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

Research on low-latency data governance and monitoring technology of energy metering whole link based on temporal fusion Transformer and neural ODE

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

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

SUMMARY: *Aiming at the problems of weak data timing coupling and lagging anomaly detection in dynamic scenarios of energy metering system, we propose a collaborative architecture of timing fusion Transformer and neural ODE to realize real-time governance and anomaly detection of the whole-link data. The multi-head self-attention mechanism is designed to dynamically aggregate voltage, current and other multi-dimensional timing correlations, and ensure the causality of timing based on causal mask; the continuous time modeling feature of neural ODE is used to replace the discrete computation and reduce the time consuming redundant inference. Experiments show that the architecture in dynamic energy metering scenarios, data governance end-to-end delay ≤ 14ms, voltage/current mutation anomaly detection accuracy of 98.25%, edge computing node memory occupancy of only 0.83GB, and real-time data reporting compliance rate of 98%, which has a high timeliness and strong scenario adaptability in the energy metering all-link governance.*

KEYWORDS: *energy metering system; temporal fusion Transformer; neural ODE; all-link data; real-time governance*

1 Introduction

In modern enterprise management, energy measurement as an important foundation for energy conservation management, more and more by the majority of enterprises attach great importance to the measurement means are also more and more modern, scientific and standardized [1]. Energy control and metering analysis system is mainly the use of intelligent instrumentation, the enterprise can manage the supply process of energy collection and analysis of intelligent application system, the main function is to the company's various users of electricity, steam, compressed air and other energy media usage management, including energy measurement, statistics, cost accounting, etc. [2, 3]. Scientific measurement of energy, first reflected in the management of various quantities of energy, and the number, the source of the volume is the purchase of energy resources into the enterprise, storage, consumption and reuse of various processes such as measurement, testing results. With accurate, timely, systematic data, planning and scheduling, quota management, supervision and evaluation, analysis and research, economic accounting will have a reliable basis. To this end, the need for production processes, a variety of energy-consuming equipment to carry out perfect daily measurement statistics and heat balance, power balance test, in order to understand the energy utilization situation, analyze the efficiency of energy-consuming equipment, found that the energy utilization rate of the weak links, so as to formulate a targeted data management and monitoring

*zhouzhenglei23@126.com
https://doi.org/10.65102/is2026030
Published: 30 April, 2026

programs, targeted energy-saving technology transformation [4, 5]. Taking the key issue of low latency data governance as an example, based on the traditional energy metering management system, it is usually only possible to make predictions based on historical electricity consumption, which is not accurate and comprehensive. With the development of big data technology, the field of data governance can be monitored to face new challenges, the governance of large-scale data and the processing of time-sensitive real-time data. The current data governance monitoring system is unable to cope with real-time demand and the system performance is not high enough to cope with the calculation of massive data. In view of the many problems of the existing data governance and monitoring system, it is necessary to study a real-time data governance and monitoring technology with perfect function, high computational efficiency, and support for massive data processing for enterprise data monitoring and management [6, 7].

Literature [8] constructed a collaborative metering system architecture based on blockchain technology to realize the management of large-scale data sets of the power grid, and safeguarded the integrity and security of the data through encryption technology. To improve the accuracy and efficiency of the smart metering system, literature [9] designed an edge-fog-cloud computing architecture and proposed a data processing framework for smart metering and IoT, which reduces the communication latency, response time, resource loading, and utilization by allocating processing at each layer. Literature [10] establishes adaptive sampling, dynamic power management and sleep scheduling algorithms to accomplish data monitoring, and the results show that the proposed algorithm reduces the energy consumption and information delay, extends the network lifetime, and improves the packet transmission rate, which is a big progress compared with the traditional methods. Literature [11] points out three limitations of multivariate time series, discrete neural architecture, high complexity, and dependence on graph prior. In addressing the existing limitations through continuous modeling, multivariate time series abstraction provides dynamic graphs for the subsequent structure, this study demonstrates that MTGODE is data accurate. Literature [12] mentions time fusion transformer as a novel recurrent neural network for forecasting short term power loads, power loads are an important part of the power system, in the control group is compared with the transformer and the transformer prediction results are shown to be better than other methods in terms of MAE and MAPE. Literature [13] designed multivariate variational modal decomposition, whale optimization algorithm along with the converter and used sample entropy to determine the appropriate number of decomposition levels and penalty factor for MVMD which is used to improve the overall performance and experiments show that this case due to other performance. Literature [14] designed cross-entity time fusion transformer, neural network model using cross-entity attention mechanism to model the correlation between entities, enhance the relationship between entities within the time window, reduce the computational complexity and reduce the computational error. In the study of literature [15], it is stated that the converter can accurately predict heat loads, efficiently operate factors such as heating and cooling and reduce emissions, can maximize the use of renewable resources, and shows relatively good results in the field of time series prediction.

Therefore, this paper proposes a low-latency data governance and monitoring technique based on time series fusion Transformer and Neural ODE for energy metering in the whole link, which is designed to dynamically aggregate multi-dimensional time series correlations such as voltage, current, etc. by designing a multi-head self-attention mechanism by fusing Transformer with Neural ODE, and guaranteeing the time series causality based on the causality mask. The continuous time modeling characteristics of neural ODE are used to replace discrete computation, reduce the time consuming redundant reasoning, and ultimately realize link-wide low-latency data governance and monitoring. The research in this paper can effectively solve the deficiencies of the existing data governance and monitoring system in dealing with massive data and real-time demand, and ensure that the data can accurately and well support the strategic decision-making of the enterprise.

2 Time-Series Fusion Transformer with NeuroODE

2.1 Overall structure

In this paper, the timing-related features are fully considered, and the proposed method accepts a variable-length input sequence to identify the distribution of low-latency data [16, 17]. The overall architecture is shown in Fig. 1, with Transformer as the backbone network, which is composed of three modules, the first part is the energy metering data extraction network, which mainly consists of lightweight MobileNet V3, and performs feature extraction on the input raw data sequence. The second part is the feature refinement network, mainly composed of Transformer, which further extracts global spatial features and global attention features as well as temporal features from the feature maps extracted in the first part. The third part is the Neural ODE model, which performs real-time governance and anomaly detection based on the features obtained in the second part.

