A Method for Hydrogen Leakage Traceability Based on Residual Neural Networks

Xiaopeng Wang [1], Hui Zhang [1], Jiaxin Zhang [2], Xuefang Li [2], Bo Sun [1]

[1] School of Control Science and Engineering, Shandong University, Jinan, China
[2] Institute of Thermal Science and Technology, Shandong University, Jinan, China

Abstract. Hydrogen leakage traceability is a key technology to ensure the safety and stability of hydrogen in the whole process of production, storage, transportation and usage. Traditional machine learning methods require manual processing of data features, which makes it difficult to cope with the high-dimensional features of multi-source hydrogen sensor data, resulting in complex model adjustments and a lack of generalization. Therefore, this study proposed a hydrogen leak tracing method based on Residual Neural Networks to improve the accuracy and efficiency of hydrogen leak tracing. Firstly, the multi-source hydrogen sensor data is combined into gray level maps and pre-processed. Then, the Residual Neural Networks model is constructed as the backbone network to adaptively extract the image features of hydrogen concentration and perform the traceability task. Finally, the effectiveness and advancement of the proposed method are tested on actual sensor signals. The comparative experimental results show that compared with other models, the Residual Neural Networks model exhibits better generalization and accuracy, with an average F1 score of 87.1%.

Keywords: Hydrogen energy; hydrogen safety; hydrogen leakage traceability; feature extraction.

1. Introduction

With the transformation of the global energy structure and the increasing demand for clean energy [1], As an efficient and clean energy carrier, the importance of hydrogen energy development is self-evident. Hydrogen not only has extremely high energy density, but also produces almost no greenhouse gases during combustion, making it an important component of the future energy system. However, with the rapid development of hydrogen energy industry, safety issues have become an important factor restricting its large-scale application [2]. Hydrogen often exists in the form of high pressure in each link of production, storage, transportation and usage, and its unique physical and chemical properties lead to its easy leakage [3]. In addition, hydrogen has a high diffusion coefficient and a wide range of flammable concentrations, which means that once hydrogen leaks and encounters an ignition source, it can evolve into a combustion or even explosion accident, endangering human life and property safety. Before the hydrogen leakage accident occurs, the rapid and accurate detection and identification of the leak location and fault, as well as taking corresponding emergency measures, will directly contribute to the further deterioration of the accident and save lives and property safety. Therefore, the detection, identification and traceability research of hydrogen leakage is crucial to ensure the safety of the whole hydrogen energy industry chain and promote the sustainable development of hydrogen energy [4].

At present, studies on hydrogen leakage source traceability have been widely reported [5]. Wang et al. [6] proposed fault diagnosis of hydrogen sensor based on wavelet singular entropy and correlation vector machine to solve the problem of fault diagnosis of hydrogen sensor under nonlinear and small sample conditions. Sun et al. [7] proposed a study on the leakage risk of hydrogen pipeline based on the dynamic Bayesian model to study the dynamic leakage risk of hydrogen pipeline. Zou [8] et al. proposed a hydrogen leakage detection method of fuel cell engine based on support vector machine, and used particle swarm optimization to optimize radial basis kernel function support vector machine to identify hydrogen leakage in the system. In [9], a data-driven high-pressure hydrogen leak diagnosis method based on actual driving conditions and probabilistic neutral network was proposed. However, the above machine learning method requires manual processing of data characteristics, which is difficult to deal with high-dimensional hydrogen concentration sensor data.

The data-driven method based on deep learning (DL) [10] has good adaptive feature extraction capability and reduces the cost of manual feature extraction and expert physical modeling. At present, deep learning is widely used in tasks such as fault diagnosis [11], pattern recognition [12], and automatic driving [13]. In terms of hydrogen leak traceability, many researchers have carried out research on deep learning methods. Yang [14] et al. proposed a deep learning method to predict the location and intensity of hydrogen leakage at hydrogen refueling stations, and used a deep learning hybrid model to predict the location and intensity of hydrogen leakage. In [15], CEEMDAN-CNN-LSTM model was proposed to predict the location of hydrogen leakage in order to reduce the noise problem existing in the data. The above methods demonstrate that deep learning can be widely applied to trace hydrogen gas leaks. However, the traceability accuracy obtained by the above methods is unsatisfactory. In order to improve traceability accuracy, this study provides a new solution from the perspective of image processing. Specifically, a hydrogen leakage tracing method based on residual network (Residual Neural Networks) is proposed. By mapping multi-source hydrogen sensor data into pixel values to form an image, introducing residual connections makes it easier for the network to learn identity maps, allowing the residual network to fully utilize feature information at different levels, Enhancing the learning ability and generalization of the network, enabling better utilization of hydrogen leak data for precise localization and traceability.

