GRU-TemporalAttention Model with Time Series Decomposition Module for Battery Fault Diagnosis in Quadrupedal Robots

Yingze Yang[2], Xiaolong Chen[1], Weirong Liu[1], Chenglong Wang[1]
Wei Yuan[1] and Jun Peng[2],
[1]School of Computer Science and Engineering, Central South University, Changsha, China
[2]School of electronic information, Central South University, Changsha, China
Corresponding Author: Weirong Liu Email: frat@csu.edu.cn

***Abstract*—Due to the dynamic work conditions of quadrupedal robots, the performance of batteries could be degraded and further cause serious safety issues. Therefore, it is crucial to design a feasible fault diagnosis method to ensure the safe and efficient operation of robots. To address this issue, a deep temporal attention network with series decomposition is proposed for fault diagnosis. The proposed network can decompose time series data into principal trends and residual volatility. By incorporating gated recurrent unit layers and temporal attention layers, the decomposed part can be fused and reconstructed for fault diagnosis. After that, a dynamic threshold method considering battery cell consistency is designed to determine the fault threshold according to various work conditions. Finally, extensive experiments are conducted to verify the effectiveness of the proposed method. The experiment results demonstrate that the proposed method is superior to other fault diagnosis models.**

***Index Terms*—lithium-ion battery, fault diagnosis, quadrupedal robot**

I. Introduction

Quadrupedal robots play a key role in fields such as mobile rescue and logistics and transportation. However, due to their special application scenarios, they need to work without external power support. Lithium-ion batteries are widely used in mobile robots due to their high energy density and cyclic charge/discharge advantages [1]. Despite the many advantages of lithium-ion batteries, operating characteristics such as frequent start-stop, power and torque variations in quadrupedal robots may have an impact on battery performance and further cause safety accidents. Therefore, fault diagnosis of quadrupedal robot batteries becomes crucial to ensure their efficient and safe operation in missions.

Typically, fault diagnosis of mobile robots can be categorized into model-based and data-driven approaches. The model-based approach establishes a mathematical model based on the physical characteristics of the robot system and other mechanisms to predict the behavior of the system. A fault is considered to occur when the actual behavior differs significantly from the predicted behavior [2]. Zhao et al. developed a mathematical model of a 6-axis robotic arm to detect its actuator faults [3]. Sabry et al. proposed a measurement-based mathematical model for energy consumption simulation for industrial robot monitoring [4]. These model-based approaches for robot fault diagnosis are very effective and suitable for specific objects such as structured machines like industrial robotic arms, which have relatively homogeneous application scenarios and it is easier to build their accurate mathematical models. However, footed mobile robots usually operate in unstructured environments, and their work scenarios and characteristics make it difficult to build accurate models [5]. Data-driven approaches, which do not require in-depth mechanisms and learn information from large amounts of data, have gained widespread attention in recent years.

There has been a great deal of research on using data-driven methods for various types of fault diagnosis in robots. Machine learning is widely used for fault diagnosis. However, as the complexity of the system increases, the data often exhibit high dimensionality and nonlinear features. Traditional ML methods are difficult to extract effective features from these data due to their limited learning ability. Deep learning methods are gradually becoming the mainstream methods for fault diagnosis nowadays. There is a review study on deep learning algorithm-based fault detection in Industrial Robot [6]. Kasap et al. using dissimilarity-based methods such as DTW for fault diagnosis of autonomous mobile robots, this math-based method is efficient, but it is difficult to identify tiny faults due to data noise and lack of learning about temporal trendiness [7]. Yun et al. using raw auto-encoder (AE) to reconstruct the robotic arm data captured by internal sound sensor, when the reconstruction error is greater than a threshold, it is detected as a fault, this approach can synthesize the relationship between different variables, but the feature extraction ability is weak [8]. Xia et al. Using CNN and adversarial domain adaptive method for joint bearing fault diagnosis of industrial robots, CNN has strong local feature extraction ability, but the global feature capture ability is weak [9]. Lu et al. using dual-module attention and CNN for fault diagnosis of industrial robot reducer, multi-scale features were extracted [10]. Wang et al. introduced CNN and LSTM for fault diagnosis of industrial robot motor drive control system, using CNN to extract the dynamic features of the data and

The data and code are available on the website.