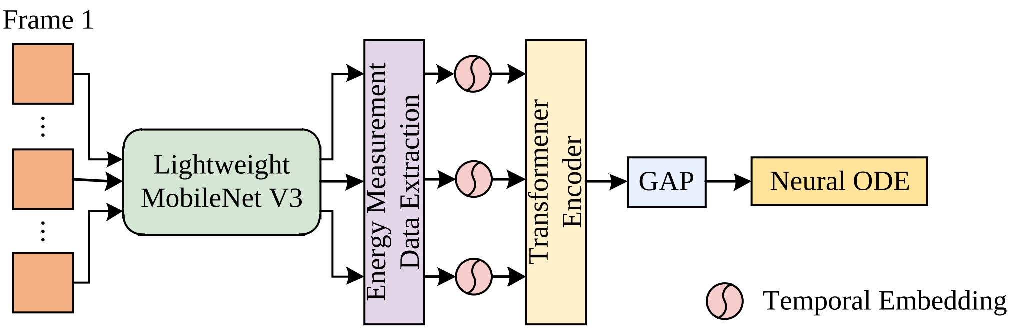


Figure 1: Overall Architecture

The entire framework can be defined as a function:

$$\Gamma = R^{m\times w\times h\times c} \to R^{n} \tag{1}$$

This function extracts the data sequences with m frame size of $w\times h$ and number of channels of c over a period of time t to obtain the low-latency data distributions of n energy metering in the set $S_t \in R^{m\times w\times h\times c}$ R^n. The first part of this function extracts spatial features from the input raw data, defined as:

$$f(t) = F(S) \tag{2}$$

where F denotes a mapping of the set $S_t \in R^{m\times w\times h\times c}$ to the set $R^{m\times k}$. This function extracts features from the input data and maps them onto h-dimensional feature vectors, and then fuses the feature vectors extracted on the m-columns of data within time t to obtain a feature set $S_t \in R^{m\times k}$. Therefore, here F can be expressed as a fusion of m feature extraction functions:

$$F = f_t^0 \oplus f_t^1 \oplus \cdots f_t^m \tag{3}$$

where $f_t^i \in G(s), i = 1, 2, 3, \cdots, m$ denotes the feature extraction of the data in the i th column within the time series t, where $G(s): R^{m \times h \times c} \to R^k$, and $\oplus$ denotes the join operator.

The second part is to further extract the timing features from the features extracted in the first part F, mainly using the Transformer architecture for the feature analysis, and this process can be expressed as the following function:

$$h_t = H(F(S_t)) \tag{4}$$

Among them, $H: R^{m \times k} \to R^t$ denotes a temporal feature mixing function, which extracts high-level abstract features including temporal features and global spatial features as well as global attentional features from the output of the first part.

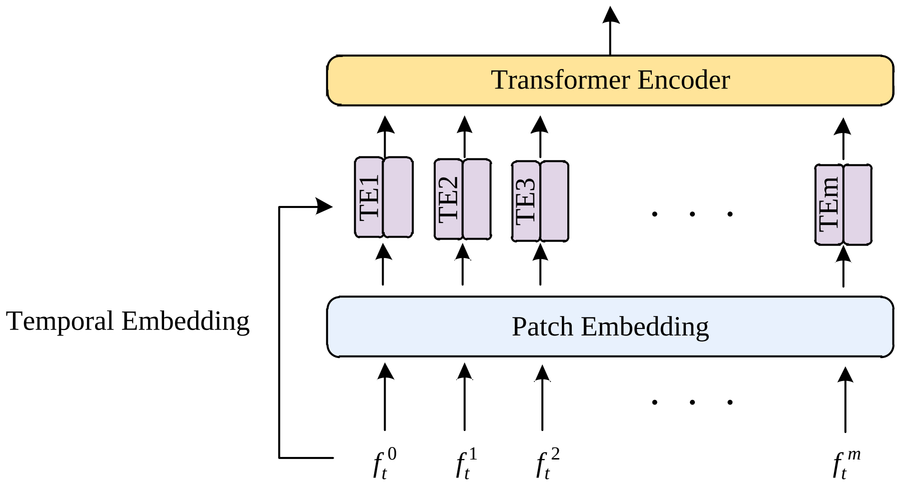
The last part is a classifier which classifies the features based on the previously extracted features, and maps the feature set R^t extracted in the previous two parts to a set of probability distributions R^n containing n data sequences. This process can be represented as the following function:

$$y_t = Y(h_t) = Y(H(F(S_t))) \tag{5}$$

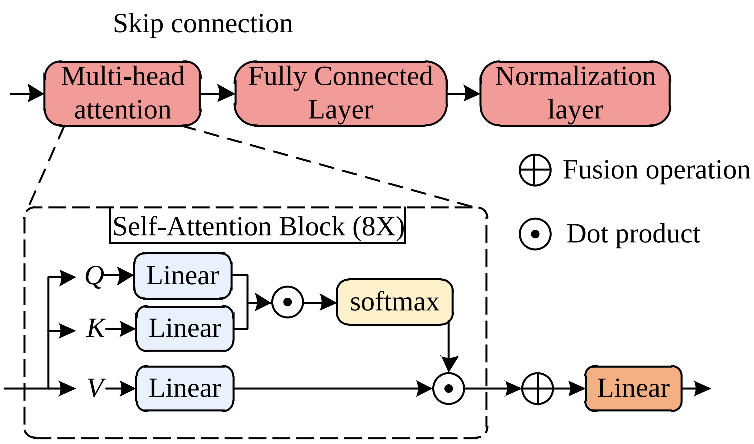
where $Y: R^t \to R^n$, y_t denotes the n-dimensional vectors of the n-data sequence.

