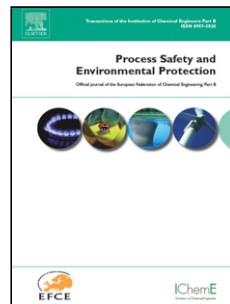


Journal Pre-proof

Enhanced Spectrum Convolutional Neural Architecture: An Intelligent Leak Detection Method for Gas Pipeline

Fangli Ning, Zhanghong Cheng, Di Meng, Shuang Duan, Juan Wei



PII: S0957-5820(20)31938-8

DOI: <https://doi.org/10.1016/j.psep.2020.12.011>

Reference: PSEP 2630

To appear in: *Process Safety and Environmental Protection*

Received Date: 8 September 2020

Revised Date: 7 December 2020

Accepted Date: 8 December 2020

Please cite this article as: Fangli Ning, Zhanghong Cheng, Di Meng, Shuang Duan, Juan Wei, Enhanced Spectrum Convolutional Neural Architecture: An Intelligent Leak Detection Method for Gas Pipeline, <![CDATA[*Process Safety and Environmental Protection*]]> (2020), doi: <https://doi.org/10.1016/j.psep.2020.12.011>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Process Safety and Environmental Protection
Enhanced Spectrum Convolutional Neural Architecture: An Intelligent Leak Detection
Method for Gas Pipeline
--Manuscript Draft--

Manuscript Number:	PSEP-D-20-00917R2
Article Type:	SI: Data-Driven Risk - Full Length Article
Keywords:	Spectrum enhancement; Convolutional neural network; Leak detection
Corresponding Author:	Fangli Ning CHINA
First Author:	Fangli Ning
Order of Authors:	Fangli Ning Zhanghong Cheng Di Meng Shuang Duan Juan Wei
Manuscript Region of Origin:	Asia Pacific
Abstract:	In this work, a novel convolutional neural architecture (SE-CNN), which combines Spectrum Enhancement (SE) and Convolutional Neural Network (CNN), is proposed to detect the leak of gas pipeline. The SE has the effect of enhancing the leak signals and reducing background noise. CNN can automatically extract leak features and realize leak diagnosis. The experimental results show that the SE-CNN can achieve an average accuracy of 94.3% for 6 categories and only requires 1.04 seconds of detection time. In this experiment, the diameters of the main pipeline and the branch pipeline are 125 mm and 25 mm. Due to its excellent accuracy and efficiency, the proposed enhanced spectrum convolutional neural architecture paves the way for real-time leak detection in industrial environments, which can ensure the process safety of gas pipeline transportation. Under strong background noise, the average accuracy of the SE-CNN can reach 94.3%, which is 33%, 3.7% higher than that of SVM and CNN. In particular, the SE can be regarded as a data compression method, which can significantly reduce the original data size. The training time of the SE-CNN is 539 seconds, reducing 90.6% compared with CNN.
Suggested Reviewers:	Zhigang Chu, Ph.D Chongqing University zgchu@cqu.edu.cn
	Liang Yu, Ph.D Shanghai Jiao Tong University liang.yu@sjtu.edu.cn
	Meng Wang, Ph.D University of Notre Dame m.wang@nd.edu
	Xiaofan Li, Ph.D Illinois Institute of Technology lix@iit.edu
Response to Reviewers:	Please view attachment.

December 8, 2020

Editor of Process Safety and Environmental Protection

Dear Editor,

We have revised the manuscript entitled “Enhanced Spectrum Convolutional Neural Architecture: An Intelligent Leak Detection Method for Gas Pipeline”.

The response letter includes the point-to-point reply. Thank you very much in advance for considering this submission.

Looking forward to hearing from you in the future.

Name, address, telephone, and email address of the corresponding author:

Fangli Ning,

School of Mechanical Engineering, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, Shaanxi, China,

Tel: +86-29-8849-3927

ningfl@nwpu.edu.cn

Sincerely yours,

Fangli Ning

RеспONSES TO THE REVIEWERS

Rplies to the Reviewers

We appreciate the editor's and reviewers' careful review of our manuscript and detailed critiques, questions and comments. These two points in the report that have helped us improve our presentation of our results and clarify the results we have obtained. In the following, we have tried to address the points in the report in order. At the same time, we have revised the manuscript carefully and all modified text has been highlighted in the revised manuscript.

1. Reply to Reviewer

Main Points:

1. It is not enough to show labels 1, 2, 3, 4, 5, 6 - please tell the reader what real life situation they correspond to i.e. what type of leak or no leak, valve leak, hole leak?

Reply:

Thank you for your suggestions. In order to tell the reader what real life situation labels 1, 2, 3, 4, 5, 6 correspond to, Table 1 has been modified in the revised manuscript according to the intensity of background noise and the types of leaks. In Tab.1, we divided the background noise into

weak background noise and strong background noise, and divided types of leaks into no leak, hole leak and valve leak. Table 1 is shown as follow:

Table 1: Description of the six categories dataset.

Label	Background noise	Types of leaks	Training Set	Test Set
1	weak noise	no leak	2688	1152
2		hole leak		
3		valve leak		
4	strong noise	no leak		
5		hole leak		
6		valve leak		

2. The diagram in Fig.7 shows two types of leaks - can you detect only when these two are simultaneously present? What if only one leak is present? What is none are present? - Not clear.

Reply:

Sorry, the unclear description in Fig.7 caused a misunderstanding. In Fig.7, we want to show the valve and the flange which are the most likely leak positions, so Fig.7 has been modified in the revised manuscript. Figure.7 is shown at the end of *Reply*.

The proposed SE-CNN can detect the leak when only one leak is present. Two different types of leaks are rarely simultaneously present in the industrial environment, and are a challenge for the intelligent

leak detection method.

If none are present, the method could give no leak results.

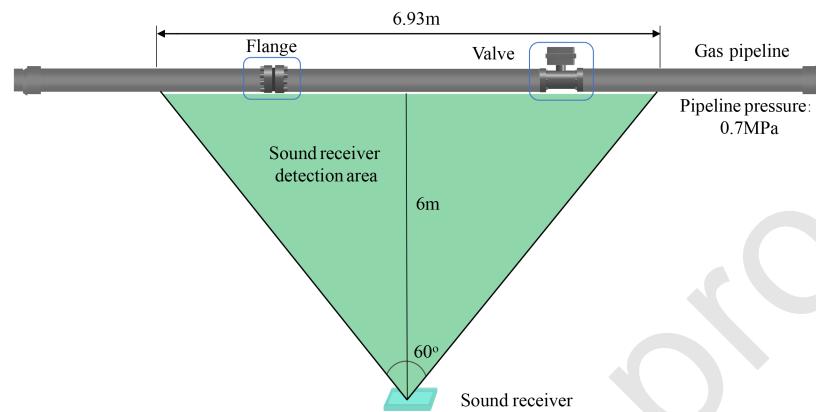


Figure 7: Schematic diagram of relationship between sampling distance and detectable pipeline length.

**Enhanced Spectrum Convolutional Neural Architecture:
An Intelligent Leak Detection Method for Gas Pipeline**

Authors: Fangli Ning^{a,*}; Zhanghong Cheng^a; Di Meng^a; Shuang Duan^a; Juan Wei^b

^aSchool of Mechanical Engineering, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, Shaanxi, China

^bSchool of Telecommunications Engineering, Xidian University, 2 South Taibai Road, Xi'an, Shaanxi, China

***Corresponding author:** Fangli Ning, School of Mechanical Engineering, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, Shaanxi, China, **Tel:** +86-29-8849-3927

, **E-mail:** ningfl@nwpu.edu.cn

Funding acknowledgement: This work was supported by National Natural Science Foundation of China (Grant No. 51675425, 52075441), Shaanxi Key Research Program Project (Grant No. 2020ZDLGY06-09), Dongguan Social Science and Technology Development(key) Project (Grant No. 20185071021600), Science and Technology on Micro-system Laboratory Foundation (Grant No. 6142804200405).

1
2
3
4
5
6
7
8
9 Enhanced Spectrum Convolutional Neural Architecture:
10 An Intelligent Leak Detection Method for Gas Pipeline
11
12

13 Fangli Ning^{a,*}, Zhanghong Cheng^a, Di Meng^a, Shuang Duan^a, Juan Wei^b
14
15

16 ^a*School of Mechanical Engineering, Northwestern Polytechnical University, 127 West
17 Youyi Road, Xi'an, Shaanxi, China*

18 ^b*School of Telecommunications Engineering, Xidian University, 2 South Taibai Road,
19 Xi'an, Shaanxi, China*

20
21
22
23
24 **Abstract**
25
26

27 In this work, a novel convolutional neural architecture (SE-CNN), which
28 combines Spectrum Enhancement (SE) and Convolutional Neural Network
29 (CNN), is proposed to detect the leak of gas pipeline. The SE has the effect
30 of enhancing the leak signals and reducing background noise. CNN can
31 automatically extract leak features and realize leak diagnosis. The
32 experimental results show that the SE-CNN can achieve an average
33 accuracy of 94.3% for 6 categories and only requires 1.04 seconds of
34 detection time. In this experiment, the diameters of the main pipeline and
35 the branch pipeline are 125 mm and 25 mm. Due to its excellent accuracy
36 and efficiency, the proposed enhanced spectrum convolutional neural
37 architecture paves the way for real-time leak detection in industrial
38 environments, which can ensure the process safety of gas pipeline
39 transportation. Under strong background noise, the average accuracy of the
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

55 *Corresponding author
56 Email address: ningf1@nwpu.edu.cn (Fangli Ning)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

SE-CNN can reach 94.3%, which is 33%, 3.7% higher than that of SVM and CNN. In particular, the SE can be regarded as a data compression method, which can significantly reduce the original data size. The training time of the SE-CNN is 539 seconds, reducing 90.6% compared with CNN.

Key words: Spectrum enhancement; Convolutional neural network; Leak detection

1. Introduction

The leak from the gas pipeline often occurs in the process of engineering practice, which not only endangers residents' safety and environment but also wasting energy affects a country's economy [1, 2]. Nowadays, gas pipeline leak detection has been becoming a hotspot in the oil and gas industry and a top priority to ensure the process safety of gas pipeline transportation. As shown in Fig.1, the current process of gas pipeline leak detection generally includes the four steps. Data collection is the first step in the process. The quantity and quality of the data determine the leak detection effect. At present, the raw data collected mainly include vibration signals [3], acoustic signals [4, 5] and images [6]. There are some unique advantages for acoustic signals, such as the long propagation distance, non-contact measurement and no blind area. At the same time, the leak acoustic signal has a frequency distribution range from infrasound to ultrasonic. Therefore, acoustic signals will be used as the raw data in this work. The second step, denoising [7], is to improve the quality of the data,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

such as the spectral subtraction [8], noise estimation [9], filtering algorithms [10, 11] and total variation (TV) method [12]. Feature extraction is the third step, which aims to extract key information and reduce the data dimension. Numerous feature extraction approaches have been developed, such as Mel-frequency cepstral coefficients (MFCC) [13], scale-invariant feature transform (SIFT) [14], histogram of oriented gradients (HOG) [15] which represent the raw acoustic signals and images. The last step is leak diagnosis to select an appropriate diagnosis algorithm as a classifier. There are also lots of machine learning algorithms, such as support vector machine (SVM) [16], hidden Markov model (HMM) [17], Bayesian network [18], deep learning (DL) [19], which have been used for leak diagnosis.

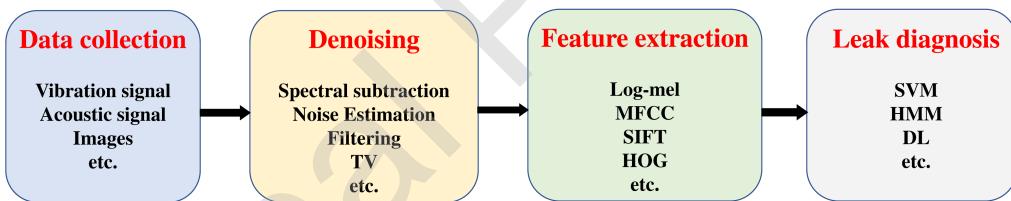


Fig. 1. The process of gas pipeline leak detection.

Based on acoustic signals, many machine learning algorithms have been proposed for gas pipeline leak detection. Li *et al.* [20] proposed a leak detection scheme based on kernel principal component analysis (kernel PCA) and SVM classifier for recognizing the leak. Cruz *et al.* [21] proposed a technique that combined acoustic sensors and machine learning algorithms to identify and locate leaks in low-pressure gas pipelines. Sun *et*

al. [22] proposed a small leak feature extraction and recognition method based on SVM. With the rapid development of artificial intelligence in recent years, the accuracy of gas pipeline leak detection has been improved greatly by deep learning technology. Zadkarami *et al.* [23] extracted wavelet features and used them as the inputs of a multi-layer perception neural network (MLPNN) classifier to determine the leak state. Wang *et al.* [24] developed a computer vision approach based on optical gas imaging (OGI) to detect the leak. Bae *et al.* [25] proposed a deep-learning leak detection technique that used trajectory-based image features extracted from time-series acoustic data. They developed root-mean-square-pattern and frequency-pattern images to reflect the leak signal characteristics and used them as the inputs of residual networks. Sun *et al.* [26] proposed an intelligent pipeline leak detection method combining compressed sensing and deep learning theory, which can achieve compressed sampling, adaptive feature extraction and recognition.

Although the above algorithms have greatly promoted the progress of gas leak detection, there is still a long way to the application in the industrial environment. The performance of the leak detection algorithm is mainly limited to two aspects. Firstly, the effect of the leak diagnosis system depends on the denoising algorithm. Many denoising methods have a significant effect on the suppression of stationary noise. Due to the continuous and stable state of the leak, the leak signal is stationary, but the background noise signal in application scenes is non-stationary, which

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

makes it difficult to apply these denoising methods to the industrial environment. What's more, these denoising methods suffer from a weakness to achieve a trade-off between noise reduction and the undistorted target signal [27]. Secondly, in the complexity of the industrial environment, the raw audio contains a lot of redundant information, which makes feature extraction difficult and challenging. Moreover, feature extraction depends on prior knowledge and is specific to the task, so it must be redesigned for every new industrial environment.