LSTM to predict the system in time domain [11].

Battery data, as time-series data, can be used in a similar way for fault diagnosis. However, quadrupedal robots usually need to carry batteries and maintain good maneuverability in a limited space and are constrained by energy consumption and do not use too many sensors to capture battery information. Therefore, how to fully extract effective information from a small number of features is the difficulty in robot battery fault diagnosis. Jiang et al. uses variational modal decomposition to decompose the battery voltage into multiple modal signals, and then uses isolated forests to diagnose faults on the high-frequency components, this method ignores the low-frequency components, and it is easy to lose the information of tiny faults [12]. Zhang et al. Use the attention mechanism network of sliding window to reconstruct the battery voltage, and diagnose the faults according to the reconstruction error. This method applies the attention mechanism to capture local features within the window, which increases the ability of local feature extraction, but limits the range of the information captured [13]. Zhao et al. uses a gated recurrent network for battery voltage prediction and diagnoses faults based on the residual difference between it and the true value, the model can only work on time series data with strong trend, and it is difficult to learn complex time series modes [14].

To address this issue, this paper embeds the sequence decomposition method into the network module so as to separate the temporal trend from the complex modality, and chooses to use the gated recurrent unit (GRU) as the trend term learning module, and uses the temporal attention mechanism as the learner of the complex modality. By appealing the three basic modules constitute the feature extraction module of the fault diagnosis model in this paper, and the deep information is mined by stacking this module. Finally, the features of different scales are concatenated and input to the fully connected layer for reconstruction to realize accurate fault diagnosis.

II. Method

For fault diagnosis of time-series data, the key is to identify the trend of time-series data and increase the noise resistance of the model. In this paper, the time-series decomposition module is embedded into the neural network, and each time we pass through the time-series decomposition module, we can get the trend term and the residual volatility term of the time-series data; the trend term is inputted into the gated recurrent unit (GRU) so as to learn the trendiness of the time-series data, and the residual volatility term is inputted into the time-series attention mechanism module to learn the higher-order hidden information contained therein. The three appealing modules form the basic module of the network, and the effect of progressive decomposition can be achieved by stacking this basic module. Finally, the trend terms and residual volatility terms output from the modules are spliced together and then fed into the fully connected layer to reconstruct back to the original input data. Finally, the residuals between the reconstructed data and the original data are calculated, and when the residuals exceed a threshold, the battery is considered to have failed.

Algorithm 1: Propsed model

Input: $\mathrm{X} \in R^{L \times d}$
Output: $\mathrm{X}^{rec} \in R^{L \times d}$
Hyper-parameters: Module block number N
Initialize: $\mathrm{X}_R = None$

$\mathrm{X}_T,\ X_R = Module\_Block\_List[1](X, X_R)$
For $(i = 2 \quad to \quad N)$
  $\mathrm{X}_T,\ X_R = Module\_Block\_List[i](X_T, X_R)$
$Multi\_Scale\_Feature = Concat(X_T,\ X_R)$
$\mathrm{X}^{rec} = Linear(Multi\_Scale\_Feature)$

Module Block

Input: $\mathrm{X}_T,\ X_R \in R^{L \times d}$
Output: $\mathrm{X}_T,\ X_R \in R^{L \times d}$
$\mathrm{X}_T,\ X'_R = SeriesDecomp(X_T)$
$\mathrm{X}'_R = TemporalAttn(Linear(X'_R))$
$\mathrm{X}_T = GRU(X_T)$
$\mathrm{X}_R = X'_R + X_R$

A. Sequence Decomposition

Sequence decomposition is a common method in time series data theory, which is done by mean filtering through a sliding window to get the trend signal, and then subtracting the input data from the trend data to get the residual data. Since it is difficult to capture the trend information completely by single mean smoothing, some trend information is still retained in the residual part, as well as many period information and residual volatility information [15]. Therefore, in this paper, the decomposition module is embedded as part of the basic module of the network, making it possible to achieve progressive decomposition by stacking the basic modules.

$$\begin{aligned} \boldsymbol{X}_T &= MovingAverage(\boldsymbol{X}), \\ \boldsymbol{X}_R &= \boldsymbol{X} - \boldsymbol{X}_T, \end{aligned} \tag{1}$$

where $\boldsymbol{X}$, $\boldsymbol{X}_T$, $\boldsymbol{X}_R$ are the original data, the trend term, and the residual volatility part, respectively, and there are $\boldsymbol{X}, \boldsymbol{X}_T, \boldsymbol{X}_R \in \mathbb{R}^{L \times d}$, L is the sequence length, and d is the number of features; $MovingAverage(\cdot)$ is to use sliding average for each feature separately, and in order to keep the sequence length unchanged, nearest neighbor padding is used before sliding average.