2.2 Transformer Encoder

Based on the above, each data sequence within the temporal sequence t is input into the lightweight MobileNet V3 feature extraction network to obtain the feature vector $f_t^i (i = 1, 2, 3, \cdots, m)$, and then these feature vectors are input as a single vector into the embedding module, and then joined into the temporal embedding for stitching The total feature vector is obtained by fusion $F(S_t)$ input into the Transformer encoder for global attention coding to obtain the relevant timing feature vector. Figure 2 shows the temporal features and encoder structure, Figure 2(a) shows the temporal feature vectors, each sequence of data within the temporal t is output through the first part of the feature extraction to obtain a module, here i.e. $f_t^i (i = 1, 2, 3, \cdots, m)$. Then each module is passed through the Transformer encoder and combined with the timing embedding to obtain a one-dimensional vector input to the Transformer encoder [18]. Figure 2(b) shows the structure of Transformer encoder, there are 6 connected Transformer encoders.



(a) Temporal feature vector



(b) Encoder structure

Figure 2: Timing Characteristics and Encoder Structure

2.3 NeuroODE model

In this section, a feature extraction component has been designed using neural ODE, and based on this component the process of interaction and matching of input energy data features has been modeled, and the neural ODE interaction and matching process is shown in Figure 3. After going through the process of obtaining the deep feature information using the feature extraction component constructed using the neural ODE, and fully integrating it with the data embedding system, and the introduction of another data source, i.e., another data sequence representation with attentional information of the three features, the final matrix information representation obtained has been able to categorize the two input data sequences [19-21].

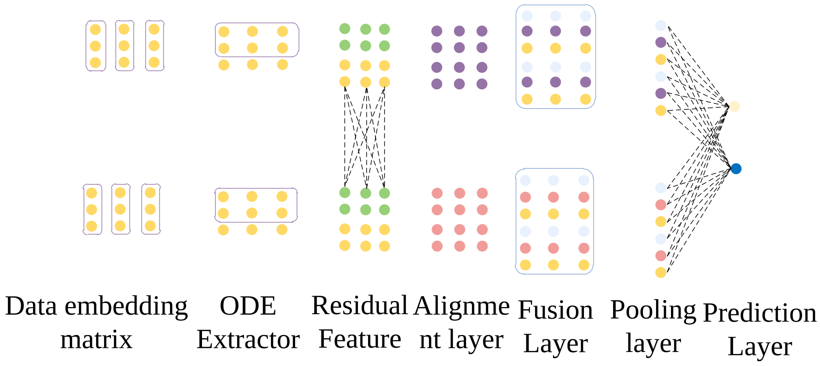


Figure 3: Structure of Neural ODE Model

After continued fusion of the matrix representation fused with global data information using the ODE-based feature component, a deeply learned feature vector representation is obtained. In order to retain the previously learned features of various dimensions, the previously obtained shallow embedding matrix, and the introduction of another feature representation with attention information are spliced to obtain the high-dimensional feature representations a'' and b'', following the principle of real-time energy metering of the whole link, this paper chooses the pairwise subtraction, pairwise multiplication, and the direct splicing of the three operations for the governance and monitoring of the energy metering in order to reflect the difference and similarity between high-dimensional features a'' and b''. The whole process can be expressed as:

$$y = W\left[a'';b'';a''-b'';a''\circ b''\right] \tag{6}$$

where W is a multilayer feed-forward neural network, while, in this paper, another simpler way of real-time data acquisition is used, i.e., only splicing operation is used, and this process can be expressed as Eq:

$$y = W\left[a'';b'';a''-b''\right] \tag{7}$$

The final obtained y is an un-normalized dataset with the dimension of the final obtained low latency data. In this paper, the $\arg\max$ function is used to obtain the final result, which is expressed as follows in Eq:

$$y = \arg\max\left(y_i\right) \tag{8}$$

3 Low-latency data governance and monitoring

3.1 Transformer Enhanced Dynamic Processing Capabilities

3.1.1 Multi-attention mechanisms to enhance temporal correlation

Transformer's multi-attention mechanism can help the energy metering system to solve the problem of a certain degree of gradient explosion or disappearance, and realize the dynamic aggregation of multi-dimensional time-sequence correlations such as voltage, current, etc., but not all of them are necessarily the most important time-sequence change features that are retained [22, 23]. Higher-level extraction of the initial extracted features of the energy metering system can further enhance the model's ability to learn the data sequences implicit in the historical data. Based on this, it is proposed to enhance the time series correlation with Transformer by taking the intermediate vectors of the energy metering system as the underlying semantics, using the self-attention mechanism in the Transformer encoder to do parallel computation on the underlying data, and transforming the computation results of the self-attention mechanism into further extracted multidimensional time series data through the fully-connected layer. In addition, residual connectivity and layer normalization are performed between each sub-layer, enabling the system to always retain a portion of the original information during the continuous passing of the gradient, and to capture the dynamic relationship of the input data in time in a timely manner.

The voltage and current extracted by the multi-head self-attention mechanism are used as aggregation vectors, and the recognition results are obtained after using the load data as a common input. Figure 4 shows the data processing process. For the load data, the input to the ODE model at the moment $(\sigma-1)$ is the sequence data $X=\{x_{\sigma-N},x_{\sigma-N+1},\cdots,x_{\sigma-1}\}\in\Re$, where N is the length of the sequence, and the real-time state $h_{\sigma-1}$ and the vector state $c_{\sigma-1}$ are updated according to the iterative process at that moment, after accepting the input sequence sequentially.