2. Problem description

In this study, the simplified model of fuel cell tank was divided into structured grids, and the hydrogen concentration at each monitoring point after leakage was obtained by numerical simulation. In order to train the multi-layer neural network, hydrogen sensor concentration data samples with different leakage locations are generated, and the leakage location of each sample is marked. To maintain universality, P is used to represent the number of data channels of sensors in the hydrogen safety monitoring system, and S is used to represent the number of sampling points for each channel. Then arrange each sample of the multi hydrogen concentration sensor data in sequence, and finally represent it as a matrix:

$$x = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,S} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ x_{P,1} & x_{P,2} & \cdots & x_{P,S} \end{bmatrix} \tag{1}$$

The goal of hydrogen leak traceability is to get a parametric model, $M(\theta;\cdot)$, mapping multi-source hydrogen sensor concentration data to specific leak locations.

$$[y^x, y^y] = M(\theta;\cdot) \tag{2}$$

where $x \in R^{P\times S}$, y^x, y^y representing the x coordinates and y coordinates of the leak location respectively, θ is the parameter of M. In addition, $y^x \in N_x$, $y^y \in N_y$. This research should address how to maintain the accuracy of hydrogen leak source tracking in the face of the lack of information in the target region.

3. Method description

3.1 Data preprocessing

Assuming obtain unlabeled multi-source hydrogen sensor sequence data from the monitoring system, denoting it as $X_p = \{x_i \mid i = 1, 2, \ldots, N\}$, where N represents the number of samples. In this study, for sample i, $x_i \in R^{P\times S}$ is seen as a grayscale image, each value in the matrix is a pixel. And the entire grayscale image x_i is normalized. Linear normalization maps the original hydrogen leakage data linearly to the interval [0,1] :

$$x_i^{\text{norm}} = \frac{x_i - x^{\min}}{x^{\max} - x^{\min}} \tag{3}$$

where x_i is the original data sample, $x^{\min}$ is the minimum value of this feature in the sample dataset, $x^{\max}$ is the maximum value of this feature in the sample dataset, x_i^{norm} is normalized data.

3.2 Network architecture

Extracting features from multi-source hydrogen concentra tion sequence data using network model based on Residual Neural Networks, its network structure is shown in Fig. 1.

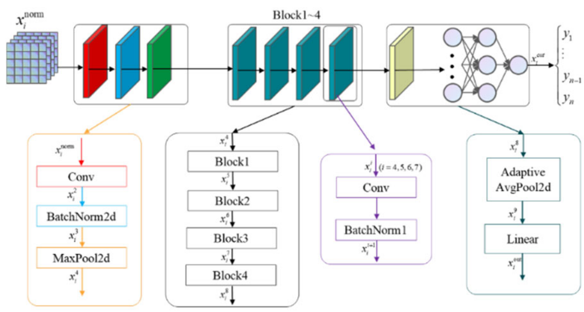

Fig. 1 Network architecture of hydrogen leak tracing method based on Residual Neural Networks

3.2.1 *Convolutional Layer*

After processing the original leaked data, it is expressed as x_i^{norm}, performing a two dimensional convolution operation on the input to obtain the output:

$$x_i^2 = \sum_{u=0}^{6}\sum_{v=0}^{6} I_{in}(x\cdot \text{stride} + \text{u - padding}, \text{y}\cdot\text{stride+v-padding}) \cdot k_c(u,v) \tag{4}$$

where $\text{stride} = 2$ is the convolution kernel that slides the step size of the input image, $\text{padding} = 3$ is the number of zero padding added around the input image, I_{in}

represents the input image tensor, $\mathrm{K_c}$ represents the output channel convolution kernel. Through the above operation, obtain the output feature map with a specific number of output channels and a specific space size.

