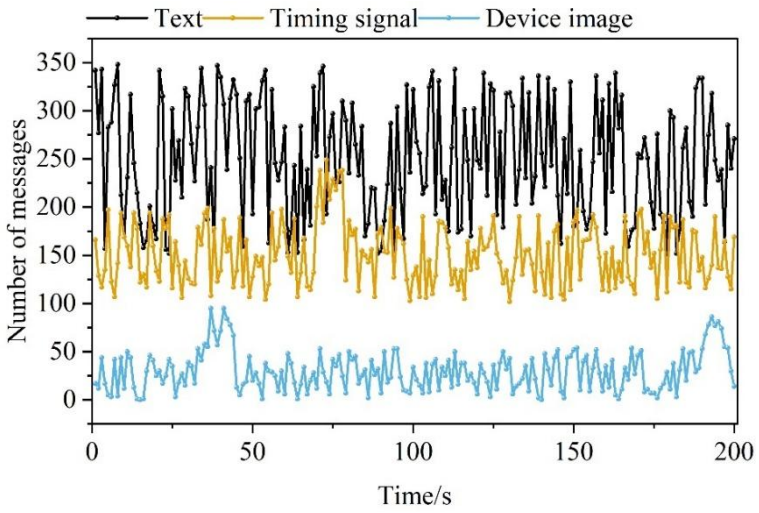

Figure 5: The distribution of the message segment under normal network traffic

Aggressive anomalous traffic is added to the network topology, at which point the number of text, timing signal and device image message segments is shown in Figure 6. The text message data is the most affected, and in phases 5 and 6, the data fluctuates dramatically, differing from the original data volume average by about 650. Timing signals and the number of equipment image message segments are less affected.

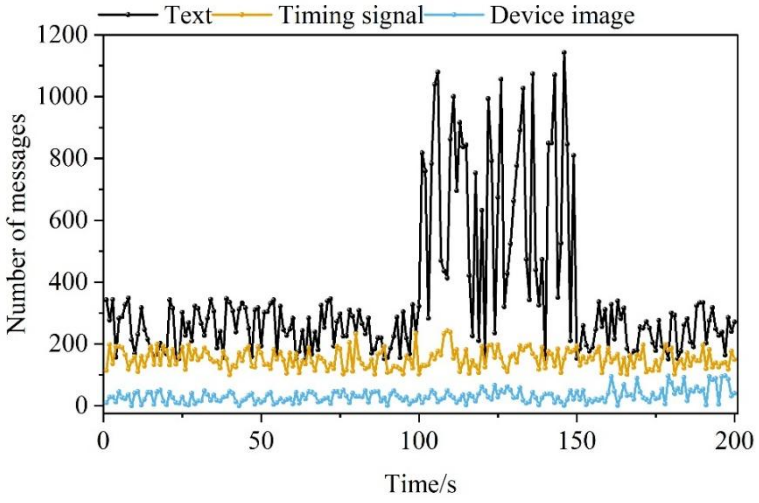

Figure 6: Abnormal message data

The cross-modal fusion of the three message data based on CLIP, anomaly detection through GMM, and the use of improved DQN blocking strategy to combat anomalous traffic, at this

time the number of the three message segments compared with the number of messages in the normal database is shown in Figure 7.

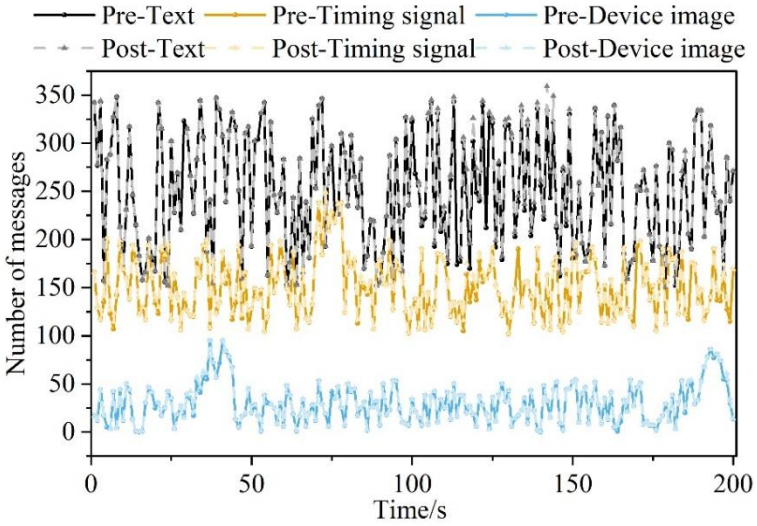

Figure 7: The comparison of the number of packets of three packets

Specific data analysis results are shown in Table 1. The detection accuracy of this paper's method is above 97.7%, and the false alarm rate is below 2.3%, the leakage rate affects the false alarm rate, and the leakage rate is the largest in the fifth stage, which is 1.8%. The average response time for each stage is 0.78 seconds.