In order to overcome the above difficulties and improve the recognition effect of the leak detection system, an architecture combining spectrum enhancement (SE) and CNN is proposed in this work. Convolution operations are used in the SE method to enhance the stationary leak signals while eliminating the non-stationary noise signals, thereby achieving the goal of enhancing the leak signal characteristics while reducing noise. With the rapid development of deep learning algorithms, convolutional neural networks have realized great achievements in many fields, such as Image Classification [28], Speech Recognition [29], and Image Segmentation [30]. There are many representative CNN models, LeNet-5 [31] is originally used as handwritten Arabic digit recognition in the bank, AlexNet [32], VGGNet [33], GoogleNet [34], ResNet [35] are the best models that stand out from ImageNet competition. CNN can achieve end-to-end leak detection, which can directly obtain the detection results from the raw audio signal. The last three steps in the process (Fig.1) are

merged into one step by the proposed algorithm, so the process of gas pipeline leak detection will become more efficient.

The rest of this paper is organized as follows. In Sec.2, the basis of CNN architecture is described in detail. In Sec.3, the method of spectrum enhancement and the SE-CNN are illustrated. The experimental results and analysis are presented in Sec.4. Conclusions and future work are summarized in Sec.5.

2. Convolutional neural network

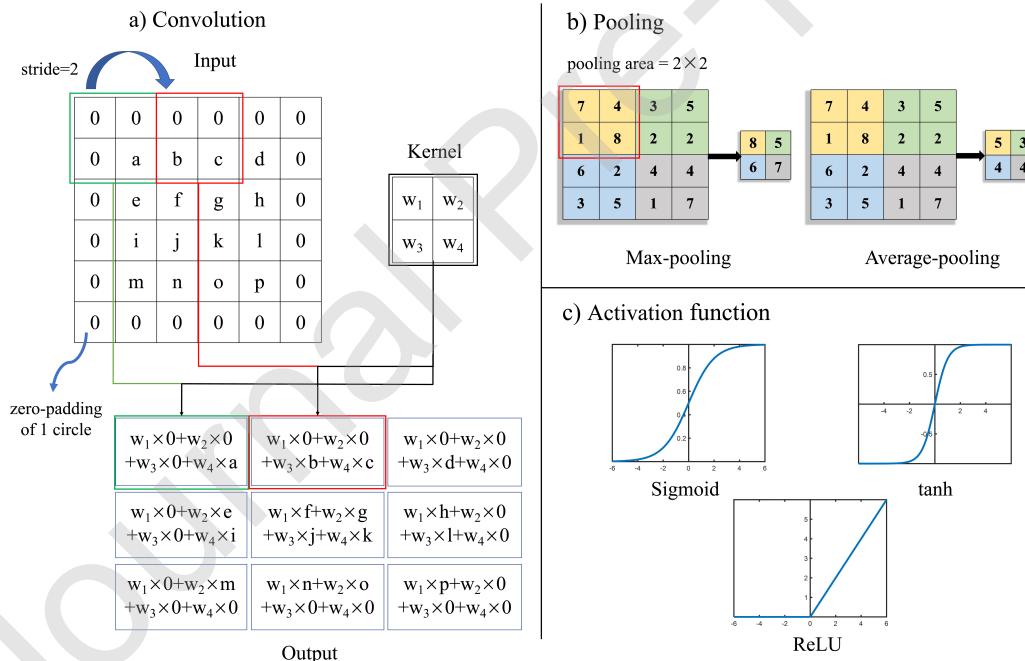


Fig. 2. Three components of convolutional neural networks. (a) Feature extraction of input by convolution operation. (b) Two main methods of pooling. (c) Three main activation functions for neural networks.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

As shown in Fig.2, convolutional neural networks have three essential components: convolution, pooling, and activation function:

a) Convolution is the most important component making neural networks gain the ability to extract the buried features. It improves the neural network's performance through three ideas: sparse connectivity, parameter sharing, and equivariant representations [36]. The convolution kernel size is much smaller than the input to achieve sparse connections, which can help reduce the number of parameters and storage requirements. Parameter sharing is realized by repeatedly using identical weight parameters during the convolution operation of the same layer, where weights are the parameters w_1, w_2, w_3, w_4 of kernel in Fig.2(a). The convolution O of A and B can be calculated as

$$O(s, t) = (A * B)(s, t) = \sum_m \sum_n A(s + m, t + n)B(m, n), \quad (1)$$

where O is the output, A is the input, B is the convolution kernel, $*$ denotes convolution operation.

Due to the parameter sharing, the convolution layer has some translation invariance, which means that no matter where the target is in the input, the same output will be obtained at the corresponding area.

b) Pooling can reduce each layer's output size, and improve the computational efficiency and robustness of the extracted features. The max-pooling $\alpha = \max\{\alpha_i\}, \alpha_i \in \mathcal{N}$ and the average-pooling

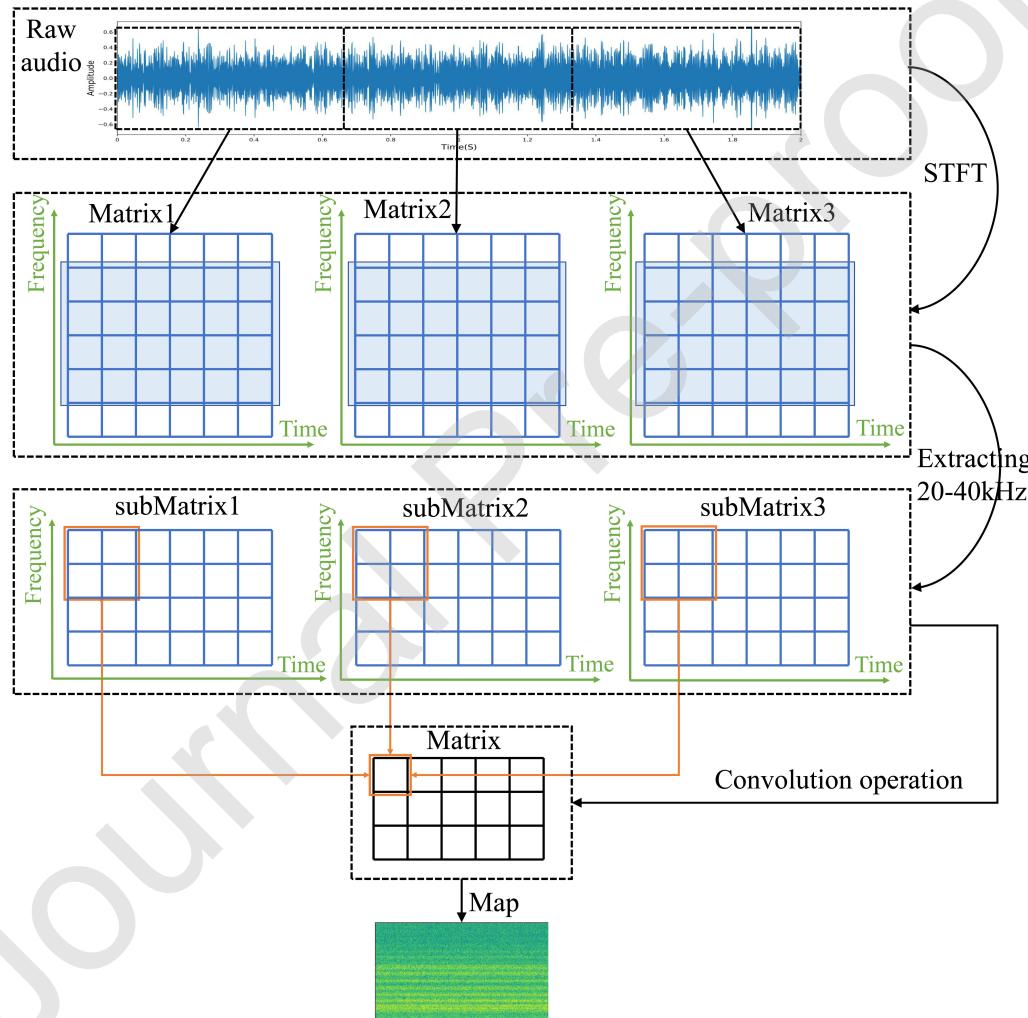
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

$\alpha = \frac{1}{k} \sum_{i=1}^k \alpha_i, \alpha_i \in \mathcal{N}$ are two of the most widely used pooling methods, where α is the output, \mathcal{N} is the area of pooling operation, k is the number of elements in the pooling area, α_i is the value in the pooling area. The max-pooling is achieved by extracting the maximum value to represent the pooling area, which has been proved to be one of the most effective pooling methods. The average-pooling is achieved using the mean in the pooling kernel representing the area, which is usually used in deep neural networks and the last layer of fully convolutional networks.

c) Activation functions give the neural networks a strong fitting ability. There are several typical activation functions for different application requirements. The sigmoid [37] is one of the most basic activation function, but it is rarely used now because of the gradient vanishing. The tanh [38] is essentially a variant of sigmoid, which has a better effect, but it still cannot solve the gradient vanishing problem. The rectified linear units (ReLU) [39] has been the most popular activation function, which can well solve the problems of gradient vanishing and gradient exploding. Simultaneously, due to its simple function and derivative, forward and backward propagation is significantly accelerated, so the overall training speed of the neural networks is greatly improved.

1
2
3
4
5
6
7
8
9 3. The new framework combining spectrum enhancement and
10
11 CNN

12
13
14 3.1. Spectrum enhancement
15
16
17



54 Fig. 3. The process of spectrum enhancement.
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

A spectrum enhancement method imposed on the matrix of the short-time Fourier Transform (STFT) is proposed in this study, which can enhance the stationary leak signals while eliminating the non-stationary noise signals. As a time-frequency analysis method, STFT is widely used in signal processing. STFT can clearly reflect the features of the raw audio than one-dimensional time series [40], it needs less processing and can save more original information than MFCC [41]. The ultrasonic domain's time-frequency spectrum (20023 Hz-40047 Hz) is extracted, which aims to reduce the impact of the noise. As shown in Fig.3, the novel method process mainly follows the four steps:

- (1) Divide each sample (audio) into B blocks of equal length.
- (2) Perform STFT on each block separately to obtain the matrix.

$$\text{STFT}\{x[n]\}(m, \omega) \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}, \quad (2)$$

where $X(m, \omega)$ is the matrix of STFT, $x[n]$ represents the original signal, $w[n-m]$ is a window function centred at the time m .

In $X(m, \omega)$, the horizontal axis represents the time, the vertical axis represents the frequency, the element of the matrix represents the power spectrum described by the decibel scale.

- (3) Extract the submatrix which represents the ultrasonic domain in each matrix.

- (4) Perform convolution operation on the elements in the kernels of size $K \times K$ to slide on each submatrix with the stride of S , the enhanced rule is calculated by,

$$M = M^{(1)} \odot \dots \odot M^{(b)} \odot \dots \odot M^{(B)}, \quad M^{(i)} \subset X^{(i)}$$

$$y = \sum_{p=1}^K \sum_{q=1}^K M(p, q), \quad (3)$$

where $X^{(i)}$ is the i -th submatrix, $M^{(i)}$ is the kernel in the i -th submatrix. \odot denotes Hardmdard product (component-wise multiplication), $M(p, q)$ is the (p, q) -th element of the matrix M , y is the output.

3.2. The architecture of SE-CNN

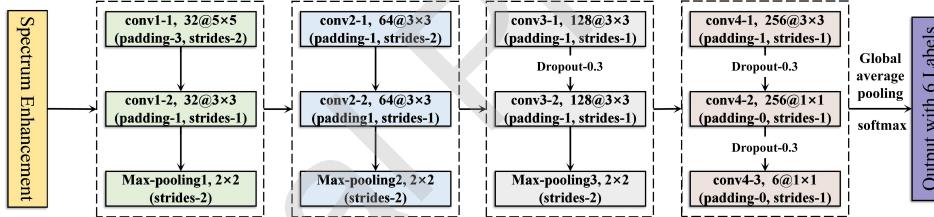


Fig. 4. The architecture of SE-CNN. Conv1-1 is the name of the convolutional layer, $32@5 \times 5$ represents that there are 32 kernels, and the size of kernels is 5×5 , padding-3 means performing zero-padding operation of 3 circles on the edge of the Input, strides-2 indicates that the stride of the convolution kernel sliding on the input is 2 pixels.

The specific architecture of SE-CNN applied in this work is illustrated in Fig.4. The architecture design is inspired by VGGNet, which has shown good mobility and scalability in many deep learning tasks. There is a consensus that the more layers a neural network, the better its performance. However, with the layers increasing, the number of

parameters increases exponentially. Overfitting, which greatly influences the algorithm generalization, usually occurs in a neural network with lots of parameters. Therefore, to reduce the generalization error of the algorithm and make it still maintain a good recognition effect for new samples (without training), the global average pooling is chosen instead of fully connected layers at the end of the network, which dramatically reduces the number of parameters. And the dropout[42] method is adopted in the SE-CNN architecture, which can randomly deactivate the connected units so that the network does not rely on some specific units, and its architecture will be simplified. The whole architecture includes spectrum enhancement, nine convolution layers, and four subsampling layers. The spectrum enhancement is added first. The max-pooling is selected in the first three subsampling layers, and the last layer is the average-pooling. ReLU is used as the activation function in all layers except the last layer, the last layer's activation function is softmax.

After the SE-CNN architecture is determined, the cost function is defined as

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k y_j^{(i)} \log(\hat{y}_j^{(i)}), \quad (4)$$

where $y_j^{(i)}$ represents the true output of the i -th sample on the j -th unit, $\hat{y}_j^{(i)}$ represents the predicted output of the i -th sample on the j -th unit, m represents the number of samples on one iteration and is set to be 32 in the experiment, k represents the number of units in the output layer.