3.2.2 Batch normalized layers

Convolutional layer output $x_i^2 = \{x_{b,c,h,\omega}\}$ as the layer input of batch normalization, where, b represents batch size, c represents the number of channels, h represents the height of the feature map, ω represents the width of the feature map.

$$\mu_c = \frac{1}{bh\omega}\sum_{b=1}^{B}\sum_{h=1}^{H}\sum_{\omega=1}^{W} x_{b,c,h,\omega} \tag{5}$$

$$\sigma_c^2 = \frac{1}{bh\omega}\sum_{b=1}^{B}\sum_{h=1}^{H}\sum_{\omega=1}^{W} (x_{b,c,h,\omega} - \mu_c)^2 \tag{6}$$

$$\hat{x}_{b,c,h,\omega} = \frac{x_{b,c,h,\omega} - \mu_c}{\sqrt{\sigma_c^2 + \varepsilon}} \tag{7}$$

$$x_i^3 = y_{b,c,h,\omega}^3 = \gamma_c \cdot \hat{x}_{b,c,h,\omega} + \beta_c \tag{8}$$

Linear transformation to obtain output x_i^3, γ_c and β_c are Learn-able parameters, ε is very small to prevent the denominator from being 0.

3.2.3 Maximumpooling layer

Batch normalized output x_i^3 as a two-dimensional maximum pooling layer input, the maximum pool layer output x_i^4 is:

$$x_i^4 = y_{h,\omega,c}^4 = \max_{i=h\times s}^{h\times s+k-1} \max_{j=\omega\times s}^{\omega\times s+k-1} x_{i,j,c}^3 \tag{9}$$

where, h represents the feature map height, ω represents the feature map width, c represents the number of channels in the feature map, slide according to the pool window size and step size in the direction of height and width, and select the maximum value in each window as the output.

3.2.4 Repeat residual block layer

Maximumpooled layer output x_i^4 as input to the four residual block layers, After under going residual processing and establishing residual connections, it serves as the input for subsequent modules.

$$x_i^5 = \mathrm{Block}(x_i^4) \tag{10}$$

$$x_i^6 = \mathrm{Block}(x_i^5) \tag{11}$$

$$x_i^7 = \mathrm{Block}(x_i^6) \tag{12}$$

$$x_i^8 = \mathrm{Block}(x_i^7) \tag{13}$$

The residual block output x_i^8 is obtained by formula (10-13).

3.2.5 3.2.5 Adaptive average pooling layer

Convert the input feature map processed from leaked data into a fixed length vector:

$$x_i^9 = y_{h,\omega,c}^9 = \frac{1}{k_h \times k_\omega}\sum_{i=h\times s_h}^{h\times s_h+k_h-1}\ \sum_{j=\omega\times s_\omega}^{\omega\times s_\omega+k_\omega-1} x_{i,j,c}^8 \tag{14}$$

where h represents the feature map height, ω represents the feature map width, c represents the number of channels in the feature map, k_h, k_ω represents the size of the pooling window in both height and width directions, s_h, s_ω represents the step size.

Through this layer operation, the sensitivity of the model to the input feature map size is reduced and the generalization ability of the model is improved.

3.2.6 3.2.6 Fully connected layer

Integrate the above extracted leakage data features and map them to the leakage target space:

$$x_i^{\text{out}} = x_i^9 \cdot \omega + \mathrm{b} \tag{15}$$

where ω is weight matrix, b is bias term, obtain the mapped leak location information to complete the leak tracing.

3.3 Data-driven learning

Regarding the Residual Neural Networks network model architecture proposed above, providing detailed steps for model learning, see Algorithm 1. Firstly, the hydrogen leakage dataset is divided into randomly assigned training sets, input the generated grayscale image of hydrogen concentration into the Residual Neural Networks network for a series of operations to obtain the predicted leak location, the model parameters were updated according to the backpropagation gradient descent to minimize the loss L.

$$L = -\sum_{i}^{C} y_i \log(\hat{y}_i) \tag{16}$$

where $\hat{y}_i$ is the predicted location, y_i is the actual location.