Table 1: Simulation test results analysis

Detection cycle	Accuracy rate	False rate	Leakage	Response time
1s~25s	98.2%	1.8%	0.6%	0.66s
26s~50s	99.5%	0.5%	0.1%	0.63s
51s~75s	98.1%	1.9%	0.3%	0.89s
76s~100s	98.6%	1.4%	0.4%	0.93s
101s~125s	97.7%	2.3%	1.8%	0.82s
126s~150s	97.9%	2.1%	1.4%	0.84s
151s~175s	99.6%	0.4%	0%	0.86s
176s~200s	98.4%	1.6%	0.5%	0.62s

The improved DQN blocking strategy is used to combat abnormal traffic, and the failure rate is adopted as the method performance evaluation index to record the failure rate of the power metering system at each stage after fault blocking. The DQN method is selected for comparison with the improved method in this paper, and the specific experimental results are shown in Fig. 8. Compared with the DQN method, the fault rate of each stage is reduced by 8.8% on average, indicating that the blocking efficiency is improved by 8.8%.

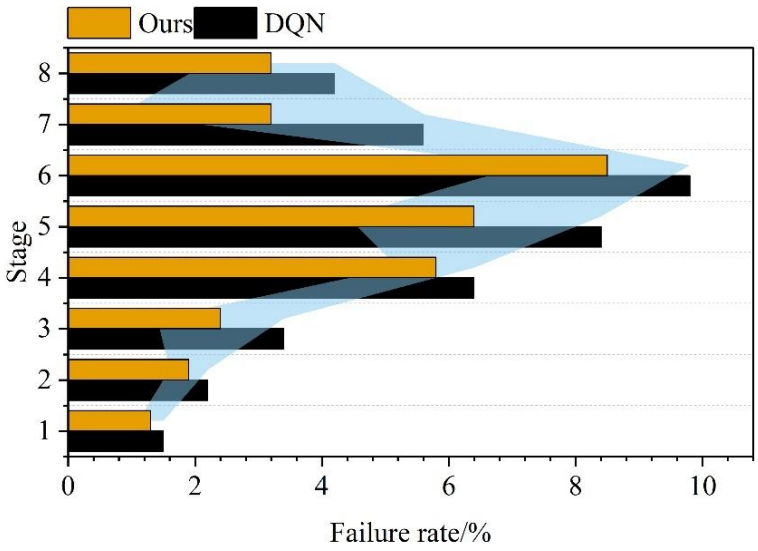

Figure 8: The DQN method and this method of the failure rate

In summary, the power metering message anomaly detection and blocking method adopted in this paper can meet the intelligent control standard of power metering system.

5 Conclusion

The study is based on CLIP model for feature fusion of text, timing signal and device image data in power metering telegrams, and the detection and blocking of abnormal information based on GMM and DQN. Through simulation tests, the detection accuracy of this paper's method for the three kinds of message data is ≥97.7%, and the average response time of each stage reaches 0.78 seconds. The real-time disposition decision generated by synthesizing the grid topology and historical data effectively improves the anomaly information blocking rate by 8.8% and reduces the occurrence of power metering system fault rate. The research method in this paper realizes the accurate identification and fast blocking of anomalous information, which provides a new idea for the security protection of power metering system.

About the Author

Jun Chen graduated from Wuhan University of Science and Technology in 2011. He is currently working in the Metering Center of Guangxi Power Grid Company. His main research fields are electric energy metering, metering automation and network security.

Zhi Xu graduated from Guangxi University in 2017 and is currently working at the Metering Center of Guangxi Power Grid Company. His main research field is metering automation.

Zhiyong Zhang graduated from Guangxi University in 2016 and is currently working at the Metering Center of Guangxi Power Grid Company. His main research areas are metering automation and network security.

Zhenglei Zhou graduated from Guangxi University in 2018 and is currently working in the Metering Center of Guangxi Power Grid Co., LTD. His main research areas are electric energy metering, metering automation and network security.

Litao Tang graduated from Guangxi University in 2010 and is currently working at the Metering Center of Guangxi Power Grid Company. His main research fields are metering automation technology research and network security management.

References

[1] Dahunsi, F. M., Olakunle, O. R., & Melodi, A. O. (2021). Evolution of electricity metering technologies in Nigeria. Nigerian Journal of Technological Development, 18(2), 152-165.

[2] Garcia, F. D., Marafão, F. P., de Souza, W. A., & da Silva, L. C. P. (2017, March). Power metering: History and future trends. In 2017 Ninth Annual IEEE Green Technologies Conference (GreenTech) (pp. 26-33). IEEE.