B. Gated Recurrent Unit

Gated Recurrent Unit (GRU) is an improved model of LSTM, the most popular current temporal model, which reduces computational cost and improves model efficiency and performance by combining the input gate (I_t) and forgetting gate (F_t) of LSTM into an updating gate (Z_t); and combining the long-term memory and short-term memory of LSTM into a hidden state, avoiding the complexity of maintaining two

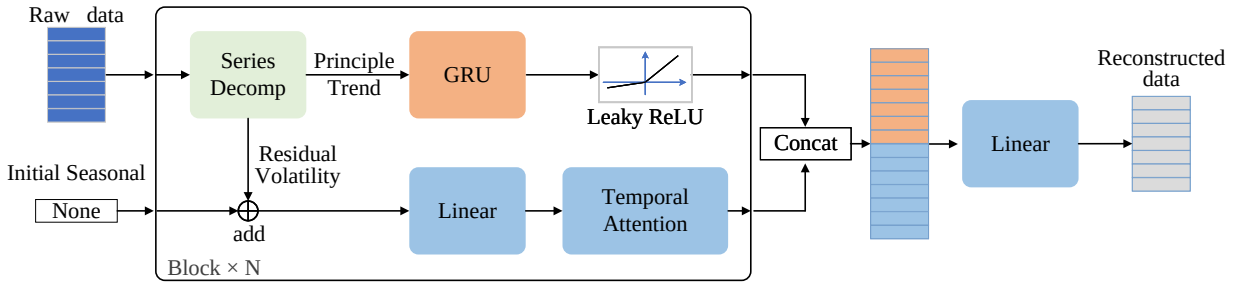

Fig. 1. Proposed fault diagnosis model

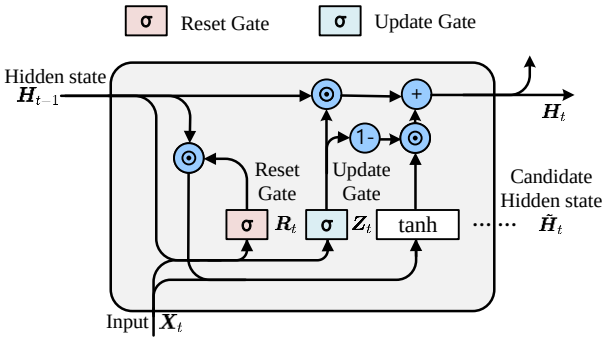

Fig. 2. Gated Recurrent Unit

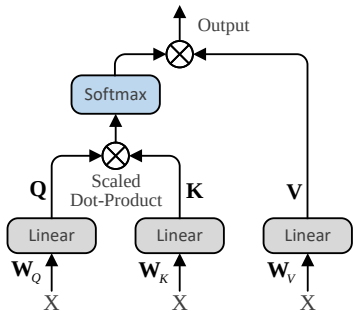

Fig. 3. Attention

states and reducing some problems such as gradient explosion and gradient dispersion. At the same time, GRU introduces a reset gate (R_t) to control the extent to which the historical information is retained.The overall model structure of GRU is shown in Fig. 2.