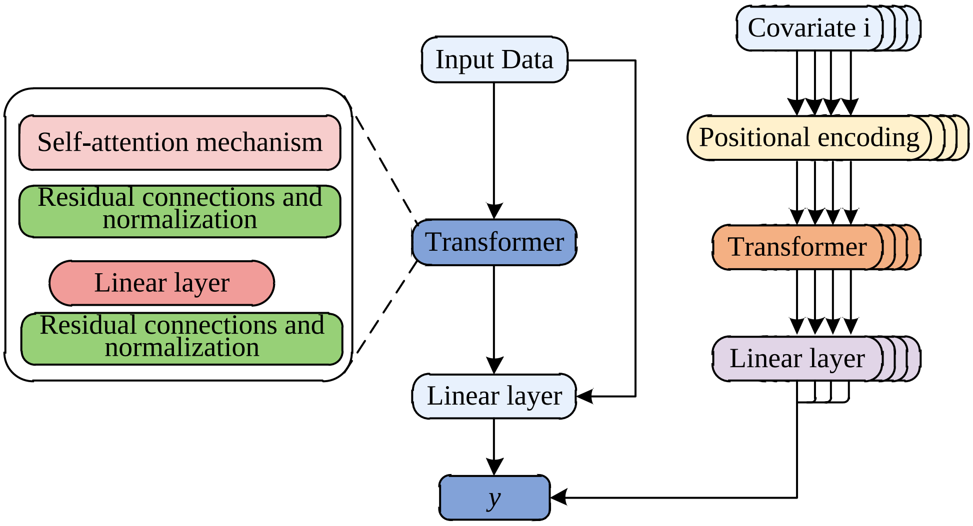


Figure 4: Data processing process

As the real-time state $h_{\sigma-1}$ and the vector state are co-passed into the Transformer encoder, within which the self-attention computation is performed, the real-time state is further extracted through the Transformer layer to obtain a sequence of features, denoted as O_T. Finally, using this extracted feature sequence as the aggregation result of the multi-head attention mechanism, the same sequence data X vector is spliced as a common input, and the data governance result y_l is obtained after the ODE model calculation:

$$O_T = Transformer*([X;O_T]) \tag{9}$$

$$y_l = W_3 O_l + b_3 \tag{10}$$

where $Transformer*$ represents the computational module of the multi-attention mechanism, O_l represents its output, and W_3 and b_3 are its corresponding voltage and current states, respectively.

For the system load as a covariate input, considering that energy systems such as smart grids usually have small and relatively flat variations over the time horizon of short-term load forecasting, a separate multi-attention mechanism is used as a covariate for data computation in order to allow Transformer to focus on learning the energy metering laws. Since Transformer does not process each data in the sequence step by time step like recurrent neural networks, which will result in the input data losing the original order information, the order information can be encoded into the original sequence by neural ODE. Eventually, the multidimensional time series correlation of voltage, current, etc. is realized.

3.1.2 Causal masks guarantee timing causality

In order to guarantee and ensure the temporal causality based on causal mask, this paper designs the dynamic hierarchical Transformer sequence recommendation algorithm architecture, and the causal mask architecture is shown in Fig. 5. It mainly consists of four parts: Embedding embedding layer, dynamic hierarchy Transformer module, Transformer layer and fully connected output layer. Through Embedding embedding layer, the distributed representation of historical behavioral sequences and energy data is obtained. The Dynamic Hierarchy Transformer module performs layer-by-layer merging of neighboring blocks from the bottom up through the attention mechanism of neighboring blocks to form a multi-scale metering link and obtains a real-time state representation of the full link.The Transformer layer models the dependencies embodied in the low-latency data sequences by using the self-attention mechanism, and ensures the temporal causality through the ODE module [24, 25].

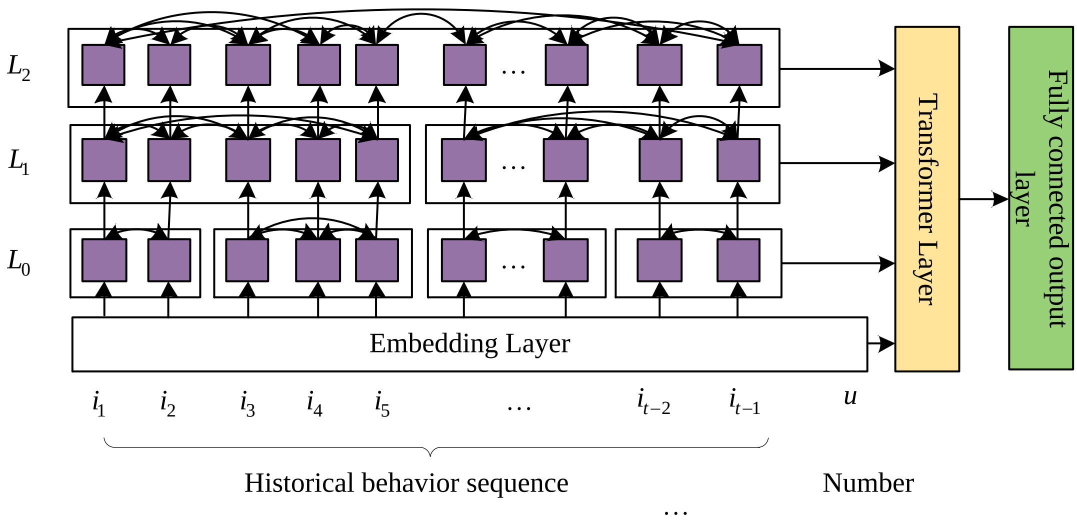


Figure 5: Causal Mask Architecture

3.2 Neural ODE modeling solution

The ODE for ordinary differential equations is based on the iterative solution of the grid equations, where the residuals of the fine-grid equations are constrained to be solved on the coarse grid, and then the resulting solution is extended over the fine grid and combined with the original solution to form an exact solution on the fine grid [26]. The solution flow is shown in Fig. 6. The proposed neural ODE framework originates from the idea of combining multiple grids to solve the equations. The numerical solution obtained by the physical information neural network and its gradient with respect to the input vectors are transferred to the new fully connected neural network, and a loss function is built for it to obtain the correction of the numerical solution, which is then combined with the numerical solution obtained by the physical information neural network to form the exact solution.