The forward propagation of ℓ -th layer is given as

$$\mathbf{x}_j^\ell = f \left(\sum_{i \in M_j} \mathbf{x}_i^{\ell-1} * \mathbf{k}_{ij}^\ell + b_j^\ell \right), \quad (5)$$

where M_j represents a selection of input maps, i is the index of input map, j is the output maps, k is the convolutional kernel. Each output map is given an additive bias b , for a particular output map, the input maps will be convolved with distinct kernels. Based on Eq.(4), the gradient term of output layer (L -th layer) can be calculated as

$$\delta^{(L)} = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k \frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \odot f'(z^{(L)}), \quad (6)$$

where $\delta^{(L)}$ is the “errors” of L -th layer, which can be thought of as “sensitivities” of each unit with respect to perturbations of the parameters.

Error back-propagation follows the chain derivation rule.

In order to speed up the convergence and improve the generalization ability of SE-CNN, Batch Normalization (BN) is applied in our proposed model. The calculation process of BN is as follows

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (7)$$

$$y_i = \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad (8)$$

Eq.(7) is the normalization process where x_i is the input, μ_B is the mean of x_i , σ_B^2 is the variance of x_i , ϵ is a constant added to variance for numerical stability. Eq.(8) is the process of scale and shift where γ and β are introduced to prevent damage to the feature distribution of the previous layer data.

The back propagation of ℓ -th layer is given as

$$\delta^\ell = W^{\ell+1} (f' (z^\ell) \odot up (\delta^{\ell+1})) , \quad (9)$$

where $\ell = 1, 2, \dots L - 1$, W is the weight matrix of ℓ layer, up denotes upsampling. Update the weights and bias with

$$\begin{aligned} W_{(i+1)}^\ell &= W_{(i)}^\ell - \eta x^{\ell-1} (\delta^\ell)^T, \\ b_{(i+1)}^\ell &= b_{(i)}^\ell - \eta (\delta^\ell), \end{aligned} \quad (10)$$

where $W_{(i+1)}^\ell$ and $b_{(i+1)}^\ell$ are the value of W^ℓ and b^ℓ at the i -th iteration, η is the learning rate.

After getting the output by global average pooling, the softmax function is used for classification.

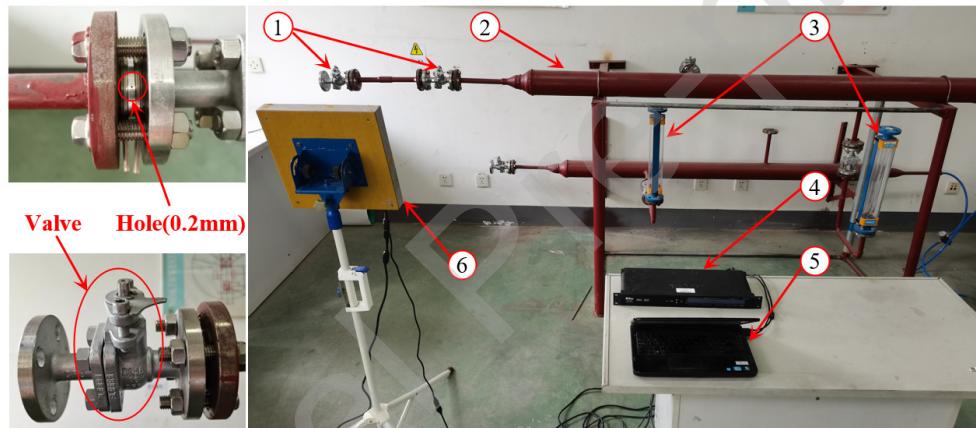
$$S_i = \frac{e^i}{\sum_{j=1}^J e^j} \quad (11)$$

where j represents the number of units in layer, i represents i -th unit of j units, S_i is the predicted probability of label i . Then, choose the largest predicted probability as the predicted result.

1
2
3
4
5
6
7
8
9 4. Experiments
10
11

12 To verify the effectiveness of the proposed method, the experimental
13 results of the SE-CNN and CNN approaches are compared in this section.
14 At the same time, in order to achieve the best performance of leak
15 detection, the mesh search method is used to search the optimal parameters
16 (B – block, S – stride, K – kernel) for spectrum enhancement.
17
18
19
20
21
22
23

24 4.1. Experiment instruments and dataset
25
26



41 Fig. 5. Experimental system: (1) valve; (2) gas pipeline; (3) flowmeter;
42 (4) signal
43 collector; (5) laptop; (6) sound receiver.
44
45

46 As shown in Fig.5, the entire experimental system consists of gas
47 pipelines, a sound receiver, a signal collector that is used to convert analog
48 signals into digital signals and the sampling rate is set to be 96 kHz, a
49 laptop which is used to analyze and process the raw data. In the process of
50 data collection, the experimental conditions, such as the valve opening, leak
51 hole size, and sampling distance (0.8m-1.2m), are changed continuously to
52
53
54
55
56
57
58

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

simulate various real leak conditions. This helps to increase the diversity of the samples. In this experiment, the diameters of the main pipeline and the branch pipeline are 125 mm and 25 mm. Collecting leak acoustic signals in the laboratory and background noise (fan noise signals) in the pipe gallery, a total of 8.5 hours real audio is obtained. Synthesizing the laboratory signals with the background noise to simulate gas pipeline leak under the industrial environment. Ultimately, a total of 13 hours audio is obtained. The audio signals are cut into short clips of 2 seconds, each of them is regarded as a sample. There are 23040 samples in total, which are randomly divided into a training set (70%) and a test set (30%). The descriptions of different classes are listed in Tab.1. The signal-to-noise ratio (SNR) of label 2, 3 is 5 dB, which can be regarded as leaks under weak noise. The SNR of synthesizing signal label 5, 6 is -25 dB, which can be regarded as leaks under strong noise. The training and predict processes of the SE-CNN and CNN are done on the Intel(R) Core(TM) i5-8300H @ 2.30 GHz system with 8GB-RAM and 4GB-GPU RAM.

In order to verify the applicability of the proposed SE-CNN in the industrial environment where the leak location is unknown, audio data were collected at 6m away from the gas pipeline. Figure 6 shows the experimental scene with a sampling distance of 6m. Under different sampling distances, 3840 hole and valve leak audio were collected, respectively, each of which is 2 seconds.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1
Description of the six categories dataset.

Label	Background noise	Types of leaks	Training Set	Test Set
1	weak noise	no leak	2688	1152
2		hole leak		
3		valve leak		
4	strong noise	no leak		
5		hole leak		
6		valve leak		

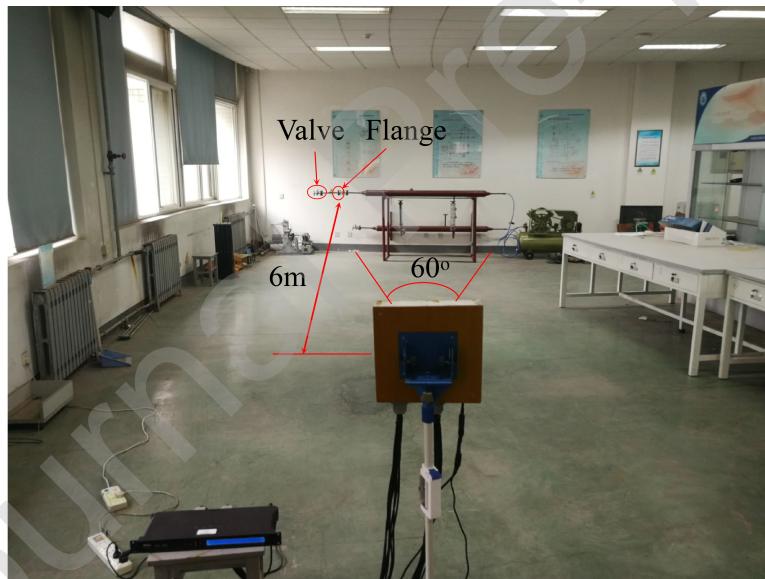


Fig. 6. The experimental system at 6m sampling distance.

Figure 7 shows the relationship between sampling distance and detectable pipeline length. The directivity of the sound receiver is 60 degrees. Therefore,

when the sampling distance is 6 m, the sound receiver can detect the pipeline up to 6.93m.

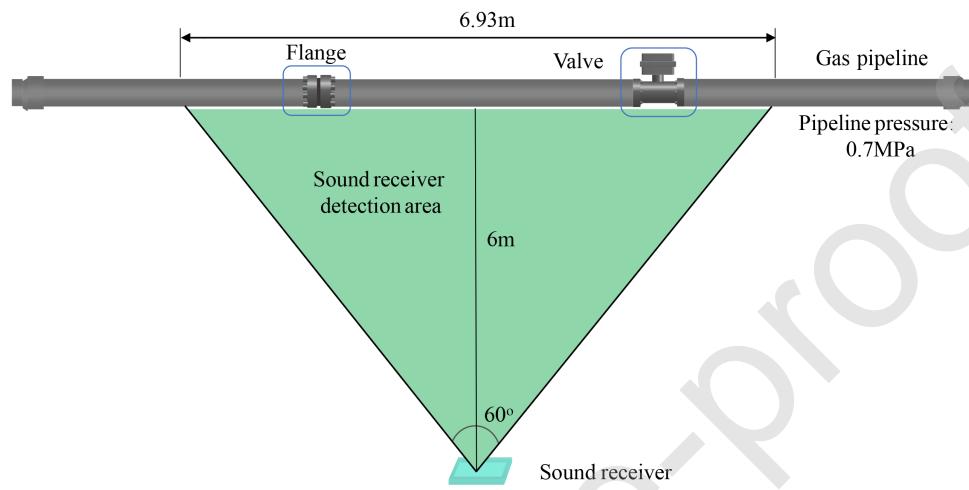


Fig. 7. Schematic diagram of relationship between sampling distance and detectable pipeline length.

The recall ratio r , precision ratio p , and f_1 -score are used to evaluate the performance of the algorithm and choose the most optimal parameters of the spectrum enhancement. The r and p are respectively defined as:

$$r = \frac{TP}{TP + FN}, \quad (12)$$

$$p = \frac{TP}{TP + FP}. \quad (13)$$

If the sample is positive and it is classified as positive, it is counted as true positive (TP); if it is classified as negative, it is considered as false negative (FN). If the sample is negative and it is classified as negative it is considered

as true negative (TN); if it is classified as positive, it is counted as false positive (FP).

The f_1 -score is a more comprehensive criterion to evaluate the performance. It is defined as

$$f_1 = \frac{2pr}{p+r} \quad (14)$$

The accuracy is the most popular to evaluate the performance of the algorithm, which is defined as

$$\text{accuracy} = \frac{N_{\text{correct}}}{N_{\text{all}}}, \quad (15)$$

where N_{correct} represents the number of samples that are predicted correctly, N_{all} is the number of all samples.

4.2. Results and analysis

The enhanced matrices are visualized to show the effect of the spectrum enhancement in Fig.8. The first column is the short-time Fourier Transform spectrums of the 6 categories data. The second column is the spectrums of corresponding data, which are obtained by the spectrum enhancement. The spectrums labeled 1, 2, 3 from weak noise environment have high similarity. The spectrums labeled 4, 5, 6 from strong noise environment also have high similarity. The spectral enhancement highlights the difference in time-frequency diagrams of different categories shown on Fig.8, which can

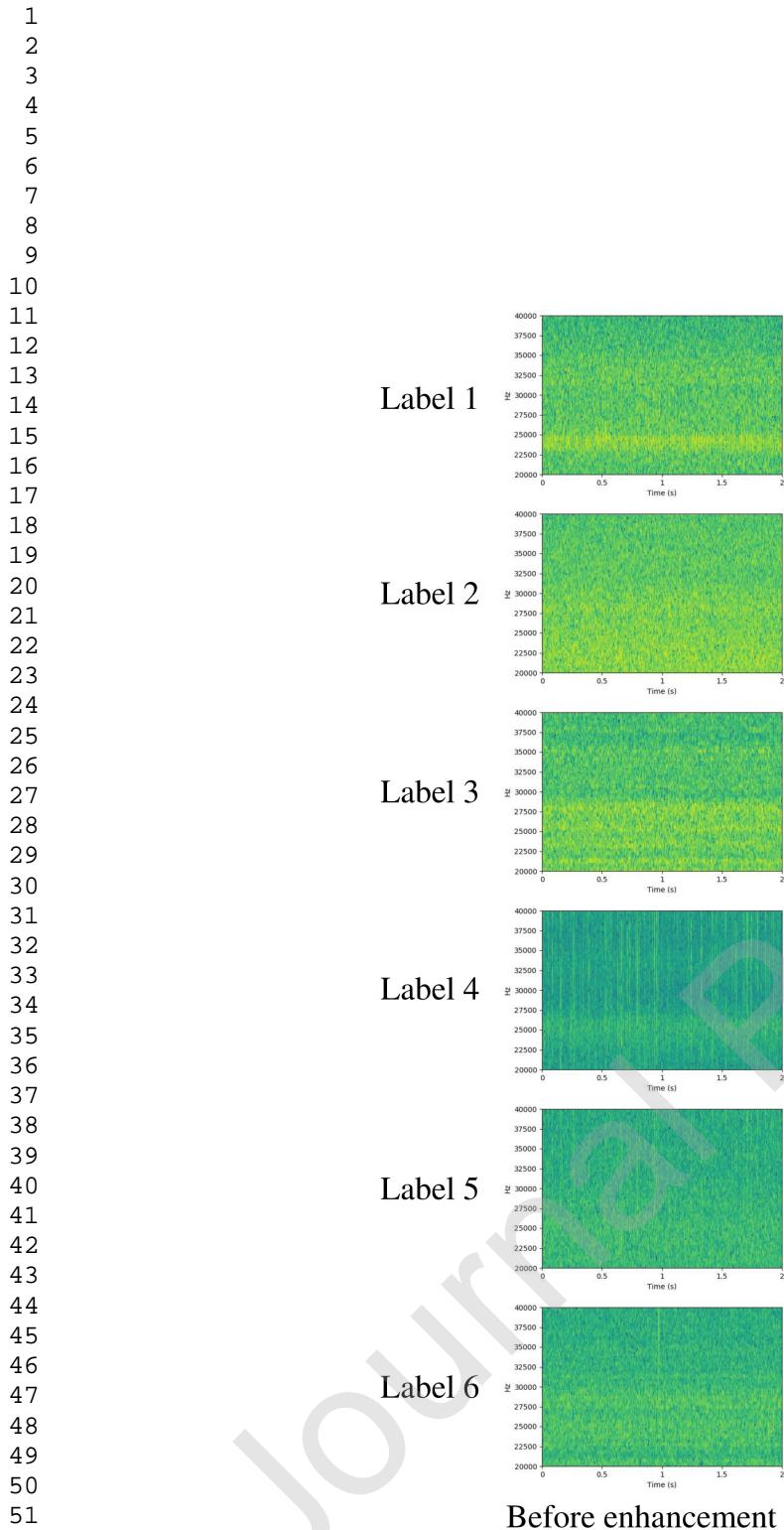


Fig. 8. Comparison of before and after spectrum enhancement for 6 categories.