[3] Shuaibu, A. S., Haq, S. U., Almadani, B., Aliyu, F., & Al-Nahari, E. (2025). Smart Metering Meets AI: Real-Time Appliance Monitoring and Anomaly Detection over DDS Middleware. IEEE Transactions on Industry Applications.

[4] Kochański, M., Korczak, K., & Skoczkowski, T. (2020). Technology innovation system analysis of electricity smart metering in the European Union. Energies, 13(4), 916.

[5] Saavedra, E., Del Campo, G., & Santamaria, A. (2020). Smart metering for challenging scenarios: A low-cost, self-powered and non-intrusive IoT device. Sensors, 20(24), 7133.

[6] Kumar, P., Lin, Y., Bai, G., Paverd, A., Dong, J. S., & Martin, A. (2019). Smart grid metering networks: A survey on security, privacy and open research issues. IEEE Communications Surveys & Tutorials, 21(3), 2886-2927.

[7] In'kov, Y. M., Rozenberg, E. N., & Maron, A. I. (2020). Simulation of the process of implementation of an intelligent electric power metering system. Russian Electrical Engineering, 91(1), 65-68.

[8] Yang, B., Liu, S., Gaterell, M., & Wang, Y. (2019). Smart metering and systems for low-energy households: challenges, issues and benefits. Advances in Building Energy Research, 13(1), 80-100.

[9] Peter, J. S. P., Babu, C. R., & Esther, B. P. (2025). Cybersecurity in ICT-Enabled Smart Metering Systems: Addressing Challenges and Implementing Solutions. Cloud Computing in Smart Energy Meter Management, 263-290.

[10] Xiong, S., Zhang, J., Zhang, B., Sun, G., Chen, Z., Qi, J., & Sun, Y. (2021). Effects of environmental and electrical factors on metering error and consistency of smart electricity meters. Applied Sciences, 11(23), 11457.

[11] Apse-Apsitis, P., Vitols, K., Grinfogels, E., Senfelds, A., & Avotins, A. (2018). Electricity meter sensitivity and precision measurements and research on influencing factors for the meter measurements. IEEE Electromagnetic Compatibility Magazine, 7(2), 48-52.

[12] Jacoba, G. L. B., & Genove, G. P. C. (2023). Cybersecurity of smart grids: Attacks and defenses of the smart meters in an advanced metering infrastructure (ami). Southeast Asian Journal of Science and Technology, 8(1), 74-83.

[13] Vlasa, I., Gligor, A., Dumitru, C. D., & Iantovics, L. B. (2020). Smart metering systems optimization for non-technical losses reduction and consumption recording operation improvement in electricity sector. Sensors, 20(10), 2947.

[14] Khalid, Z., Kazmi, S. A. A., Hassan, M., Ali Shah, S. A., Anwar, M., Yousif, M., & Tariq, A. H. (2025). Socio-Economic Analysis for Adoption of Smart Metering System in SAARC Region: Current Challenges and Future Perspectives. Sustainability, 17(15), 6786.

[15] Ma, X., Zhang, X., Pun, M. O., & Liu, M. (2024). A multilevel multimodal fusion transformer for remote sensing semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-15.

[16] Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. ACM computing surveys, 56(9), 1-36.

[17] Qu, Z., Li, Y., & Tiwari, P. (2023). QNMF: A quantum neural network based multimodal fusion system for intelligent diagnosis. Information Fusion, 100, 101913.

[18] Zhang, C., Yang, Z., He, X., & Deng, L. (2020). Multimodal intelligence: Representation learning, information fusion, and applications. IEEE Journal of Selected Topics in Signal Processing, 14(3), 478-493.

[19] Lee, J., Kim, J., Shon, H., Kim, B., Kim, S. H., Lee, H., & Kim, J. (2022). Uniclip: Unified framework for contrastive language-image pre-training. Advances in Neural Information Processing Systems, 35, 1008-1019.

[20] Tu, W., Deng, W., & Gedeon, T. (2023). A closer look at the robustness of contrastive language-image pre-training (clip). Advances in Neural Information Processing Systems, 36, 13678-13691.

[21] Wang, S., Cheng, N., & Hu, Y. (2025). Comprehensive environmental monitoring system for industrial and mining enterprises using multimodal deep learning and clip model. IEEE Access.

[22] Deng, X., Shi, H., Huang, R., Li, C., Xu, H., Han, J., ... & Liang, X. (2023). GrowCLIP: Data-aware automatic model growing for large-scale contrastive language-image pre-training. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 22178-22189).

[23] Losada, A., & Bernardos, A. M. (2023, August). Image Classification Using Contrastive Language-Image Pre-training: Application. In 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023): Salamanca, Spain, September 5–7, 2023, Proceedings, Volume 2 (Vol. 750, p. 13). Springer Nature.