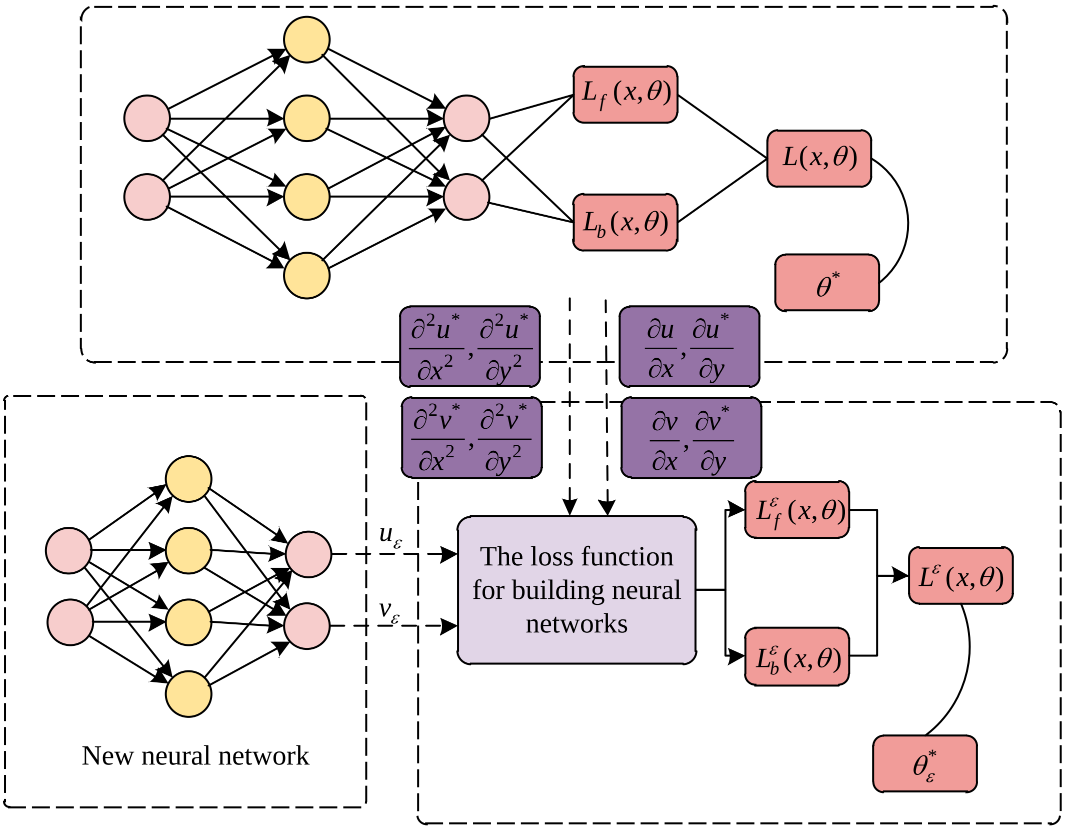


Figure 6: Solution Process

For solving the constant partial differential equation, assume that the numerical solution obtained by the physically informative neural network solution is $u^*(x) = f(x;\theta^*)_L$, which is obtained by substituting it into the constant partial differential equation:

$$\begin{aligned} &N_x[u^*] \approx h(x) \\ &u^* \approx g(x) \end{aligned} \tag{11}$$

According to Eq. (11), it can be seen that the numerical solution obtained by the physically informative neural network solution $u^*(x)$ has an error $\varepsilon = u - u^*(x)$ from the true solution u. This section considers a new fully connected neural network ε for the error $\varepsilon(x;\theta)_L$ so that it satisfies the following equation:

$$N_x\left[u^* + \varepsilon(x;\theta)_L\right] = h(x) \tag{12}$$

$$u^*(x) + \varepsilon(x;\theta)_L = g(x) \tag{13}$$

According to equation (13), the residual equation between the fully connected neural network $\varepsilon(x;\theta)_L$ and the numerical solution $u*(x)$ is constructed $r_\theta^*(x;\theta)$:

$$r_\theta^*(x;\theta) = N_x\left[\varepsilon(x;\theta)_L + u*\right] - h(x) \tag{14}$$

According to Equation (14), the loss function of the neural network $\varepsilon(x;\theta)_L$ is defined as:

$$L^\varepsilon(x;\theta) = \lambda_f^\varepsilon L_f^\varepsilon(\theta,x) + \lambda_b^\varepsilon L_b^\varepsilon(x;\theta) \tag{15}$$

Among them:

$$L_f^\varepsilon(\theta,x) = \frac{1}{N_f^\varepsilon}\sum_{i=1}^{N_f^\varepsilon}\left(r_\theta^*\left(x_f^i;\theta\right)\right)^2 \tag{16}$$

where λ_f^ε is the weight factor between the loss functions, and x_f^i denotes the training set that satisfies the residual equations and boundary conditions, and in order to improve the computational accuracy, the training set is independent of each other from the training set selected by the neural network, and the loss functions in the neural network $\varepsilon(x;\theta)_L$ loss function $L^\varepsilon(x;\theta)$ that satisfies the numerical solution of the constant partial differential equation to replace the discrete computation and to reduce the time-consumption of redundant reasoning.