$$\begin{aligned}
\boldsymbol{R}_t &= \sigma\left(\boldsymbol{X}_t \boldsymbol{W}_{xr} + \boldsymbol{H}_{t-1} \boldsymbol{W}_{hr} + \boldsymbol{b}_r\right), \\
\boldsymbol{Z}_t &= \sigma\left(\boldsymbol{X}_t \boldsymbol{W}_{xz} + \boldsymbol{H}_{t-1} \boldsymbol{W}_{hz} + \boldsymbol{b}_z\right), \\
\tilde{\boldsymbol{H}}_t &= \tanh\left(\boldsymbol{X}_t \boldsymbol{W}_{xh} + \left(\boldsymbol{R}_t \odot \boldsymbol{H}_{t-1}\right) \boldsymbol{W}_{hh} + \boldsymbol{b}_h\right), \\
\boldsymbol{H}_t &= \boldsymbol{Z}_t \odot \boldsymbol{H}_{t-1} + \left(1 - \boldsymbol{Z}_t\right) \odot \tilde{\boldsymbol{H}}_t,
\end{aligned} \tag{2}$$

where $R_t, Z_t, \tilde{H}_t, H_t$ denote the output of the reset gate, the output of the update gate, the candidate hidden state, and the output hidden state (the model output), respectively; σ denotes the activation function, which is usually a $sigmoid$ function; $tanh$ is the hyperbolic tangent function, and $\odot$ is the element-level multiplication.

C. Temporal Attention

The attention mechanism mimics the human visual system or attentional mindset, enabling the model to focus on the important parts when processing information. The attention mechanism dynamically allocates attention to inputs at different temporal locations when processing temporal data, allowing it to better capture key information. This helps to improve the model's ability to understand and represent the input data, thus compensating for the shortcomings of time-series models such as GRU that have difficulty in capturing complex time-series patterns. The temporal attention structure used in this paper is shown in Fig. 3, and the calculation formula is shown in Eq. (3).

$$\begin{aligned}
&\boldsymbol{Q} = \boldsymbol{X} \times \boldsymbol{W}_q, \quad \boldsymbol{K} = \boldsymbol{X} \times \boldsymbol{W}_k, \quad \boldsymbol{V} = \boldsymbol{X} \times \boldsymbol{W}_v \\
&Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = softmax(\frac{\boldsymbol{Q} \times \boldsymbol{K}^{\boldsymbol{T}}}{\sqrt{d_k}})\boldsymbol{V},
\end{aligned} \tag{3}$$

where $\boldsymbol{X} \in \mathbb{R}^{L \times d}$ are the original data, L is the length of the sequence and d is the number of features, and since the self-attention mechanism is used here, the hidden features have the same dimensions, both d_k, and there is $\boldsymbol{W}_q, \boldsymbol{W}_k, \boldsymbol{W}_v \in \mathbb{R}^{d \times d_k}$.

D. Dynamic threshold

This paper introduces a dynamic threshold method based on the 3 rule by considering the consistency of battery cells within a battery pack. Initially, a fundamental threshold is established. When the average reconstruction loss L_{mean} of the battery pack voltage exceeds the basic threshold τ_{basic}, the dynamic threshold τ is calculated to replace the basic threshold using the Eq. (4) and (5).

$$\begin{aligned}
&L = |X - X^{rec}|, \ L \in R^{L \times d}, \\
&L_{cells} = Mean(L, dim = 0) \in R^{1 \times d}, \\
&L_{mean} = Mean(L_{cells}, dim = 1) \in R^{1 \times 1}, \\
&L_{std} = Std(L_{cells}, dim = 1) \in R^{1 \times 1},
\end{aligned} \tag{4}$$

where X represents voltage data collected by sensors, X^{rec} is the data reconstructed by the model, and L denotes their