1
2
3
4
5
6
7
8
9
10
11
12
13
14

only give a qualitative, but not quantitative, spectral enhancement effectiveness. Next, the quantitative verification will be carried out by the comparison of experimental results.

Figure 9 shows that the results of the architectures when B (the number of blocks) are different values. $B = 1$ means there is no spectrum enhancement, which represents the results of the CNN approach. As shown in Fig.9(a), the framework of SE-CNN has better performance than CNN in every class. Especially under the strong background noise environment, the improvement is reaching nearly 5%, the most notable is that the improvement of label 5 increases more than 8%. The performance of SE-CNN under the strong noise environment is excellent, which confirms that the spectrum enhancement can indeed play a specific role in boosting the performance. Simultaneously, the performances of SE-CNNs with different B are close, and the difference in overall performance is within 0.4%. Increasing the value of B can make the spectrum more effectively enhanced, but the size of the spectrum will be reduced correspondingly, which will inevitably lose some vital information. So the algorithm performance cannot be improved only by increasing B . It can be seen from Fig.9(b), the r of CNN is less than the SE-CNN in each class except label 4, which is more than 1.6% compared with that of the SE-CNN. However, the r of label 5 is 68.1%, which is lower than 15.1% compared to the SE-CNN. On the contrary, as shown in Fig.9(c), the p of label 4 is 73.3% and significantly lower than 80.6% of the SE-CNN, but the p of label 5 is 99.8% which is slightly higher than 97.1% of the SE-CNN. The reason for

the low precision ratio of label 4 is that the features of strong background noise are not obvious in the ultrasonic frequency band. It can be seen from the Fig.9(d), the improvement in accuracy of the label 1, 2, 3 is between 1.0% ~ 1.8%. The improvement in accuracy of the label 4, 5, 6 is between 3.5% ~ 4.3%. It confirms that the SE-CNN is more suitable for industrial environments than CNN. As far as the experimental results are concerned, The number of blocks is chosen as 4 to achieve a good performance in this work.

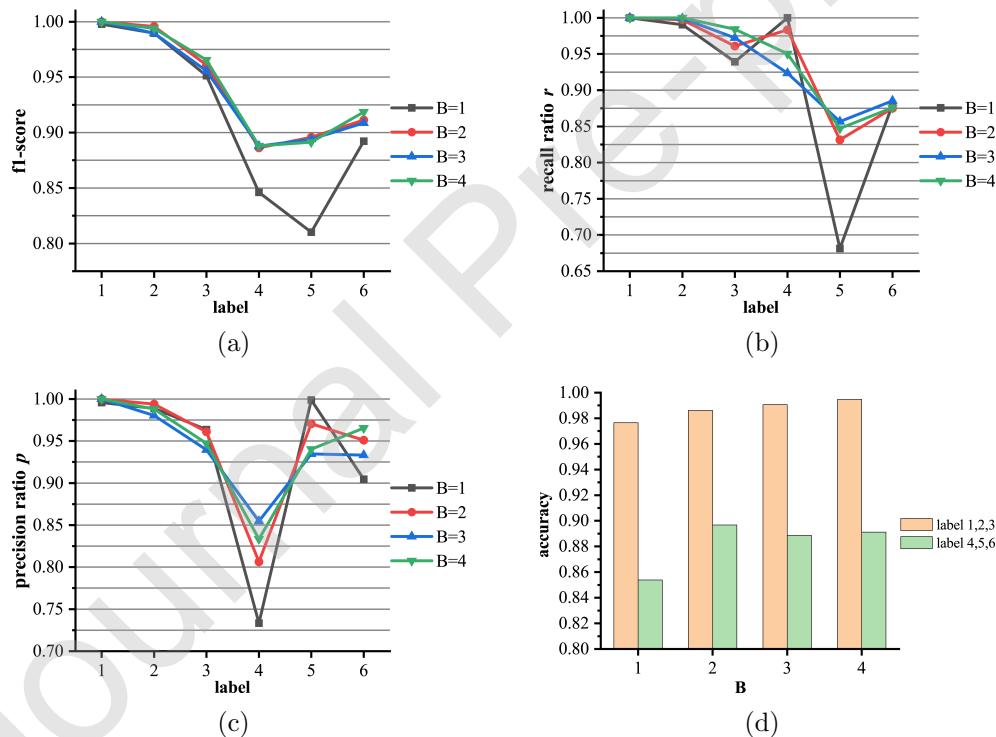


Fig. 9. (a) f_1 -score (b) Recall ratio r (c) Precision ratio p (d) Accuracy with different B (the number of blocks).

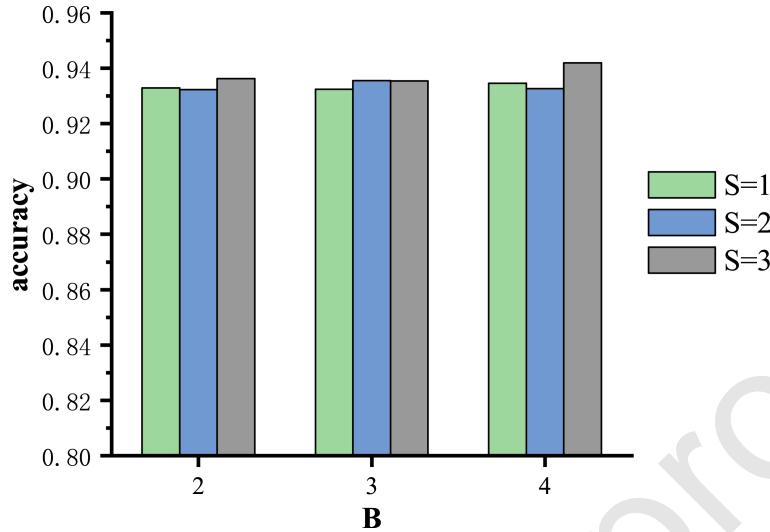


Fig. 10. Accuracy with different S (the number of strides).

Table 2
Comparison of computing cost with different S (stride).

Parameters	Training time(s) 16128 samples	Test time(s) 6912 samples	Predict time(s) 1 sample
$S = 1 (B = 2, K = 3)$	6140.26	74.32	1.23
$S = 2 (B = 2, K = 3)$	1587.78	58.93	1.07
$S = 3 (B = 2, K = 3)$	811.24	38.54	1.05

As shown in Fig.10, the overall accuracy of SE-CNNs with different S (the number of strides) ranges from 93.2% to 94.1%, which indicates that S does not have much impact on the performance of SE-CNNs. Noticeably, the larger S can greatly reduce the training time of the neural network, which is shown in Tab.2. In fact, S changes the size of the neural network's inputs, the size of the sample is 188×214 when $S = 1$, the size of the sample is 94×107

when $S = 2$, and then the size of the sample is 63×72 when $S = 3$. Changing the size of the sample by changing the S can greatly reduce the computing cost and speed up the training process. Nevertheless, if S is too large, the sample will lose lots of key information, which is definitely detrimental to the performance of SE-CNNs. In order to minimize the consumption of computing power, S is finally set to be 3 in this work.

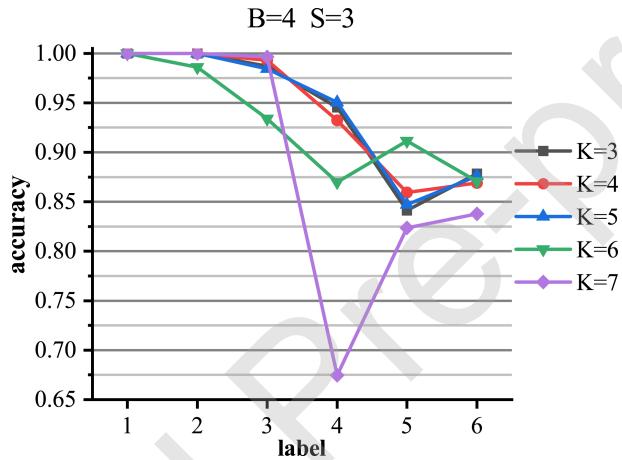


Fig. 11. The accuracy of SE-CNNs with different K (the number of kernels).

The value of K (the number of kernels) mainly depends on the value of S . When $K < S$, some elements in the matrix are not involved in the convolution operation when spectrum enhancement is performed, which will inevitably waste some information and undermine the performance of the SE-CNN, when $K \geq 2S$, the difference between adjacent elements of the matrix after spectrum enhancement will be reduced, so the feature of the spectrum be weakened. Therefore, it is also harmful to the performance of the SE-CNN. As shown in Fig.11, there is a dramatic decrease in the accuracy of

SE-CNNs when $K \geq 2S$. The experiment result shows that a SE-CNN can obtain an excellent performance when $K = S + 1$ or $K = S + 2$. This is due to the fact that K is slightly larger than S , which can maximize the effect of spectrum enhancement. K is finally set to be 5 in this work.

Table 3
Comparison results of SVM, CNN and SE-CNN.

Label	1	2	3	4	5	6	f_1 -score	Acc	AvgAcc
SVM (The accuracy of 6 categories = 76.1%)									
1	1117	25	10	0	0	0	0.979	96.9%	
2	2	1078	72	0	0	0	0.948	93.6%	96.1%
3	0	20	1126	0	0	0	0.947	97.7%	
4	1	0	0	867	170	114	0.611	75.3%	
5	4	0	0	605	393	150	0.402	34.1%	56.1%
6	0	0	19	215	243	677	0.647	58.8%	
CNN (The accuracy of 6 categories = 91.5%)									
1	1152	0	0	0	0	0	0.998	100%	
2	5	1141	0	6	0	0	0.990	99%	97.7%
3	0	1	1082	0	0	69	0.951	93.9%	
4	0	0	0	1152	0	0	0.846	100%	
5	0	11	1	317	785	38	0.810	68.1%	85.4%
6	0	1	40	96	1	1014	0.892	88%	
SE-CNN (B = 4, S = 3, K = 5; The accuracy of 6 categories = 94.3%)									
1	1152	0	0	0	0	0	1.000	100%	
2	0	1152	0	0	0	0	1.000	100%	99.5%
3	0	1	1134	0	0	17	0.966	98.4%	
4	0	1	0	1095	47	10	0.888	95.1%	
5	0	13	0	154	976	9	0.891	84.7%	89.1%
6	0	0	63	65	15	1009	0.919	87.3%	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Some scholars have used a support vector machine (SVM) for pipeline leakage detection [43, 44]. The classification results of SVM, CNN and SE-CNN with the optimal parameters are shown in Tab.3. The results in Tab.3 are obtained on a dataset with sampling distance of 0.8m-1.2m. Compared SVM, CNN with SE-CNN, the accuracy of leak detection under the weak background noise increases from 96.1%, 97.7% to 99.5%, which achieves a 3.4%, 1.8% improvement, respectively. In the industrial environment, strong background noise makes it difficult to detect the leak in the gas pipeline, but the improvement of leak detection under strong background noise is significant. The accuracy of label 4, 5, 6 increases from 56.1%, 85.4% to 89.1%, which achieves a 33%, 3.7% improvement.

Table 4 illustrates the computing cost of CNN and the SE-CNN. Compared with CNN, the training time, testing time and predict time of the SE-CNN are reduced by 90.6%, 62.5% and 15.4%, respectively. While using two-dimensional discrete convolution operation to reduce noise and enhance the striations of spectrums, increasing the stride (S) to speed up the convolution operation process, so that the accuracy and efficiency of the SE-CNN are significantly improved.

According to the parameters determined above ($B = 4$, $S = 3$, $K = 5$), these audio data collected at 6m away from the gas pipeline were input into SE-CNN as a test set to confirm its practicality in the industrial environment. Table 5 lists the results of these data on CNN and SE-CNN.