3.3 Performance indicators

In order to verify the accuracy of this paper's full-link low-latency data governance and monitoring technique for current/voltage mutation detection, four commonly used evaluation metrics, namely, $Accuracy$, $Recall$, $Precision$, and F1-value, are used to evaluate the detection accuracy. It is assumed that the data features in the dataset are categorized into two classes, positive and negative. TP denotes the number of positive class detections predicted by the algorithm into positive classes, TN denotes the number of positive class detections predicted into negative classes, FP denotes the number of negative class detections predicted into positive classes, and FN denotes the number of negative class detections predicted into negative classes. The formulae for each of the four rating indicators are:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{17}$$

$$Recall = \frac{TP}{TP+FN} \tag{18}$$

$$Precision = \frac{TP}{TP+FP} \tag{19}$$

$$F1 = \frac{2\cdot Precision\cdot Recall}{Precision+Recall} \tag{20}$$

4 Analysis of the results of the energy measurement experiment

4.1 Delayed data governance

In order to verify the effectiveness of this paper's link-wide low-latency data governance and monitoring method in latency data governance, Spark-based link-wide low-latency data governance and monitoring, Zigbee network hybrid routing-based link-wide low-latency data governance and monitoring, Transformer-based link-wide low-latency data governance and monitoring, and neural ODE-based link-wide low-latency data governance and monitoring are used as the comparison items of this paper's temporal fusion Transformer and neural ODE-based energy metering low-latency data governance and monitoring. data governance and monitoring, as a comparison item of this paper's timing fusion Transformer and neural ODE based energy metering all-link low-latency data governance and monitoring. The full-link latency data is governed by five full-link low-latency data governance and monitoring methods, including end-to-end data latency, peak processing latency, and model inference time consuming three aspects, and Table 1 shows the results of full-link latency data governance. It can be seen that, in the end-to-end data delay governance and peak processing delay process, Spark-based governance and monitoring method and Zigbee hybrid routing-based governance and monitoring method take more than 100ms, so the delay data processing time is too long, and it is very easy to cause the problem of anomaly detection lag; Transformer-based governance and monitoring and neural ODE-based governance and monitoring methods, end-to-end data delay data is more than 100ms, and it is very easy to cause the problem of anomaly detection lag; Transformer-based governance and monitoring and neural ODE-based governance and monitoring methods, end-to-end data delay data is more than 100ms. monitoring methods, the end-to-end data delay governance and peak processing delay time is less than 50ms, and the governance speed is obviously higher than the former, but in comparison, in this paper, under the governance and monitoring method based on the temporal fusion of Transformer and neural ODE, the end-to-end data delay governance and peak processing delay time is less than 15ms, and the model reasoning time is only 9.53ms, which indicates that the method in this paper has the highest efficiency of delayed data governance. The method in this paper has the highest efficiency in delayed data governance, which helps in the real-time detection of data anomalies.

Table 1: Results of End-to-End Latency Data Governance

Governance and Monitoring Methods	End-to-end data latency/ms	Peak processing latency/ms	Model inference time/ms
Spark-based Governance and Monitoring	147.46	189.33	-
Zigbee-based Hybrid Routing Governance and Monitoring	112.54	127.75	-
Transformer-based Governance and Monitoring	49.57	75.52	25.48
Neural ODE-based Governance and Monitoring	38.33	46.41	16.37
Time-series Fusion Governance and Monitoring of Transformer and Neural ODE	13.76	14.28	9.53

4.2 Detection accuracy

The current/voltage mutation detection results are shown in Table 2, and the metrics evaluation results of the whole-link low-latency data governance and monitoring methods based on Spark, Zigbee network hybrid routing, Transformer, and neural ODE are all lower than 85%, and thus the overall accuracy of the current/voltage mutation detection is not satisfactory. However, the results of the current/voltage mutation detection metrics are around 98% after adopting the governance and monitoring methods of time-sequence fusion Transformer and neural ODE, thus indicating that the governance and monitoring methods based on time-sequence fusion Transformer and neural ODE in this paper can accurately detect current/voltage mutation.

Table 2: Detection results of sudden current/voltage changes

Governance and Monitoring Methods	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
Spark-based Governance and Monitoring	80.87	76.22	79.36	78.48
Zigbee-based Hybrid Routing Governance and Monitoring	83.68	77.46	81.54	80.45
Transformer-based Governance and Monitoring	81.46	80.75	82.33	79.47
Neural ODE-based Governance and Monitoring	79.12	84.35	83.41	83.49
Time-series Fusion Governance and Monitoring of Transformer and Neural ODE	97.95	98.37	98.25	97.91

4.3 Resource utilization rate

Resource occupancy is an important indicator of the performance of the full-link low-latency data governance and monitoring method, and the lower the edge-point memory occupancy, the smoother the operation will be in the device. The lower average CPU utilization will represent the lower complexity of the governance and monitoring algorithms, and the results of the resource occupancy rate are shown in Table 3. The governance and monitoring methods of Spark, Zigbee hybrid routing, and Transformer all occupy more than 3.6 GB of video memory at the edge point, and the average CPU utilization is higher than 60%, so the resource occupancy is high, which affects the running speed of the device; the governance and monitoring method based on neural ODE occupies 1.7 GB of video memory at the edge point, and the average CPU utilization rate is 52.4%, which is significantly better than the previous methods. However, in comparison, the governance and monitoring method based on temporal fusion Transformer and neural ODE occupies only 0.83GB of edge point video memory, with an average CPU utilization rate of only 37.6%. Therefore, it can be seen that the governance and monitoring method based on time-series fusion Transformer with neural ODE has the ideal performance advantage.

Table 3: Resource Utilization Results

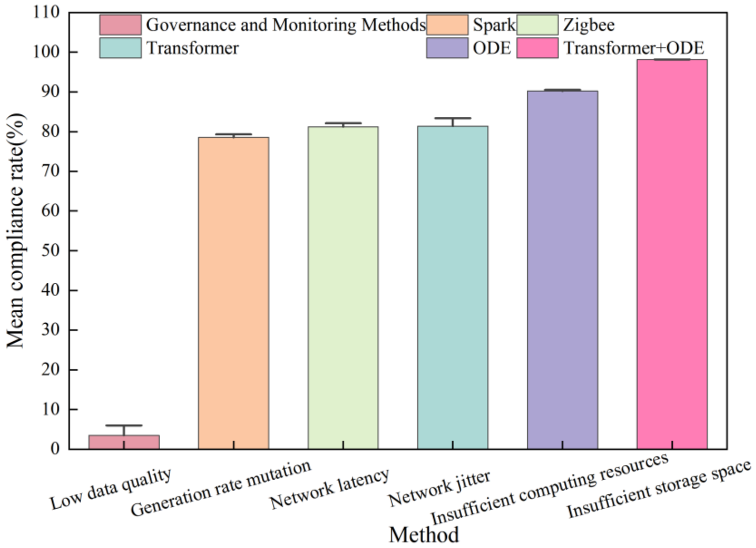
Governance and Monitoring Methods	Edge point video memory usage /GB	Average CPU utilization /%
Spark-based Governance and Monitoring	4.7	69.8
Zigbee-based Hybrid Routing Governance and Monitoring	4.1	60.3
Transformer-based Governance and Monitoring	3.6	63.6
Neural ODE-based Governance and Monitoring	1.7	52.4