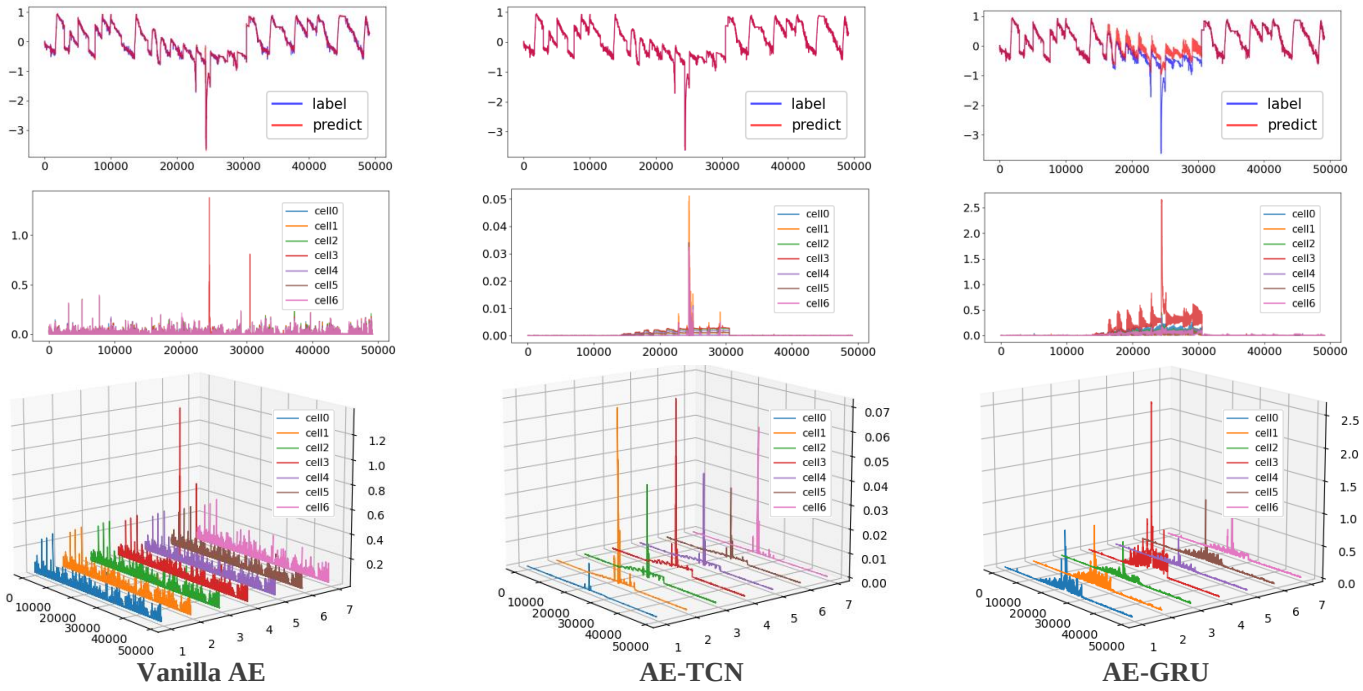

Fig. 4. Reconstruction results for vanilla AE, AE-TCN and AE-GRU. The first row displays a comparative chart of original and reconstructed data, while the second and third rows depict the plane figure and the three-dimensional figure of the reconstruction loss, respectively.

absolute error. $L_{cells} = [L_{cell1}, L_{cell2}, ..., L_{celld}]$ represents the average reconstruction error on individual battery cells, with d being the number of battery cells. Meanwhile, L_{mean} and L_{std} respectively signify the mean and standard deviation of the reconstruction errors among individual battery cells.

$$\tau = \begin{cases} L_{\text{mean}} + 3 \times L_{\text{std}}, & \text{if } L_{\text{mean}} \geq \tau_{\text{basic}} \\ \tau_{\text{basic}}, & \text{if } L_{\text{mean}} < \tau_{\text{basic}} \end{cases} \tag{5}$$

Finally, by comparing L_{cells} with a dynamic threshold, to determine whether a fault has occurred in battery cells.

III. Experiment and Analysis

A. Battery Dataset

The battery used in this paper for the quadrupedal robot consists of seven battery cells connected in series with a charging limiting voltage of DC 33.6v and charging currents ranging from 3.5A to 9A. The main parameters of the battery packs are shown in Table I. The dataset used in this paper was obtained from five quadrupedal robots operating for five months, where two of the battery packs produced failures, one pack exhibited a relatively slight undervoltage, and the other pack exhibited a significant voltage drop.

B. Data preprocessing

For the real data collected, environmental conditions or sensor errors may result in large noise in the data, which may mislead the fault diagnosis algorithm. Therefore, in this paper, we adopt a sliding window approach to preprocess the raw data to remove outliers caused by environmental or sensor errors.

TABLE I
Battery Pack Profiles

Parameter name	Parameter information
Rated voltage of battery cell	4.2v
Rated voltage of battery pack	DC 28.8v
Battery pack rated capacity	15000mAh, 432Wh

C. Method comparison and analysis

To verify the superiority of the method proposed in this paper, we compared it with vanilla AE, AE-GRU, AE-TCN, AE-CrossAttention, and AE-TemporalAttention. The final experimental results are shown in Fig. 4 and Fig. 5, and the related data are listed in Table II.