9
10 **Table 4**
11 Comparison of computing cost between SE-CNN and CNN.
12

Algorithm	Training time(s) 16128 samples	Test time(s) 6912 samples	Predict time(s) 1 sample
CNN	5746.12	70.75	1.23
SE-CNN (B = 4, S = 3, K = 5)	539.08	26.56	1.04

24 **Table 5**
25 Comparison results of CNN and SE-CNN at sampling distance of 6m.
26

Model	Metrics	Label					
		1	2	3	4	5	6
CNN	f_1 -score	0.997	0.615	0.764	0.639	0.277	0.658
	Accuracy	99.5%	47.1%	100%	95.5%	16.9%	50%
	Avg Acc			68.2%			
SE-CNN	f_1 -score	1	0.813	0.900	0.907	0.734	0.858
	Accuracy	100%	68.5%	83.1%	98.9%	82.3%	87.8%
	Avg Acc			86.8%			

44 As the sampling distance increases, the average accuracy of CNN and
45 SE-CNN for 6 categories both decreases. However, the average accuracy
46 of SE-CNN can reach 86.8%, which is 18.6% higher than CNN when the
47 sampling distance is 6m. The experimental results indicate that SE-CNN
48 can still effectively detect leaks even at long sampling distances.
49
50
51
52
53

5. Conclusions

The intelligent architecture (SE-CNN) for leak detection of the gas pipeline has been presented. Experiments were conducted to train the SE-CNN and test the performance of SE-CNN. In this experiment, the diameters of the main pipeline and the branch pipeline are 125 mm and 25 mm. The critical features of the study can be summarized as follow

- The spectrum enhancement method can effectively enhance the stationary leak signals and reduce noise in the aliasing signals. Therefore, the SE-CNN can achieve an average accuracy of 94.3%, which is 18.2%, 2.8% higher than SVM, CNN. Especially, the accuracy is 33%, 3.7% higher than SVM, CNN under strong background noise (SNR = -25 dB).
- The spectrum enhancement can also be regarded as a data compression method, it can significantly reduce the size of the original signal, which is good for speeding up the training process of the neural network and reducing the computing cost. The training time of SE-CNN is 539 seconds, and the predict time is 1.04 seconds, reducing 90.6% and 15.4% computing cost, respectively, compared with CNN.
- The SE-CNN keeps all the advantages of CNN. One important characteristic of CNN is to avoid the challenging feature engineering problem and achieve end-to-end processing of the raw data. On the

other hand, The SE-CNN uses the average-pooling instead of fully connected layers in the end, which dramatically reduces the risk of overfitting.

The proposed SE-CNN can satisfy two major requirements of gas pipeline leak detection, which are accuracy and efficiency. The good performance of SE-CNN brings prospects to industrial applications of gas pipeline leak detection. However, due to the leak in the industrial environments are diverse, the six categories samples in this study cannot fully cover all leak types. Therefore, more experiments need to be conducted in the future to obtain more types of leak signals such as crack leak, weld defect leak and so on. The SE-CNN can perform the real-time detection of gas pipelines to ensure the process safety of transportation. Besides, timely response to pipeline leaks is also important to risk engineering for the gas pipeline. Therefore, SE-CNN provides potential applications in industrial environments for both process safety and risk engineering.

Conflict of interest statement

The authors declared that there is no conflict of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 51675425, 52075441), Shaanxi Key Research Program

1
2
3
4
5
6
7
8
9 Project (Grant No. 2020ZDLGY06-09), Dongguan Social Science and
10 Technology Development(key) Project (Grant No. 20185071021600),
11 Science and Technology on Micro-system Laboratory Foundation (Grant
12 No. 6142804200405).
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

- [1] S. Datta, S. Sarkar, A review on different pipeline fault detection methods, *Journal of Loss Prevention in the Process Industries* 41 (2016) 97–106.
- [2] S. Bonvicini, G. Antonioni, P. Morra, V. Cozzani, Quantitative assessment of environmental risk due to accidental spills from onshore pipelines, *Process Safety and Environmental Protection* 93 (2015) 31 – 49.
- [3] Q. Wang, L. Han, X. Fan, J. Zhu, Distributed fiber optic vibration sensor based on polarization fading model for gas pipeline leakage testing experiment, *Journal of Low Frequency Noise Vibration and Active Control* 37 (3) (2018) 468–476.
- [4] Q. Xu, L. Zhang, W. Liang, Acoustic detection technology for gas pipeline leakage, *Process Safety and Environmental Protection* 91 (4) (2013) 253 – 261.
- [5] Y. An, X. Wang, B. Yue, S. Jin, L. Wu, Z. Qu, A novel method for natural gas pipeline safety online monitoring based on acoustic pulse

1
2
3
4
5
6
7
8
9 compression, Process Safety and Environmental Protection 130 (2019)
10
11 174 – 181.
12
13

- 14 [6] J. Shi, Y. Chang, C. Xu, F. Khan, G. Chen, C. Li, Real-time
15 leak detection using an infrared camera and faster r-cnn technique,
16 Computers & Chemical Engineering 135 (2020) 106780.
17
18 [7] W. Lu, W. Liang, L. Zhang, W. Liu, A novel noise reduction method
19 applied in negative pressure wave for pipeline leakage localization,
20 Process Safety and Environmental Protection 104 (2016) 142 – 149.
21
22 [8] S. Boll, Suppression of acoustic noise in speech using spectral
23 subtraction, IEEE Transactions on acoustics, speech, and signal
24 processing 27 (2) (1979) 113–120.
25
26 [9] I. Cohen, Noise spectrum estimation in adverse environments: improved
27 minima controlled recursive averaging, IEEE Transactions on Speech
28 and Audio Processing 11 (5) (2003) 466–475.
29
30 [10] J. Chen, J. Benesty, Y. Huang, S. Doclo, New insights into the
31 noise reduction wiener filter, IEEE Transactions on audio, speech, and
32 language processing 14 (4) (2006) 1218–1234.
33
34 [11] P. Yu, J. Cao, V. Jegatheesan, L. Shu, Activated sludge process faults
35 diagnosis based on an improved particle filter algorithm, Process Safety
36 and Environmental Protection 127 (2019) 66 – 72.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- [12] Q. Yuan, L. Zhang, H. Shen, Hyperspectral image denoising employing a spectral-spatial adaptive total variation model, *IEEE Transactions on Geoscience and Remote Sensing* 50 (10) (2012) 3660–3677.
- [13] T. Xu, S. Chen, S. Guo, X. Huang, J. Li, Z. Zeng, A small leakage detection approach for oil pipeline using an inner spherical ball, *Process Safety and Environmental Protection* 124 (2019) 279 – 289.
- [14] L. Zheng, Y. Yang, Q. Tian, Sift meets cnn: A decade survey of instance retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (5) (2018) 1224–1244.
- [15] B. Tan, Q. Peng, X. Yao, C. Hu, Z. Xu, Z. Zhang, Character recognition based on corner detection and convolution neural network (2017) 503–507.
- [16] F. Li, W. Wang, J. Xu, J. Yi, Q. Wang, Comparative study on vulnerability assessment for urban buried gas pipeline network based on svm and ann methods, *Process Safety and Environmental Protection* 122 (2019) 23 – 32.
- [17] P. Arpaia, U. Cesaro, M. Chadli, H. Coppier, L. De Vito, A. Esposito, F. Gargiulo, M. Pezzetti, Fault detection on fluid machinery using hidden markov models, *Measurement* 151 (2020) 107126.
- [18] X. Li, G. Chen, H. Zhu, Quantitative risk analysis on leakage failure of

1
2
3
4
5
6
7
8
9 submarine oil and gas pipelines using bayesian network, Process Safety
10 and Environmental Protection 103 (2016) 163 – 173.
11
12

- 13
14 [19] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553)
15 (2015) 436–444.
16
17
18 [20] Z. Li, H. Zhang, D. Tan, X. Chen, H. Lei, A novel acoustic emission
19 detection module for leakage recognition in a gas pipeline valve, Process
20 Safety and Environmental Protection 105 (2017) 32–40.
21
22
23
24 [21] R. P. D. Cruz, F. V. D. Silva, A. M. F. Fileti, Machine learning and
25 acoustic method applied to leak detection and location in low-pressure
26 gas pipelines, Clean Technologies and Environmental Policy (2020) 1–12.
27
28
29
30
31 [22] J. Sun, Q. Xiao, J. Wen, F. Wang, Natural gas pipeline small leakage
32 feature extraction and recognition based on lmd envelope spectrum
33 entropy and svm, Measurement 55 (2014) 434–443.
34
35
36
37 [23] M. Zadkarami, M. Shahbazian, K. Salahshoor, Pipeline leak diagnosis
38 based on wavelet and statistical features using dempster–shafer classifier
39 fusion technique, Process safety and environmental protection 105
40 (2017) 156–163.
41
42
43
44
45
46
47
48
49
50 [24] J. Wang, L. P. Tchapmi, A. P. Ravikumar, M. McGuire, C. S. Bell,
51 D. Zimmerle, S. Savarese, A. R. Brandt, Machine vision for natural gas
52 methane emissions detection using an infrared camera, Applied Energy
53 257 (2020) 113998.
54
55
56
57
58
59
60
61
62
63
64
65

- [25] J. Bae, D. Yeo, D. Yoon, S. W. Oh, G. J. Kim, N. Kim, C. Pyo, Deep-learning-based pipe leak detection using image-based leak features, in: 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 2361–2365.
- [26] J. SUN, Y. QIAO, J. WEN, Intelligent aperture identification combining compressed data acquisition with sparse filtering-based deep learning towards natural gas pipeline leak, Structural Health Monitoring 2017 (shm) (2017).
- [27] A. Fukane, S. Sahare, Noise estimation algorithms for speech enhancement in highly non-stationary environments, International Journal of Computer Science Issues 8 (03 2011).
- [28] M. Rastegari, V. Ordonez, J. Redmon, A. Farhadi, Xnor-net: Imagenet classification using binary convolutional neural networks, in: European Conference on Computer Vision, Springer, 2016, pp. 525–542.
- [29] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu, Convolutional neural networks for speech recognition, IEEE/ACM Transactions on audio, speech, and language processing 22 (10) (2014) 1533–1545.
- [30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets,

1
2
3
4
5
6
7
8
9 atrous convolution, and fully connected crfs, IEEE transactions on
10 pattern analysis and machine intelligence 40 (4) (2017) 834–848.
11
12

- 13 [31] A. El-Sawy, E.-B. Hazem, M. Loey, Cnn for handwritten arabic digits
14 recognition based on lenet-5, in: International Conference on Advanced
15 Intelligent Systems and Informatics, Springer, 2016, pp. 566–575.
16
17 [32] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with
18 deep convolutional neural networks, in: Advances in neural information
19 processing systems, 2012, pp. 1097–1105.
20
21 [33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-
22 scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
23
24 [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov,
25 D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with
26 convolutions, in: Proceedings of the IEEE conference on computer vision
27 and pattern recognition, 2015, pp. 1–9.
28
29 [35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image
30 recognition, in: Proceedings of the IEEE conference on computer vision
31 and pattern recognition, 2016, pp. 770–778.
32
33 [36] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.
34
35 [37] J. Han, C. Moraga, The influence of the sigmoid function parameters on
36 the speed of backpropagation learning, in: International Workshop on
37 Artificial Neural Networks, Springer, 1995, pp. 195–201.

- [38] B. L. Kalman, S. C. Kwasny, Why tanh: choosing a sigmoidal function, in: International Joint Conference on Neural Networks, 1992.
- [39] A. F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375 (2018).
- [40] F. Demir, M. Turkoglu, M. Aslan, A. Sengur, A new pyramidal concatenated cnn approach for environmental sound classification, Applied Acoustics 170 (2020) 107520.
- [41] K. Zhao, H. Jiang, Z. Wang, P. Chen, B. Zhu, X. Duan, Long-term bowel sound monitoring and segmentation by wearable devices and convolutional neural networks, IEEE Transactions on Biomedical Circuits and Systems (2020).
- [42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014) 1929–1958.
- [43] Z. Jia, L. Ren, H. Li, T. Jiang, W. Wu, Pipeline leakage identification and localization based on the fiber bragg grating hoop strain measurements and particle swarm optimization and support vector machine, Structural Control and Health Monitoring 26 (2) (2019) e2290.
- [44] Z. Qu, H. Feng, Z. Zeng, J. Zhuge, S. Jin, A svm-based pipeline leakage detection and pre-warning system, Measurement 43 (4) (2010) 513–519.

Enhanced Spectrum Convolutional Neural Architecture: An Intelligent Leak Detection Method for Gas Pipeline

Fangli Ning^{a,*}, Zhanghong Cheng^a, Di Meng^a, Shuang Duan^a, Juan Wei^b

^a*School of Mechanical Engineering, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, Shaanxi, China*

^b*School of Telecommunications Engineering, Xidian University, 2 South Taibai Road, Xi'an, Shaanxi, China*

Abstract

In this work, a novel convolutional neural architecture (SE-CNN), which combines Spectrum Enhancement (SE) and Convolutional Neural Network (CNN), is proposed to detect the leak of gas pipeline. The SE has the effect of enhancing the leak signals and reducing background noise. CNN can automatically extract leak features and realize leak diagnosis. The experimental results show that the SE-CNN can achieve an average accuracy of 94.3% for 6 categories and only requires 1.04 seconds of detection time. In this experiment, the diameters of the main pipeline and the branch pipeline are 125 mm and 25 mm. Due to its excellent accuracy and efficiency, the proposed enhanced spectrum convolutional neural architecture paves the way for real-time leak detection in industrial environments, which can ensure the process safety of gas pipeline transportation. Under strong background noise, the average accuracy of the

*Corresponding author

Email address: ningf1@nwpu.edu.cn (Fangli Ning)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

SE-CNN can reach 94.3%, which is 33%, 3.7% higher than that of SVM and CNN. In particular, the SE can be regarded as a data compression method, which can significantly reduce the original data size. The training time of the SE-CNN is 539 seconds, reducing 90.6% compared with CNN.

Key words: Spectrum enhancement; Convolutional neural network; Leak detection

1. Introduction

The leak from the gas pipeline often occurs in the process of engineering practice, which not only endangers residents' safety and environment but also wasting energy affects a country's economy [1, 2]. Nowadays, gas pipeline leak detection has been becoming a hotspot in the oil and gas industry and a top priority to ensure the process safety of gas pipeline transportation. As shown in Fig.1, the current process of gas pipeline leak detection generally includes the four steps. Data collection is the first step in the process. The quantity and quality of the data determine the leak detection effect. At present, the raw data collected mainly include vibration signals [3], acoustic signals [4, 5] and images [6]. There are some unique advantages for acoustic signals, such as the long propagation distance, non-contact measurement and no blind area. At the same time, the leak acoustic signal has a frequency distribution range from infrasound to ultrasonic. Therefore, acoustic signals will be used as the raw data in this work. The second step, denoising [7], is to improve the quality of the data,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

such as the spectral subtraction [8], noise estimation [9], filtering algorithms [10, 11] and total variation (TV) method [12]. Feature extraction is the third step, which aims to extract key information and reduce the data dimension. Numerous feature extraction approaches have been developed, such as Mel-frequency cepstral coefficients (MFCC) [13], scale-invariant feature transform (SIFT) [14], histogram of oriented gradients (HOG) [15] which represent the raw acoustic signals and images. The last step is leak diagnosis to select an appropriate diagnosis algorithm as a classifier. There are also lots of machine learning algorithms, such as support vector machine (SVM) [16], hidden Markov model (HMM) [17], Bayesian network [18], deep learning (DL) [19], which have been used for leak diagnosis.