[24] Bhattacharjee, S., & Das, S. K. (2018). Detection and forensics against stealthy data falsification in smart metering infrastructure. IEEE Transactions on Dependable and Secure Computing, 18(1), 356-371.

[25] Jain, S., Choksi, K. A., & Pindoriya, N. M. (2019). Rule-based classification of energy theft and anomalies in consumers load demand profile. IET Smart Grid, 2(4), 612-624.

[26] Jiang, C., Wang, J., Wang, Y., & Zhao, W. (2023). Anomaly Diagnosis Method and Condition Assessment of Power Metering Device Based on SSD Algorithm. Scalable

Computing: Practice and Experience, 24(4), 1177-1184.

[27] Venkatakrishnan, G. R., Rengaraj, R., Viswavardinii, S., & NM, V. (2025, April). Anomaly Detection in Smart Metering: Clustering-Based Identification of Energy Theft. In 2025 International Conference on Computing and Communication Technologies (ICCCT) (pp. 1-4). IEEE.

[28] Sida, Z., Meiying, Z., & Ying, L. (2025). Research on anomaly detection and correction of power metering data based on machine learning algorithm. Science and Technology for Energy Transition, 80, 6.

[29] Lee, S., Nengroo, S. H., Jin, H., Doh, Y., Lee, C., Heo, T., & Har, D. (2023). Anomaly detection of smart metering system for power management with battery storage system/electric vehicle. ETRI Journal, 45(4), 650-665.

[30] Liu, K., Jia, X., Wang, J., & Ma, X. (2025). Real-Time Monitoring and Simulation of Multi-User Electricity Metering Anomaly Data Based on Distributed System. IEEE Access.

[31] Guo, Q., Shi, Y., Zhou, S., Xie, C., & Li, X. (2024). Missing-Tolerant Anomaly Detection for Gateway Electrical Energy Metering Device Based on Improved Transformer. IEEE Transactions on Instrumentation and Measurement, 73, 1-10.

[32] Chen, L., Zheng, X., Liu, Y., & Zheng, H. (2024, December). Abnormal Handling Evaluation Model of Electric Energy Metering Device Based on Multi-Source Information Fusion. In 2024 IEEE 7th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (pp. 168-174). IEEE.

[33] Chen, L., Zhou, X., Zhou, P., Sun, X., & Zheng, S. (2025). Anomaly detection method for power system information based on multimodal data. PeerJ Computer Science, 11, e2976.

[34] Ayub, M., & El-Alfy, E. S. M. (2025). Household Appliance Identification Using Vision Transformers and Multimodal Data Fusion. IEEE Transactions on Consumer Electronics.

[35] Yang, W., Gao, J., & Mirzasoleiman, B. (2023). Robust contrastive language-image pretraining against data poisoning and backdoor attacks. Advances in Neural Information Processing Systems, 36, 10678-10691.

[36] Zeng, S., Chen, Y., Li, M. N., Wu, Y., & Tian, J. (2025, July). CLIP-DSA: A CLIP-Based Discriminative and Self-supervised Framework for Few-Shot Anomaly Detection. In International Conference on Intelligent Computing (pp. 27-40). Singapore: Springer Nature Singapore.

[37] Losada, A., Bernardos, A. M., & Besada, J. (2023, August). Image Classification Using Contrastive Language-Image Pre-training: Application to Aerial Views of Power Line Infrastructures. In International Conference on Soft Computing Models in Industrial and Environmental Applications (pp. 13-23). Cham: Springer Nature Switzerland.

[38] Jiayu Zhang,Qingji Guan,Junbo Liu,Yaping Huang & Jianyong Guo. (2025). Railway-CLIP: A multimodal model for abnormal object detection in high-speed railway. High-

speed Railway,3(3),194-204. https://doi.org/10.1016/J.HSPR.2025.06.001.

[39] Mustafa Demetgul, Yicheng Zhao, Minjie Gu, Jonas Hillenbrand & Jürgen Fleischer. (2022). Motor Current Based Misalignment Diagnosis on Linear Axes with Short- Time Fourier Transform (STFT). Procedia CIRP,106,239-243. https://doi.org/10.1016/J.PROCIR.2022.02.185.

[40] Subhan Ullah, Pervaiz Akhtar & Ghasem Zaefarian. (2018). Dealing with endogeneity bias: The generalized method of moments (GMM) for panel data. Industrial Marketing Management,71,69-78. https://doi.org/10.1016/j.indmarman.2017.11.010.

[41] Pengqiang Nie, Yanxia Wu, Zhenlin Wang, Song Xu, Seiji Hashimoto & Takahiro Kawaguchi. (2025). The Voltage Regulation of Boost Converters via a Hybrid DQN-PI Control Strategy Under Large-Signal Disturbances. Processes, 13(7), 2229-2229. https://doi.org/10.3390/PR13072229.