Due to the limitation of the length of the article, we only show the reconstruction results on the No.9 battery pack. In Fig. 4 and Fig. 5, each column represents the results of one model. The first row shows the result of the reconstruction on cell 4 of battery pack No.9 (cell 4 is the one that fails). The second row shows a 2D plan view of the reconstruction errors for all battery cells on pack No.9, while the third row shows the corresponding 3D figure.

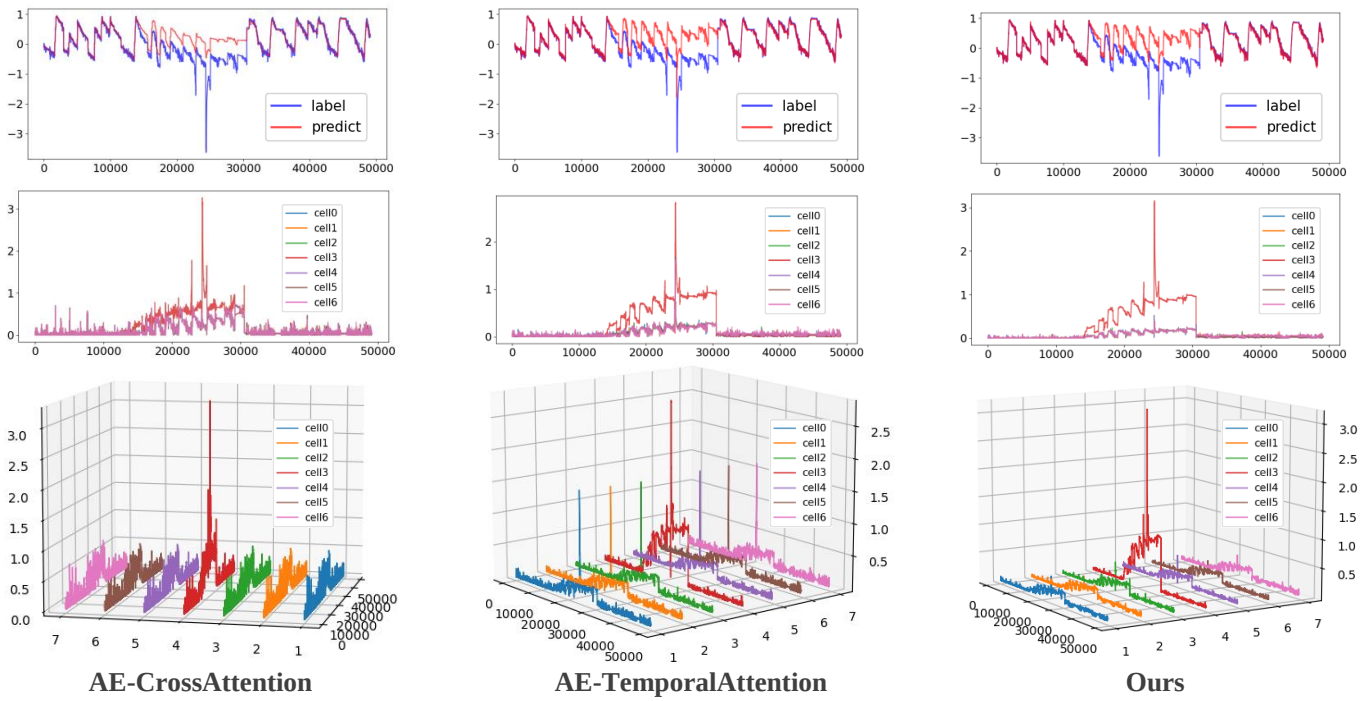

Fig. 5. Reconstruction results for AE-CrossAttention, AE-TemporalAttention and proposed model. The first row displays a comparative chart of original and reconstructed data, while the second and third rows depict the plane figure and the three-dimensional figure of the reconstruction loss, respectively.