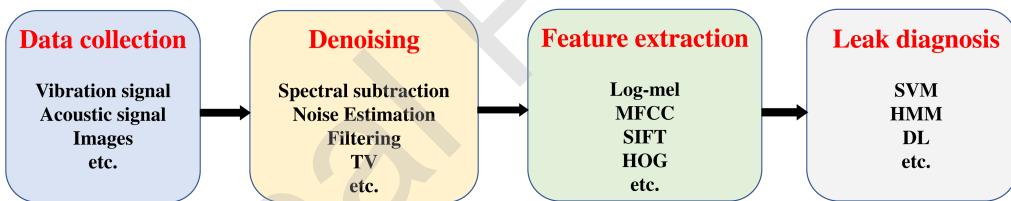


Fig. 1. The process of gas pipeline leak detection.

Based on acoustic signals, many machine learning algorithms have been proposed for gas pipeline leak detection. Li *et al.* [20] proposed a leak detection scheme based on kernel principal component analysis (kernel PCA) and SVM classifier for recognizing the leak. Cruz *et al.* [21] proposed a technique that combined acoustic sensors and machine learning algorithms to identify and locate leaks in low-pressure gas pipelines. Sun *et*

al. [22] proposed a small leak feature extraction and recognition method based on SVM. With the rapid development of artificial intelligence in recent years, the accuracy of gas pipeline leak detection has been improved greatly by deep learning technology. Zadkarami *et al.* [23] extracted wavelet features and used them as the inputs of a multi-layer perception neural network (MLPNN) classifier to determine the leak state. Wang *et al.* [24] developed a computer vision approach based on optical gas imaging (OGI) to detect the leak. Bae *et al.* [25] proposed a deep-learning leak detection technique that used trajectory-based image features extracted from time-series acoustic data. They developed root-mean-square-pattern and frequency-pattern images to reflect the leak signal characteristics and used them as the inputs of residual networks. Sun *et al.* [26] proposed an intelligent pipeline leak detection method combining compressed sensing and deep learning theory, which can achieve compressed sampling, adaptive feature extraction and recognition.

Although the above algorithms have greatly promoted the progress of gas leak detection, there is still a long way to the application in the industrial environment. The performance of the leak detection algorithm is mainly limited to two aspects. Firstly, the effect of the leak diagnosis system depends on the denoising algorithm. Many denoising methods have a significant effect on the suppression of stationary noise. Due to the continuous and stable state of the leak, the leak signal is stationary, but the background noise signal in application scenes is non-stationary, which

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

makes it difficult to apply these denoising methods to the industrial environment. What's more, these denoising methods suffer from a weakness to achieve a trade-off between noise reduction and the undistorted target signal [27]. Secondly, in the complexity of the industrial environment, the raw audio contains a lot of redundant information, which makes feature extraction difficult and challenging. Moreover, feature extraction depends on prior knowledge and is specific to the task, so it must be redesigned for every new industrial environment.

In order to overcome the above difficulties and improve the recognition effect of the leak detection system, an architecture combining spectrum enhancement (SE) and CNN is proposed in this work. Convolution operations are used in the SE method to enhance the stationary leak signals while eliminating the non-stationary noise signals, thereby achieving the goal of enhancing the leak signal characteristics while reducing noise. With the rapid development of deep learning algorithms, convolutional neural networks have realized great achievements in many fields, such as Image Classification [28], Speech Recognition [29], and Image Segmentation [30]. There are many representative CNN models, LeNet-5 [31] is originally used as handwritten Arabic digit recognition in the bank, AlexNet [32], VGGNet [33], GoogleNet [34], ResNet [35] are the best models that stand out from ImageNet competition. CNN can achieve end-to-end leak detection, which can directly obtain the detection results from the raw audio signal. The last three steps in the process (Fig.1) are

merged into one step by the proposed algorithm, so the process of gas pipeline leak detection will become more efficient.

The rest of this paper is organized as follows. In Sec.2, the basis of CNN architecture is described in detail. In Sec.3, the method of spectrum enhancement and the SE-CNN are illustrated. The experimental results and analysis are presented in Sec.4. Conclusions and future work are summarized in Sec.5.

2. Convolutional neural network

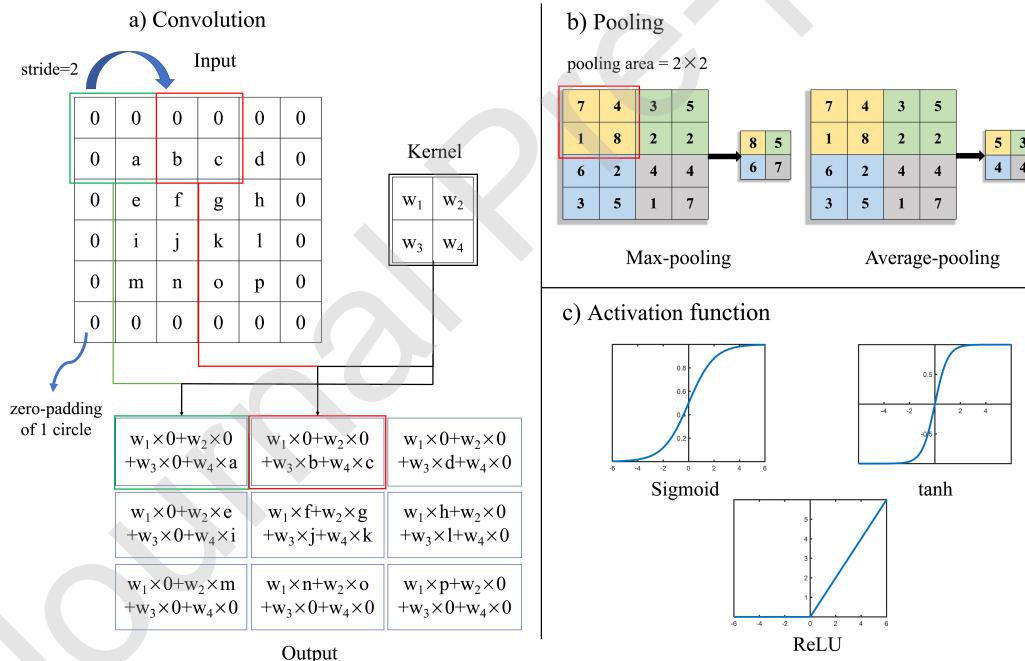


Fig. 2. Three components of convolutional neural networks. (a) Feature extraction of input by convolution operation. (b) Two main methods of pooling. (c) Three main activation functions for neural networks.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

As shown in Fig.2, convolutional neural networks have three essential components: convolution, pooling, and activation function:

a) Convolution is the most important component making neural networks gain the ability to extract the buried features. It improves the neural network's performance through three ideas: sparse connectivity, parameter sharing, and equivariant representations [36]. The convolution kernel size is much smaller than the input to achieve sparse connections, which can help reduce the number of parameters and storage requirements. Parameter sharing is realized by repeatedly using identical weight parameters during the convolution operation of the same layer, where weights are the parameters w_1, w_2, w_3, w_4 of kernel in Fig.2(a). The convolution O of A and B can be calculated as

$$O(s, t) = (A * B)(s, t) = \sum_m \sum_n A(s + m, t + n)B(m, n), \quad (1)$$

where O is the output, A is the input, B is the convolution kernel, $*$ denotes convolution operation.

Due to the parameter sharing, the convolution layer has some translation invariance, which means that no matter where the target is in the input, the same output will be obtained at the corresponding area.

b) Pooling can reduce each layer's output size, and improve the computational efficiency and robustness of the extracted features. The max-pooling $\alpha = \max\{\alpha_i\}, \alpha_i \in \mathcal{N}$ and the average-pooling

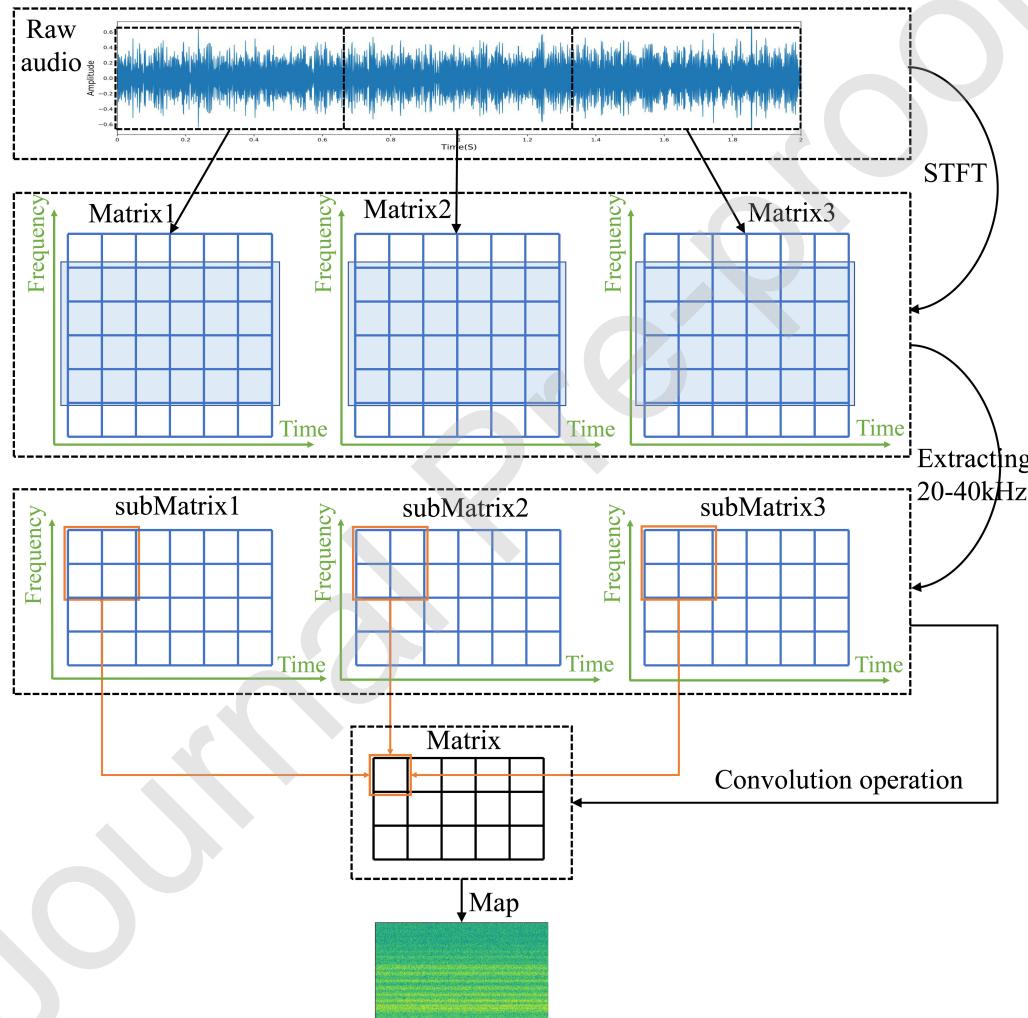
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

$\alpha = \frac{1}{k} \sum_{i=1}^k \alpha_i, \alpha_i \in \mathcal{N}$ are two of the most widely used pooling methods, where α is the output, \mathcal{N} is the area of pooling operation, k is the number of elements in the pooling area, α_i is the value in the pooling area. The max-pooling is achieved by extracting the maximum value to represent the pooling area, which has been proved to be one of the most effective pooling methods. The average-pooling is achieved using the mean in the pooling kernel representing the area, which is usually used in deep neural networks and the last layer of fully convolutional networks.

c) Activation functions give the neural networks a strong fitting ability. There are several typical activation functions for different application requirements. The sigmoid [37] is one of the most basic activation function, but it is rarely used now because of the gradient vanishing. The tanh [38] is essentially a variant of sigmoid, which has a better effect, but it still cannot solve the gradient vanishing problem. The rectified linear units (ReLU) [39] has been the most popular activation function, which can well solve the problems of gradient vanishing and gradient exploding. Simultaneously, due to its simple function and derivative, forward and backward propagation is significantly accelerated, so the overall training speed of the neural networks is greatly improved.

1
2
3
4
5
6
7
8
9 3. The new framework combining spectrum enhancement and
10
11 CNN

12
13
14 3.1. Spectrum enhancement
15
16
17



54 Fig. 3. The process of spectrum enhancement.
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

A spectrum enhancement method imposed on the matrix of the short-time Fourier Transform (STFT) is proposed in this study, which can enhance the stationary leak signals while eliminating the non-stationary noise signals. As a time-frequency analysis method, STFT is widely used in signal processing. STFT can clearly reflect the features of the raw audio than one-dimensional time series [40], it needs less processing and can save more original information than MFCC [41]. The ultrasonic domain's time-frequency spectrum (20023 Hz-40047 Hz) is extracted, which aims to reduce the impact of the noise. As shown in Fig.3, the novel method process mainly follows the four steps:

- (1) Divide each sample (audio) into B blocks of equal length.
- (2) Perform STFT on each block separately to obtain the matrix.

$$\text{STFT}\{x[n]\}(m, \omega) \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}, \quad (2)$$

where $X(m, \omega)$ is the matrix of STFT, $x[n]$ represents the original signal, $w[n-m]$ is a window function centred at the time m .

In $X(m, \omega)$, the horizontal axis represents the time, the vertical axis represents the frequency, the element of the matrix represents the power spectrum described by the decibel scale.

- (3) Extract the submatrix which represents the ultrasonic domain in each matrix.