Algorithm 1 The proposed hydrogen leakage traceability method based on Residual Neural Networks.

Require: D : dataset composed of gray level images of hy drogen concentration collected by unlabeled multi-source hydrogen sensor; $D=\{(x_i,y_i)\mid i=1,\ldots,N\}$, $x_i,y_i\in R^{P\times S}$; N_{bs} : the mini-batch size; K : the number of iterations; θ : model parameter, is a parameter list, represents all learnable parameters in the above content, such as γ_c , β_c ; Adam optimizer;

Initialize: The model parameter θ is randomly initialized;

Compute:

1: for $k=1,2,\ldots,K$ do

2: Randomly select N_{bs} small batch training samples from D , $d=\{(x_i,y_i)\mid i=1,2,\ldots,N_{bs}\}\in D$

3: for $i=1,2,\ldots,N_{bs}$ do

4: Preprocess X_i to obtain x_i^{norm} using formula(3);

5: $\hat{y}_i\leftarrow \text{Model}(\theta;x_i)$ obtain the predicted location of hydrogen leakage $\hat{y}_i$;

6: Calculate the loss function value L of data-driven learning using formula(16);

7: The gradient parameter is reset to zero, $g_\theta\leftarrow 0$;

8: Calculate parameter gradient, $g_\theta\leftarrow\dfrac{\partial L}{\partial\theta}$;

9: Calculate the mean, $m_k=\beta_1 m_{k-1}+(1-\beta_1)\cdot g_\theta$, β_1 set the value of 0.9;

10: Calculate non central variance, $v_k=\beta_2\cdot v_{k-1}+(1-\beta_2)\cdot g_k^2$, β_2 set the value of 0.999;

11: Bias correction, $\hat{m}_k=\dfrac{m_k}{1-\beta_1^k}$, $\hat{v}_t=\dfrac{v_k}{1-\beta_2^k}$;

12: Update parameter θ_k , $\theta_k=\theta_{k-1}-\alpha\cdot\dfrac{\hat{m}_k}{\sqrt{\hat{v}_t}+\varepsilon}$

13: **end for**

14: **end for**

4. Experiment

4.1 Comparison methods and indicators

4.1.1 Comparison method

In order to fully evaluate the performance of the proposed Residual Neural Networks-based hydrogen leak traceability method, comparing Residual Neural Networks to some common machine learning methods, which are Multi Layer Perceptron (MLP), Wavelet Discrete Transform (K-DWT), Support Vector Machine (SVM) and Extreme Learning Machine (ELM). In the comparative experiment, the original leakage concentration data were smoothed and normalized.

4.1.2 Evaluation indicators

To accurately measure the superior performance of the Residual Neural Networks-based hydrogen leak traceability method proposed in this study, four different performance indicators were used in the comparative experiment to comprehensively measure the accuracy of training and testing, which are Accuracy, Precision, Recall and F1 score.

4.2 The result of the proposed Residual Neural Networks

In order to effectively help us understand the learning bias of the proposed method at different stages, the loss changes during the training and testing processes were plotted. As shown in Fig. 2. both training losses and test losses can decrease gradually, while training losses decline rapidly at the beginning, indicating that the Residual Neural Networks model can effectively learn valuable rule features from the trained hydrogen concentration gray scale image data of hydrogen leakage, which also means that the leakage data and the model have a good fit. In addition, Adam as an adaptive optimization algorithm, can dynamically update the learning rate, allowing the method to quickly decrease in the gradient direction of the loss function.