TABLE II
THE COMPARISON OF DIFFERENT METHODS

	Battery Pack Number	Vanilla AE	AE-TCN	AE-GRU	AE-CrossAttn	AE-TemporalAttn	Proposed
Reconstruction Loss (Test)	No.9 No.10	3.907 e-4 2.424 e-4	2.242 e-7 2.260 e-8	1.577 e-3 1.574 e-4	1.633 e-2 5.574 e-3	1.005 e-2 2.273 e-3	9.303 e-3 1.953 e-3
False Alarm Rate (Mis/Correct diagnosis)	No.9 No.10	0/1 0/0	0/0 0/0	90/82 0/0	713/225 967/400	6/163 0/250	0/162 0/236
Training Time	No.1~No.6	12.5*29	29.6*12	20.1*17	47.6*47	32.4*24	63.2*9
Reconstruction Loss (Validation)	No.7~No.8	3.514 e-4	1.696 e-10	5.688 e-7	2.873 e-3	2.786 e-4	4.863 e-5
Model parameters		hid_size=28 —	hid_size=28 —	— hid_len=28	hid_size=96 hid_len=64	hid_size=28 —	kernel_size=3 block_num=3

It can be seen that all the models used in this paper are able to effectively learn the features of time-series data and reconstruct the original normal data. However, AE-GRU shows higher sensitivity to abnormal data, and its reconstruction error for abnormal data is relatively large as shown in Fig 4. In contrast, vanilla AE and AE-TCN exhibit better reconstruction ability for abnormal data. This is due to that vanilla AE and AE-TCN are more capable in capturing localized information, whereas GRU focuses more on learning the long-time dependencies of the data and maintains these relationships during the reconstruction process. However, the memory units of GRU are susceptible to faulty data, which leads to changes in the long-time dependencies, resulting in large differences between the reconstructed data and the original data.Therefore, GRU is more suitable as a module for fault diagnosis of temporal data than the convolution-based TCN network and vanilla AE using only full connectivity.

In addition, this paper adds the attention mechanism from the current superior Transformer model into the model, the attention approach is different from the above methods which allows each part of the input to capture the information globally. As shown in Fig. 5, both AE-TemporalAttention and

AE-CrossAttention can detect faults, but Cross Attention can misidentify multiple batteries under the fault time as faulty batteries, while TemporalAttention can accurately identify a faulty battery cell. This is because TemporalAttention learns the key information on the timing, while Cross Attention considers the relationship of different battery cells in addition to learning the relationship on the timing, which will affect the accuracy of localizing the faulty cell.

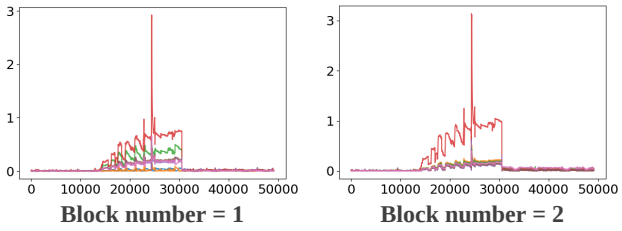

Fig. 6. The effect of model reconstruction with different number of blocks.

By comparing with the most effective AE-GRU and AE-TemporalAttn, the method proposed in this paper can clearly distinguish normal batteries from faulty batteries, and the reconstruction error on normal batteries is much smaller, which can be used to detect faults earlier for replacement and maintenance by adjusting the fault diagnosis threshold.

In addition, this paper also compares the effect of the number of modules on the performance of the proposed model, as shown in Fig. 6. In the third column of Fig. 5, the effect of stacking the three-layer model is demonstrated. With stacking the GRU-TemporalAttention module with time series decomposition capability, it is able to distinguish the faulty battery cell from the normal batteries more distinctly, and the reconstruction error on the normal battery is lower. Therefore, the module designed in this paper with the attribute of time series decomposition shows better results and is able to mine the temporal characteristics more effectively, which leads to more accurate fault diagnosis.

To ensure the fairness of the experiments in this paper, we uniformly used an input window of length 96 and used the voltage data of the seven battery cells of the battery pack as inputs. All models were experimented with hidden layer dimensions of 96, 64, and 28 settings, respectively, and for a fair comparison, we selected the results of each model that performs best under these dimensions.