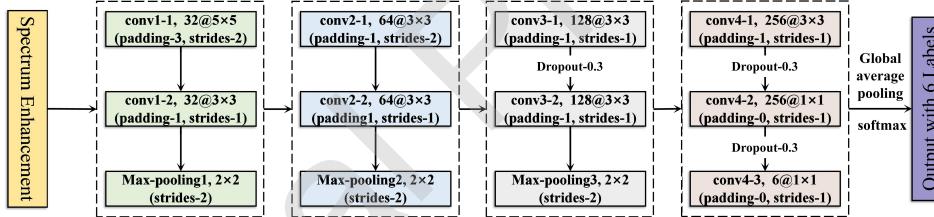
- 1
2
3
4
5
6
7
8
9 (4) Perform convolution operation on the elements in the kernels of size
10
11 $K \times K$ to slide on each submatrix with the stride of S , the enhanced
12 rule is calculated by,
13
14

$$M = M^{(1)} \odot \dots \odot M^{(b)} \odot \dots \odot M^{(B)}, \quad M^{(i)} \subset X^{(i)}$$

$$y = \sum_{p=1}^K \sum_{q=1}^K M(p, q), \quad (3)$$

22
23 where $X^{(i)}$ is the i -th submatrix, $M^{(i)}$ is the kernel in the i -th submatrix.
24
25 \odot denotes Hardmdard product (component-wise multiplication), $M(p, q)$
26
27 is the (p, q) -th element of the matrix M , y is the output.
28
29

30 3.2. The architecture of SE-CNN
31
32



41
42 **Fig. 4.** The architecture of SE-CNN. Conv1-1 is the name of the convolutional layer,
43 32@5x5 represents that there are 32 kernels, and the size of kernels is 5x5, padding-3
44 means performing zero-padding operation of 3 circles on the edge of the Input, strides-2
45 indicates that the stride of the convolution kernel sliding on the input is 2 pixels.
46
47

48 The specific architecture of SE-CNN applied in this work is illustrated
49 in Fig.4. The architecture design is inspired by VGGNet, which has shown
50 good mobility and scalability in many deep learning tasks. There is a
51 consensus that the more layers a neural network, the better its
52 performance. However, with the layers increasing, the number of
53
54
55
56
57
58

parameters increases exponentially. Overfitting, which greatly influences the algorithm generalization, usually occurs in a neural network with lots of parameters. Therefore, to reduce the generalization error of the algorithm and make it still maintain a good recognition effect for new samples (without training), the global average pooling is chosen instead of fully connected layers at the end of the network, which dramatically reduces the number of parameters. And the dropout[42] method is adopted in the SE-CNN architecture, which can randomly deactivate the connected units so that the network does not rely on some specific units, and its architecture will be simplified. The whole architecture includes spectrum enhancement, nine convolution layers, and four subsampling layers. The spectrum enhancement is added first. The max-pooling is selected in the first three subsampling layers, and the last layer is the average-pooling. ReLU is used as the activation function in all layers except the last layer, the last layer's activation function is softmax.

After the SE-CNN architecture is determined, the cost function is defined as

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k y_j^{(i)} \log(\hat{y}_j^{(i)}), \quad (4)$$

where $y_j^{(i)}$ represents the true output of the i -th sample on the j -th unit, $\hat{y}_j^{(i)}$ represents the predicted output of the i -th sample on the j -th unit, m represents the number of samples on one iteration and is set to be 32 in the experiment, k represents the number of units in the output layer.

The forward propagation of ℓ -th layer is given as

$$\mathbf{x}_j^\ell = f \left(\sum_{i \in M_j} \mathbf{x}_i^{\ell-1} * \mathbf{k}_{ij}^\ell + b_j^\ell \right), \quad (5)$$

where M_j represents a selection of input maps, i is the index of input map, j is the output maps, k is the convolutional kernel. Each output map is given an additive bias b , for a particular output map, the input maps will be convolved with distinct kernels. Based on Eq.(4), the gradient term of output layer (L -th layer) can be calculated as

$$\delta^{(L)} = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k \frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \odot f'(z^{(L)}), \quad (6)$$

where $\delta^{(L)}$ is the “errors” of L -th layer, which can be thought of as “sensitivities” of each unit with respect to perturbations of the parameters.

Error back-propagation follows the chain derivation rule.

In order to speed up the convergence and improve the generalization ability of SE-CNN, Batch Normalization (BN) is applied in our proposed model. The calculation process of BN is as follows

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (7)$$

$$y_i = \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad (8)$$

Eq.(7) is the normalization process where x_i is the input, μ_B is the mean of x_i , σ_B^2 is the variance of x_i , ϵ is a constant added to variance for numerical stability. Eq.(8) is the process of scale and shift where γ and β are introduced to prevent damage to the feature distribution of the previous layer data.

The back propagation of ℓ -th layer is given as

$$\delta^\ell = W^{\ell+1} (f' (z^\ell) \odot up (\delta^{\ell+1})) , \quad (9)$$

where $\ell = 1, 2, \dots L - 1$, W is the weight matrix of ℓ layer, up denotes upsampling. Update the weights and bias with

$$\begin{aligned} W_{(i+1)}^\ell &= W_{(i)}^\ell - \eta x^{\ell-1} (\delta^\ell)^T, \\ b_{(i+1)}^\ell &= b_{(i)}^\ell - \eta (\delta^\ell), \end{aligned} \quad (10)$$

where $W_{(i+1)}^\ell$ and $b_{(i+1)}^\ell$ are the value of W^ℓ and b^ℓ at the i -th iteration, η is the learning rate.

After getting the output by global average pooling, the softmax function is used for classification.

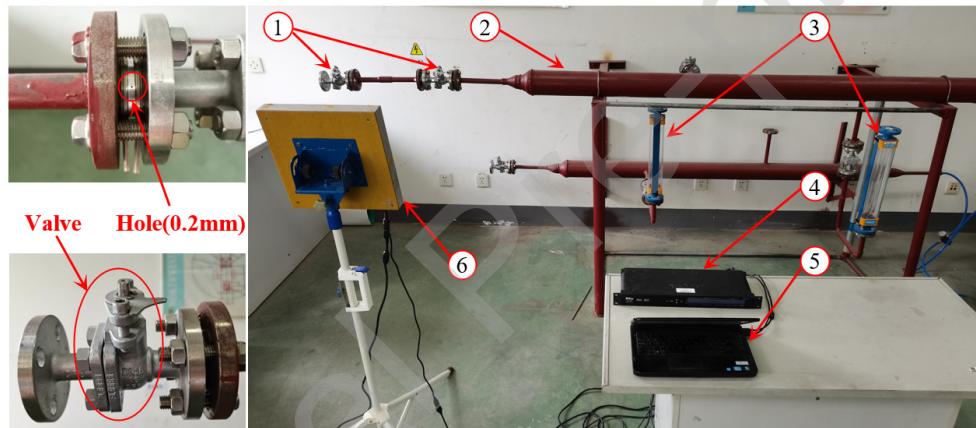
$$S_i = \frac{e^i}{\sum_{j=1}^J e^j} \quad (11)$$

where j represents the number of units in layer, i represents i -th unit of j units, S_i is the predicted probability of label i . Then, choose the largest predicted probability as the predicted result.

1
2
3
4
5
6
7
8
9 4. Experiments
10
11

12 To verify the effectiveness of the proposed method, the experimental
13 results of the SE-CNN and CNN approaches are compared in this section.
14 At the same time, in order to achieve the best performance of leak
15 detection, the mesh search method is used to search the optimal parameters
16 (B – block, S – stride, K – kernel) for spectrum enhancement.
17
18
19
20
21
22
23

24 4.1. Experiment instruments and dataset
25
26



41 Fig. 5. Experimental system: (1) valve; (2) gas pipeline; (3) flowmeter;
42 (4) signal
43 collector; (5) laptop; (6) sound receiver.
44
45

46 As shown in Fig.5, the entire experimental system consists of gas
47 pipelines, a sound receiver, a signal collector that is used to convert analog
48 signals into digital signals and the sampling rate is set to be 96 kHz, a
49 laptop which is used to analyze and process the raw data. In the process of
50 data collection, the experimental conditions, such as the valve opening, leak
51 hole size, and sampling distance (0.8m-1.2m), are changed continuously to
52
53
54
55
56
57
58

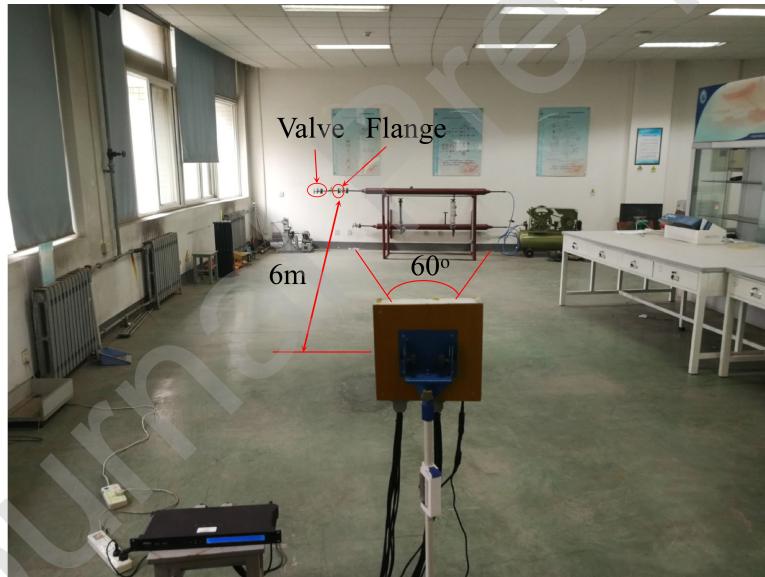
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

simulate various real leak conditions. This helps to increase the diversity of the samples. In this experiment, the diameters of the main pipeline and the branch pipeline are 125 mm and 25 mm. Collecting leak acoustic signals in the laboratory and background noise (fan noise signals) in the pipe gallery, a total of 8.5 hours real audio is obtained. Synthesizing the laboratory signals with the background noise to simulate gas pipeline leak under the industrial environment. Ultimately, a total of 13 hours audio is obtained. The audio signals are cut into short clips of 2 seconds, each of them is regarded as a sample. There are 23040 samples in total, which are randomly divided into a training set (70%) and a test set (30%). The descriptions of different classes are listed in Tab.1. The signal-to-noise ratio (SNR) of label 2, 3 is 5 dB, which can be regarded as leaks under weak noise. The SNR of synthesizing signal label 5, 6 is -25 dB, which can be regarded as leaks under strong noise. The training and predict processes of the SE-CNN and CNN are done on the Intel(R) Core(TM) i5-8300H @ 2.30 GHz system with 8GB-RAM and 4GB-GPU RAM.

In order to verify the applicability of the proposed SE-CNN in the industrial environment where the leak location is unknown, audio data were collected at 6m away from the gas pipeline. Figure 6 shows the experimental scene with a sampling distance of 6m. Under different sampling distances, 3840 hole and valve leak audio were collected, respectively, each of which is 2 seconds.

1
2
3
4
5
6
7
8
9 **Table 1**
10 Description of the six categories dataset.
11
12

Label	Background noise	Types of leaks	Training Set	Test Set
1	weak noise	no leak	2688	1152
2		hole leak		
3		valve leak		
4	strong noise	no leak		
5		hole leak		
6		valve leak		



49 **Fig. 6.** The experimental system at 6m sampling distance.
50
51
52

53 Figure 7 shows the relationship between sampling distance and detectable
54 pipeline length. The directivity of the sound receiver is 60 degrees. Therefore,
55
56
57
58

when the sampling distance is 6 m, the sound receiver can detect the pipeline up to 6.93m.

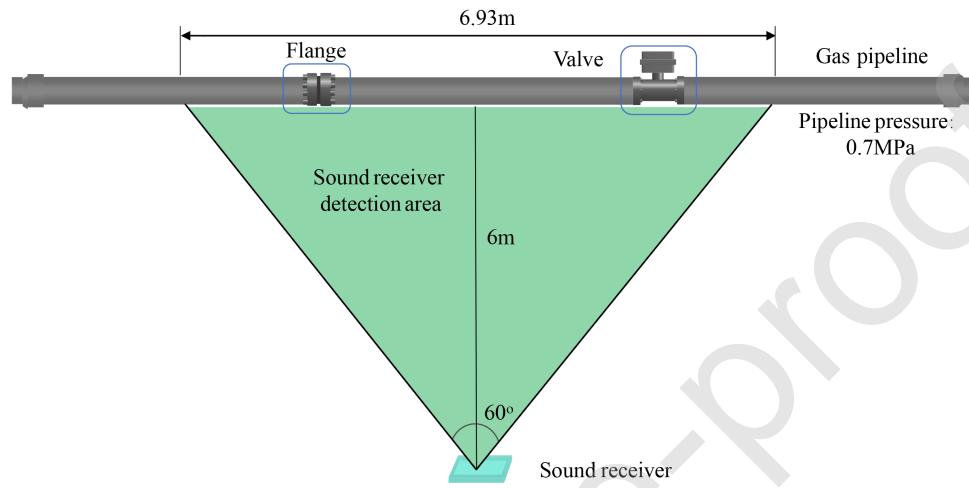


Fig. 7. Schematic diagram of relationship between sampling distance and detectable pipeline length.

The recall ratio r , precision ratio p , and f_1 -score are used to evaluate the performance of the algorithm and choose the most optimal parameters of the spectrum enhancement. The r and p are respectively defined as:

$$r = \frac{TP}{TP + FN}, \quad (12)$$

$$p = \frac{TP}{TP + FP}. \quad (13)$$

If the sample is positive and it is classified as positive, it is counted as true positive (TP); if it is classified as negative, it is considered as false negative (FN). If the sample is negative and it is classified as negative it is considered

as true negative (TN); if it is classified as positive, it is counted as false positive (FP).

The f_1 -score is a more comprehensive criterion to evaluate the performance. It is defined as

$$f_1 = \frac{2pr}{p+r} \quad (14)$$

The accuracy is the most popular to evaluate the performance of the algorithm, which is defined as

$$\text{accuracy} = \frac{N_{\text{correct}}}{N_{\text{all}}}, \quad (15)$$

where N_{correct} represents the number of samples that are predicted correctly, N_{all} is the number of all samples.