Time-series Fusion Governance and Monitoring of Transformer and Neural ODE	0.83	37.6

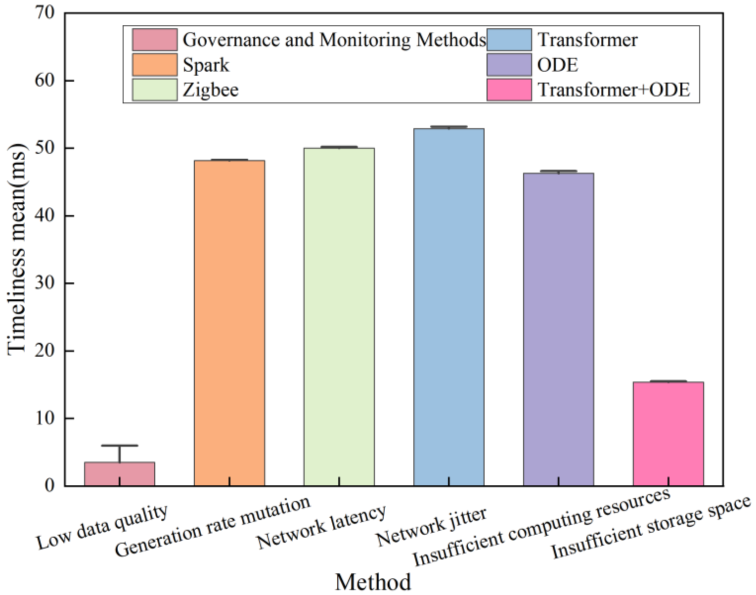
4.4 Regularity and timeliness of data reporting

Although low-latency data in the full link can realize data transmission in a very short time, the data is likely to be damaged and lost in the rapid transmission, so although the low-latency data can guarantee the real-time data transmission, it cannot guarantee the integrity and high quality of the data. In the process of governance and monitoring of all-link low-latency data, it is necessary to determine whether there is any abnormality in the data through real-time reporting of data, but if there is incomplete, inaccurate, untimely, and unstandardized data reported, it is impossible to accurately determine the abnormality of the data through the reported data, so the rate of completeness, accuracy, standardization, and timeliness is an important index for determining the qualification rate of the data reported. Figure 7 shows the data compliance and and timeliness, the data reporting pass rate under different governance and monitoring methods is shown in Figure 7(a), Spark, Zigbee hybrid routing, Transformer's governance and monitoring methods data reporting pass rate in are lower than 90%, and can not provide reliable data for the whole-link low-latency data governance and monitoring. While the reported data qualification rate of the neural ODE governance and monitoring methods are all around 90% and the data integrity rate is more satisfactory, the reported data qualification rate of the timing fusion Transformer and the neural ODE governance and monitoring methods are all around 98%, which can provide more complete, accurate, standardized, and reliable data for the whole-link low-latency data governance and monitoring, in order to promote the data in the whole-link It can provide more complete, more accurate, more standardized and more reliable data for the governance and monitoring of all-link low-latency data, so as to promote the governance and monitoring of data in the whole-link transmission process and realize the real-time detection of data anomaly.

Reporting data timeliness is shown in Fig. 7(b), the neural ODE governance and monitoring method will have an average delay of 46.3ms in the process of reporting data, which is significantly better than the delay of Spark, Zigbee hybrid routing, and Transformer's governance and monitoring method, but it still fails to satisfy the requirements of real-time governance and monitoring of the whole-link data. In contrast, this paper's timing fusion Transformer and neural ODE governance and monitoring method has an average delay of 15.4ms when reporting data, and the low latency and high data quality fully satisfy the requirements of real-time governance and monitoring of all-link data.



(a) Data reporting pass rate under different governance and monitoring methods



(b) Timeliness of data reporting

Figure 7: Data reporting patterns and timeliness

5 Conclusion

In this paper, in order to carry out perfect daily metering statistics and heat balance and power balance tests for various energy-consuming equipment, so as to understand the energy utilization situation, we propose a research on low-latency data governance and monitoring technology for energy metering whole link based on timing fusion Transformer and neural ODE. And we analyze the effectiveness of the method proposed in this paper from four dimensions: delayed data governance, detection accuracy, resource occupation rate, and data reporting qualification rate. Under the governance and monitoring method based on timing fusion Transformer and neural ODE, the end-to-end data latency governance and peak processing latency are lower than 15ms, the model inference takes only 9.53ms, and the detection accuracy of current/voltage anomalies is 98.25%. Under the governance and monitoring method based on neural ODE, the edge point memory occupied is 1.7GB, and the average CPU utilization rate is 52.4%, although the overall situation is more ideal, but in comparison, under the governance and monitoring method based on the timing fusion Transformer and neural ODE in this paper, the edge point memory occupied is only 0.83GB, and the average CPU utilization rate is only 37.6%. It can be seen that based on the timing fusion Transformer and neural ODE can realize the efficient and stable operation of the energy metering system, ensure the timeliness and completeness of the data, and facilitate the real-time governance and monitoring of the whole-link low-latency data.

About the Authors

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.

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.

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.

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.