IV. Conclusion

Battery fault diagnosis for quadrupedal robots is crucial to ensure their safe and efficient operation. However, due to factors such as motion characteristics, the collected battery data often shows large fluctuations, which can easily lead to misjudgment by the fault diagnosis algorithm. In this paper, by embedding the temporal decomposition method into the network module, introducing the GRU to learn the temporal trend, and combining the attention mechanism to capture the key information dynamically, the model is equipped with the ability to decompose and mine the deeper information progressively, which improves the accuracy and reliability of fault diagnosis. The experimental comparison results show that this paper's method significantly outperforms multiple fault diagnosis methods based on the auto-encoder structure in terms of fault diagnosis effectiveness, and the misdiagnosis rate is lower than that of the effective AE-GRU model. In summary, the method in this study provides an accurate and reliable solution for the fault diagnosis of quadrupedal robot batteries, which provides a strong support for improving the safety and performance of robotic systems.

References

[1] D. McNulty, A. Hennessy, M. Li, E. Armstrong, and K. M. Ryan, "A review of li-ion batteries for autonomous mobile robots: Perspectives and outlook for the future," *Journal of Power Sources*, vol. 545, p. 231943, 2022.

[2] O. N. Şahin and M. İ. C. Dede, "Model-based detection and isolation of the wheel slippage and actuator faults of a holonomic mobile robot," *Industrial Robot: the international journal of robotics research and application*, vol. 49, no. 6, pp. 1202–1217, 2022.

[3] J. Zhao, K. Zhang, M. Hou, H. Zhang, Y. Bai, Y. Huang, and J. Li, "Actuator fault detection for masonry robot manipulator arm with the interval observer," *Journal of Field Robotics*, vol. 40, no. 2, pp. 147–160, 2023.

[4] A. H. Sabry, F. H. Nordin, A. H. Sabry, and M. Z. A. Ab Kadir, "Fault detection and diagnosis of industrial robot based on power consumption modeling," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7929–7940, 2019.

[5] Z. Miao, F. Zhou, X. Yuan, Y. Xia, and K. Chen, "Multi-heterogeneous sensor data fusion method via convolutional neural network for fault diagnosis of wheeled mobile robot," *Applied Soft Computing*, vol. 129, p. 109554, 2022.

[6] P. Kumar, S. Khalid, and H. S. Kim, "Prognostics and health management of rotating machinery of industrial robot with deep learning applications—a review," *Mathematics*, vol. 11, no. 13, p. 3008, 2023.

[7] M. Kasap, M. Yılmaz, E. Çinar, and A. Yazıcı, "Unsupervised dissimilarity-based fault detection method for autonomous mobile robots," *Autonomous Robots*, pp. 1–16, 2023.

[8] H. Yun, H. Kim, Y. H. Jeong, and M. B. Jun, "Autoencoder-based anomaly detection of industrial robot arm using stethoscope based internal sound sensor," *Journal of Intelligent Manufacturing*, vol. 34, no. 3, pp. 1427–1444, 2023.

[9] B. Xia, K. Wang, A. Xu, P. Zeng, N. Yang, and B. Li, "Intelligent fault diagnosis for bearings of industrial robot joints under varying working conditions based on deep adversarial domain adaptation," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[10] K. Lu, C. Chen, T. Wang, L. Cheng, and J. Qin, "Fault diagnosis of industrial robot based on dual-module attention convolutional neural network," *Autonomous Intelligent Systems*, vol. 2, no. 1, p. 12, 2022.

[11] T. Wang, L. Zhang, and X. Wang, "Fault detection for motor drive control system of industrial robots using cnn-lstm-based observers," *CES Transactions on Electrical Machines and Systems*, 2023.

[12] J. Jiang, T. Li, C. Chang, C. Yang, and L. Liao, "Fault diagnosis method for lithium-ion batteries in electric vehicles based on isolated forest algorithm," *Journal of Energy Storage*, vol. 50, p. 104177, 2022.

[13] H. Zhang, J. Hong, Z. Wang, and G. Wu, "State-partial accurate voltage fault prognosis for lithium-ion batteries based on self-attention networks," *Energies*, vol. 15, no. 22, p. 8458, 2022.

[14] H. Zhao, Z. Chen, X. Shu, J. Shen, Y. Liu, and Y. Zhang, "Multi-step ahead voltage prediction and voltage fault diagnosis based on gated recurrent unit neural network and incremental training," *Energy*, vol. 266, p. 126496, 2023.

[15] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.