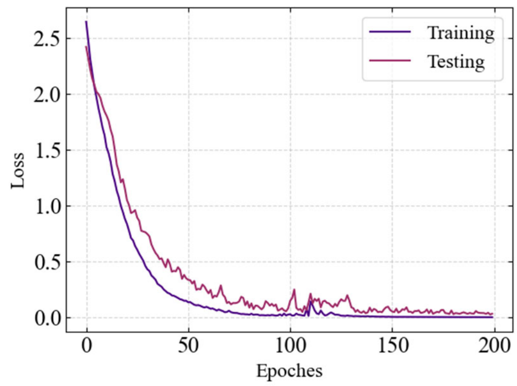

Fig. 2 Training and test loss results

Although the test loss decreased more slowly than the training loss at the beginning of the iteration, this meant that the distribution of the trained and tested hydrogen sensor concentration data was quite different, as shown in Fig. 2. However, at the later stage of training, the proposed method can be well adapted to this difference, so the test loss can be rapidly reduced, which indicates that the proposed method has good generalization.

4.3 Compare the results and analysis

A comprehensive evaluation was conducted on the performance of six methods, including the proposed method, in hydrogen leak tracing tasks, and specific evaluation indicators were provided, as shown in Table 1 and Fig. 3. The proposed method achieved the best performance in hydrogen leak tracing task, with an F1 score of 87.1%.

As is shown in Table 1, deep learning-based methods are often superior to machine learning-based methods. For example, the Accuracy index of LSTM and Residual Neural Networks methods can reach 100% in training, while K-DWT, ANN, SVM and ELM can only reach 60%-90%, because the powerful adaptive feature extraction capability of deep learning is crucial to accurately identify information about leaks.

The Precision index of Residual Neural Networks is the highest among all learning methods, that is, the proportion of positive samples is the largest, which means that the proposed method has superior test performance. The Recall index and F1 score fully represent the model's ability to find leakage data and significant generalization ability, that is, it can obtain satisfactory positioning accuracy for untrained hydrogen leakage data.

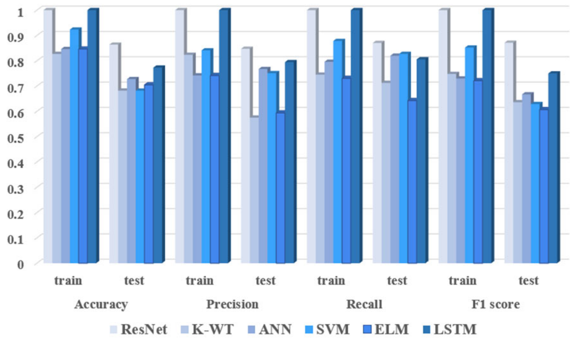

Fig. 3 Performance indicators of different comparison methods

Table 1. Performance index

		K-DWT	ANN	SVM	ELM	LSTM	Residual Neural Networks
Accuracy	train	0.8654	0.9423	0.8654	0.8462	1.0000	1.0000
	test	0.6818	0.7272	0.6363	0.7272	0.7727	0.8636
Precision	train	0.8349	0.7034	0.7058	0.7285	1.0000	1.0000
	test	0.6316	0.7906	0.8034	0.7044	0.7890	0.7738
Recall	train	0.7971	0.7811	0.7742	0.7581	1.0000	1.0000
	test	0.7750	0.7698	0.8132	0.7241	0.8612	0.8469
F1 score	train	0.7976	0.7332	0.7283	0.7344	1.0000	1.0000
	test	0.7068	0.7161	0.8788	0.7068	0.7540	0.7921

Fig. 4 Radar chart of the performance indicators of all the comparison algorithms

5. Summary

In the research on hydrogen leak traceability, traditional methods based on machine learning are faced with technical challenges, that is, the high dimension of hydrogen sensor concentration data leads to difficult feature extraction and poor generalization. To solve this problem, this study proposes a Residual Neural Networks based hydrogen leakage tracing method. The main task of this research is to solve the problem of multi-source sensor data fusion and high-dimensional data feature extraction, and use feature mapping to achieve accurate localization. Specifically, a hydrogen leak tracing method based on Residual Neural Networks was proposed and applied to tracing tasks in practical application scenarios, achieving superior accuracy. The F1 score of the comprehensive positioning index can reach 87.1%. Compared with machine learning based methods, Residual Neural Networks can demonstrate superior generalization performance in unknown hydrogen leakage data.Therefore, this study demonstrates the effectiveness and efficiency of the proposed hydrogen leak tracing method, which is beneficial for the development of hydrogen safety. These studies are expected to promote hydrogen safety research and the development of the hydrogen energy industry.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant number: 2023YFB4004500).