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., Eniola, S. O., Ponnle, A. A., Agbolade, O. A., Udekwe, C. N., & Melodi, A. O. (2021). A review of smart energy metering system projects. *Jurnal Elektronika dan Telekomunikasi*, 21(1), 70-78.

[2] Rind, Y. M., Raza, M. H., Zubair, M., Mehmood, M. Q., & Massoud, Y. (2023). Smart energy meters for smart grids, an internet of things perspective. *Energies,* 16(4), 1974.

[3] Avancini, D. B., Rodrigues, J. J., Rabêlo, R. A., Das, A. K., Kozlov, S., & Solic, P. (2021). A new IoT-based smart energy meter for smart grids. *International Journal of Energy Research*, 45(1), 189-202.

[4] Zhou, K., Meng, E., Jin, Q., Luo, B., & Tian, B. (2023). Evaluation of data governance effectiveness in power grid enterprises using deep neural network. *Soft Computing*, 27(23), 18333-18351.

[5] Collins, A., Hamza, O., Eweje, A., & Babatunde, G. O. (2024). Challenges and solutions in data governance and privacy: A conceptual model for telecom and business intelligence systems. *Telecom and Business Intelligence Journal*, 9(3), 110-125.

[6] Pestana, G., & Sofou, S. (2024). Data governance to counter hybrid threats against critical

infrastructures. Smart Cities, 7(4), 1857-1877.

[7] Marcucci, S., Alarcón, N. G., Verhulst, S. G., & Wüllhorst, E. (2023). Informing the global data future: benchmarking data governance frameworks. *Data & Policy*, 5, e30.

[8] Zanghi, E., Brown Do Coutto Filho, M., & Stacchini de Souza, J. C. (2024). Collaborative smart energy metering system inspired by blockchain technology. *International Journal of Innovation Science,* 16(2), 227-243.

[9] Oprea, S. V., & Bâra, A. (2023). An Edge-Fog-Cloud computing architecture for IoT and smart metering data. *Peer-to-Peer Networking and Applications,* 16(2), 818-845.

[10] Niu, Y., Vugar, A., Wang, H., & Jia, Z. (2023). Optimization of energy-efficient algorithms for real-time data administering in wireless sensor networks for precision agriculture. *Khwarizmia*, 2023, 24-36.

[11] Jin, M., Zheng, Y., Li, Y. F., Chen, S., Yang, B., & Pan, S. (2022). Multivariate time series forecasting with dynamic graph neural odes. *IEEE Transactions on Knowledge and Data Engineering*, 35(9), 9168-9180.

[12] Huy, P. C., Minh, N. Q., Tien, N. D., & Anh, T. T. Q. (2022). Short-term electricity load forecasting based on temporal fusion transformer model. *Ieee Access*, 10, 106296-106304.

[13] Ye, H., Zhu, Q., & Zhang, X. (2024). Short-term load forecasting for residential buildings based on multivariate variational mode decomposition and temporal fusion transformer. *Energies*, 17(13), 3061.

[14] Wang, Z., Zhu, Z., Xiao, G., Bai, B., & Zhang, Y. (2022). A transformer-based multi-entity load forecasting method for integrated energy systems. *Frontiers in Energy Research*, 10, 952420.

[15] Behrens, F., Leiprecht, S., Brantl, J., & Finkenrath, M. (2022). Temporal Fusion Transformer for thermal load prediction in district heating and cooling networks. *Scandinavian Simulation Society*, 333-338.

[16] Bhatt, N., & Thakkar, A. (2021). An efficient approach for low latency processing in stream data. *PeerJ Computer Science,* 7, e426.

[17] Chaccour, C., Soorki, M. N., Saad, W., Bennis, M., & Popovski, P. (2022). Can terahertz provide high-rate reliable low-latency communications for wireless VR?. *IEEE Internet of Things Journal*, 9(12), 9712-9729.

[18] Labaien, J., Idé, T., Chen, P. Y., Zugasti, E., & De Carlos, X. (2023). Diagnostic spatio-temporal transformer with faithful encoding. *Knowledge-based systems,* 274, 110639.

[19] McDonald, T., & Álvarez, M. (2021). Compositional modeling of nonlinear dynamical systems with ODE-based random features. *Advances in neural information processing systems*, 34, 13809-13819.

[20] Lai, Z., Liu, W., Jian, X., Bacsa, K., Sun, L., & Chatzi, E. (2022). Neural modal ordinary differential equations: Integrating physics-based modeling with neural ordinary

differential equations for modeling high-dimensional monitored structures. *Data-Centric Engineering*, 3, e34.

[21] Bradley, W., & Boukouvala, F. (2021). Two-stage approach to parameter estimation of differential equations using neural odes. *Industrial & Engineering Chemistry Research,* 60(45), 16330-16344.

[22] Xuan, Y., Pang, C., Yu, H., Zeng, X., & Chen, Y. (2024). Load energy decomposition algorithm based on improved bidirectional transformer combined with time-sensing self-attention. *IEEE Access*, 12, 75625-75639.

[23] Zheng, K., Sun, L., Wang, X., Zhou, S., Li, H., Li, S., ... & Gong, Q. (2021). Named entity recognition in electric power metering domain based on attention mechanism. *IEEE Access,* 9, 152564-152573.

[24] Kim, S., Ji, W., Deng, S., Ma, Y., & Rackauckas, C. (2021). Stiff neural ordinary differential equations. *Chaos: An Interdisciplinary Journal of Nonlinear Science,* 31(9).

[25] Gong, C., Zhang, C., Yao, D., Bi, J., Li, W., & Xu, Y. (2024). Causal discovery from temporal data: An overview and new perspectives. *ACM Computing Surveys*, 57(4), 1-38.

[26] Zou, Z., Meng, X., Psaros, A. F., & Karniadakis, G. E. (2024). NeuralUQ: A comprehensive library for uncertainty quantification in neural differential equations and operators. *SIAM Review*, 66(1), 161-190.