References

1. Sun Han, Meng Zhenghao, Li Meilin, Yang Xiaohui, Wang Xiaoxue. Global supply sustainability assessment of critical metals for clean energy technology. Resources Policy, 2023. DOI: 10.1016/J.RESOURPOL.2023.103994
2. K. Harri, W. Achim, D. Georgios, C. Neil, J. Daniel, H. I. Young, and F. Sahand. Industry rd needs in hydrogen safety. Applications in Energy and Combustion Science, 2024, 18.
3. W. Ke, L. Changjun, J. Wenlong, C. Yong W. Jie. Under-expanded jet and diffusion characteristics for small-hole leakage of hydrogen-blended natural gas in high-pressure pipelines. Process Safety and Environmental Protection, 2024, 190: 195-211. DOI: 10.1016/J.PSEP.2024.07.045
4. K. Mahdi, M. Francesco, and A. M. Alessandra. Fault detection and diagnosis methods for green hydrogen production: A review. International Journal of Hydrogen Energy, 2022, 47(65): 27747–27774.
5. Zhao Mingbin, Huang Teng, Liu Chenhui, Chen Mingjia, Ji Shui, D. M. Christopher, Li Xuefang. Leak localization using distributed sensors and machine learning for hydrogen releases from a fuel cell vehicle in a parking garage. International Journal of Hydrogen Energy, 2021,46(1): 1420-1433. DOI: 10.1016/j.ijhydene.2020.09.218
6. Wang Bing, Diao Ming Song Kai. Fault diagnosis of hydrogen sensor based on wavelet singular entropy and relevance vector machine. Electric Machines and Control, 2015, 19(1): 96–101.
7. Sun Dongliang Chen Shiyue. Research on the leakage risk of hydrogen pipelines based on dynamic bayesian model. Journal of Safety and Environment, 2022, 22(128): 901–909. DOI: 10.13637/j.issn.1009-6094.2020.1762
8. Zou Qiang, Tian Ying, Li Hongsong, Qin Shunshun. Hydrogen leakage detection method for fuel cell engine based on support vector machine. Journal of Beijing Jiaotong University, 2020, 44(1): 84–90.
9. Li Jianwei, Weng Hongda, Yang Qingqing, Shen Jun. Data-driven diagnosis method of high-pressure hydrogen leakage based on actual driving conditions and probabilistic neutral network. International Journal of Hydrogen Energy. 2024, 71: 411–421. DOI: 10.1016/J.IJHYDENE.2024.04.204
10. Lei Jiawei, Xu Chengkan, Liu Chaofeng, Qin Feng, Zhang He. A data driven prediction for concrete crack propagation path based on deep learning method. Case Studies in Construction Materials, 2024, 21.
11. H. Ahmad, Cheng Wei, Xing Ji, Wang Wentao, Du Shuhong et al. Deep learning-based fault diagnosis of planetary gearbox: A systematic review. Journal Of Manufacturing Systems, 2024, 77: 249–266. DOI: 10.1016/J.JMSY.2024.10.004
12. Z. Amiri, A. Heidari, N. Jafari, M. Hosseinzadeh. Deep study on autonomous learning techniques for complex pattern recognition in interconnected information systems. Computer Science Review, 2024, 54.
13. F. Wang. Automatic driving decision model based on deep reinforcement learning algorithms in complex environments. 2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), 2024, 187–91.
14. Yang Guodong, Kong Depeng, He Xu, Yu Xirui Jiang Kai. Prediction of hydrogen leakage location and intensity in hydrogen refueling stations based on deep learning. International Journal of Hydrogen Energy, 2024, 68: 209–220. DOI: 10.1016/J.IJHYDENE.2024.04.234
15. Bi Yubo, Wu Qiulan, Wang Shilu, Shi Jihao, Cong Haiyong, Ye Lili, Gao Wei, Bi Mingshu. Hydrogen leakage location prediction at hydrogen refueling stations based on deep learning. Energy, 2023, 284. DOI: 10.1016/J.ENERGY.2023.129361