4.2. Results and analysis

The enhanced matrices are visualized to show the effect of the spectrum enhancement in Fig.8. The first column is the short-time Fourier Transform spectrums of the 6 categories data. The second column is the spectrums of corresponding data, which are obtained by the spectrum enhancement. The spectrums labeled 1, 2, 3 from weak noise environment have high similarity. The spectrums labeled 4, 5, 6 from strong noise environment also have high similarity. The spectral enhancement highlights the difference in time-frequency diagrams of different categories shown on Fig.8, which can

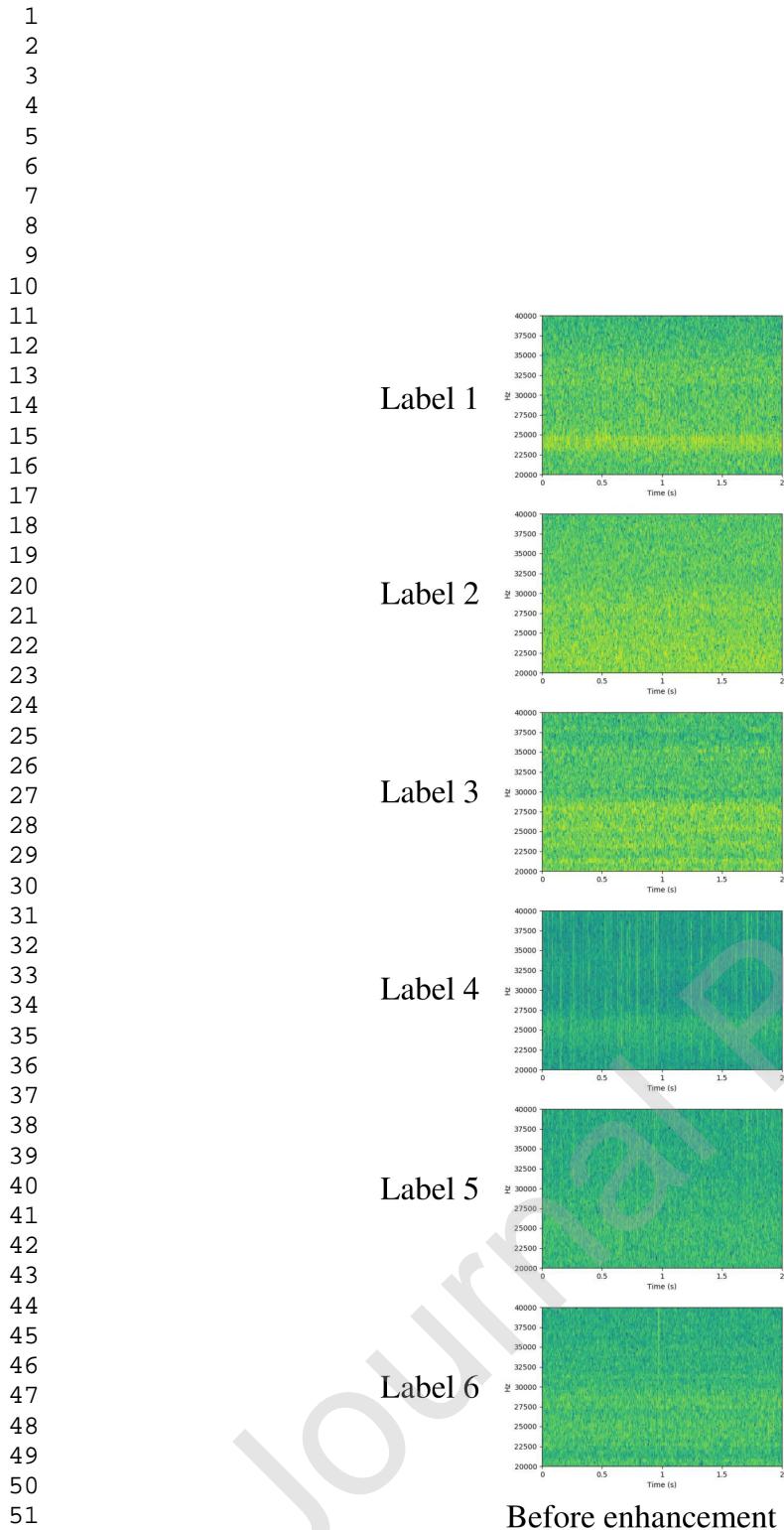


Fig. 8. Comparison of before and after spectrum enhancement for 6 categories.

1
2
3
4
5
6
7
8
9
10
11
12
13
14

only give a qualitative, but not quantitative, spectral enhancement effectiveness. Next, the quantitative verification will be carried out by the comparison of experimental results.

Figure 9 shows that the results of the architectures when B (the number of blocks) are different values. $B = 1$ means there is no spectrum enhancement, which represents the results of the CNN approach. As shown in Fig.9(a), the framework of SE-CNN has better performance than CNN in every class. Especially under the strong background noise environment, the improvement is reaching nearly 5%, the most notable is that the improvement of label 5 increases more than 8%. The performance of SE-CNN under the strong noise environment is excellent, which confirms that the spectrum enhancement can indeed play a specific role in boosting the performance. Simultaneously, the performances of SE-CNNs with different B are close, and the difference in overall performance is within 0.4%. Increasing the value of B can make the spectrum more effectively enhanced, but the size of the spectrum will be reduced correspondingly, which will inevitably lose some vital information. So the algorithm performance cannot be improved only by increasing B . It can be seen from Fig.9(b), the r of CNN is less than the SE-CNN in each class except label 4, which is more than 1.6% compared with that of the SE-CNN. However, the r of label 5 is 68.1%, which is lower than 15.1% compared to the SE-CNN. On the contrary, as shown in Fig.9(c), the p of label 4 is 73.3% and significantly lower than 80.6% of the SE-CNN, but the p of label 5 is 99.8% which is slightly higher than 97.1% of the SE-CNN. The reason for

the low precision ratio of label 4 is that the features of strong background noise are not obvious in the ultrasonic frequency band. It can be seen from the Fig.9(d), the improvement in accuracy of the label 1, 2, 3 is between 1.0% ~ 1.8%. The improvement in accuracy of the label 4, 5, 6 is between 3.5% ~ 4.3%. It confirms that the SE-CNN is more suitable for industrial environments than CNN. As far as the experimental results are concerned, The number of blocks is chosen as 4 to achieve a good performance in this work.

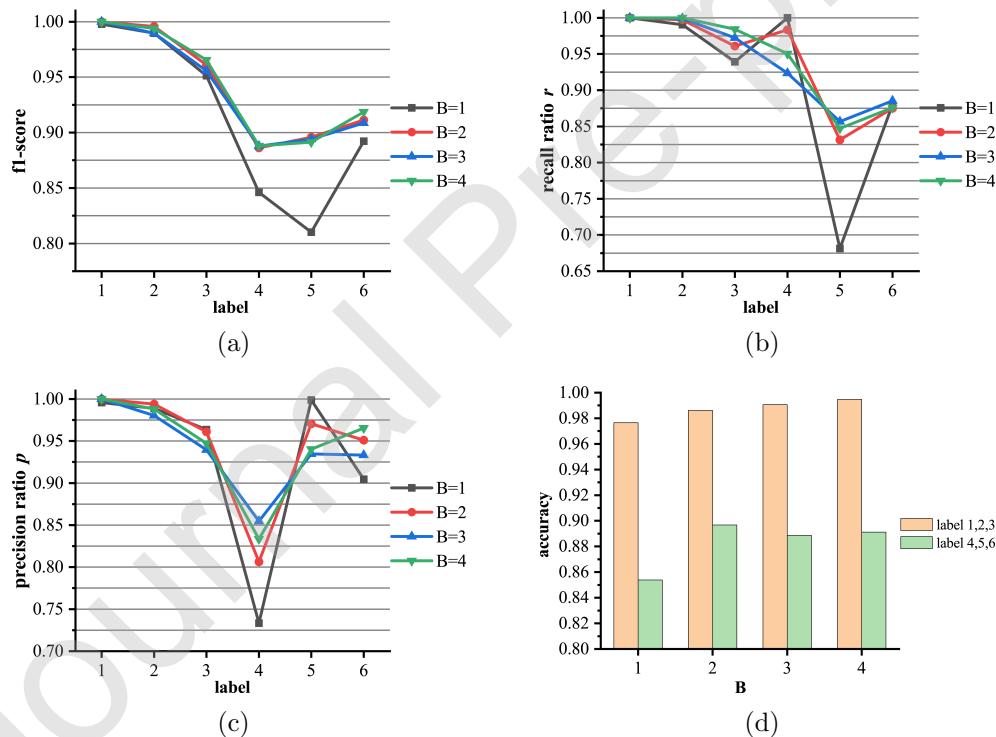


Fig. 9. (a) f_1 -score (b) Recall ratio r (c) Precision ratio p (d) Accuracy with different B (the number of blocks).

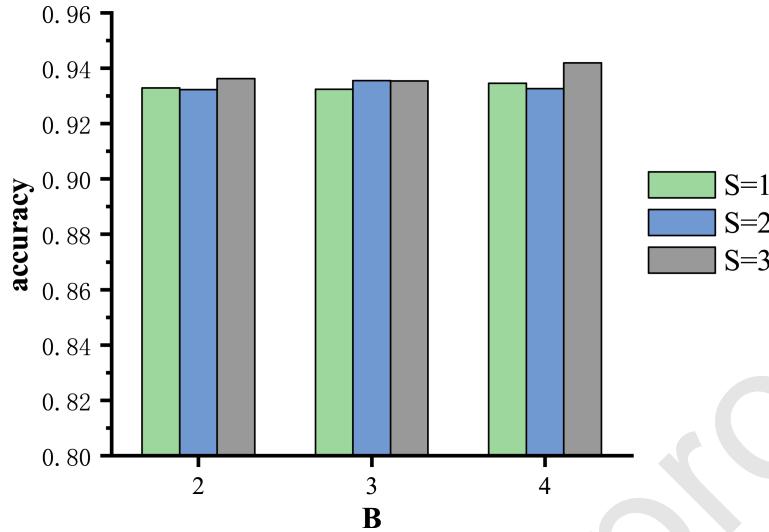


Fig. 10. Accuracy with different S (the number of strides).

Table 2
Comparison of computing cost with different S (stride).

Parameters	Training time(s) 16128 samples	Test time(s) 6912 samples	Predict time(s) 1 sample
$S = 1 (B = 2, K = 3)$	6140.26	74.32	1.23
$S = 2 (B = 2, K = 3)$	1587.78	58.93	1.07
$S = 3 (B = 2, K = 3)$	811.24	38.54	1.05

As shown in Fig.10, the overall accuracy of SE-CNNs with different S (the number of strides) ranges from 93.2% to 94.1%, which indicates that S does not have much impact on the performance of SE-CNNs. Noticeably, the larger S can greatly reduce the training time of the neural network, which is shown in Tab.2. In fact, S changes the size of the neural network's inputs, the size of the sample is 188×214 when $S = 1$, the size of the sample is 94×107

when $S = 2$, and then the size of the sample is 63×72 when $S = 3$. Changing the size of the sample by changing the S can greatly reduce the computing cost and speed up the training process. Nevertheless, if S is too large, the sample will lose lots of key information, which is definitely detrimental to the performance of SE-CNNs. In order to minimize the consumption of computing power, S is finally set to be 3 in this work.

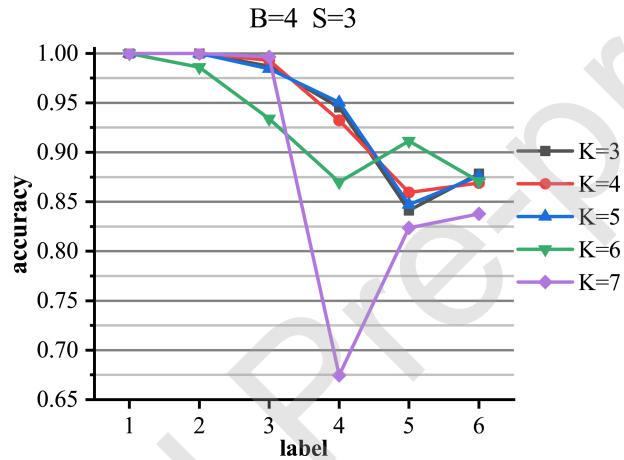


Fig. 11. The accuracy of SE-CNNs with different K (the number of kernels).

The value of K (the number of kernels) mainly depends on the value of S . When $K < S$, some elements in the matrix are not involved in the convolution operation when spectrum enhancement is performed, which will inevitably waste some information and undermine the performance of the SE-CNN, when $K \geq 2S$, the difference between adjacent elements of the matrix after spectrum enhancement will be reduced, so the feature of the spectrum be weakened. Therefore, it is also harmful to the performance of the SE-CNN. As shown in Fig.11, there is a dramatic decrease in the accuracy of

SE-CNNs when $K \geq 2S$. The experiment result shows that a SE-CNN can obtain an excellent performance when $K = S + 1$ or $K = S + 2$. This is due to the fact that K is slightly larger than S , which can maximize the effect of spectrum enhancement. K is finally set to be 5 in this work.

Table 3
Comparison results of SVM, CNN and SE-CNN.

Label	1	2	3	4	5	6	f_1 -score	Acc	AvgAcc
SVM (The accuracy of 6 categories = 76.1%)									
1	1117	25	10	0	0	0	0.979	96.9%	
2	2	1078	72	0	0	0	0.948	93.6%	96.1%
3	0	20	1126	0	0	0	0.947	97.7%	
4	1	0	0	867	170	114	0.611	75.3%	
5	4	0	0	605	393	150	0.402	34.1%	56.1%
6	0	0	19	215	243	677	0.647	58.8%	
CNN (The accuracy of 6 categories = 91.5%)									
1	1152	0	0	0	0	0	0.998	100%	
2	5	1141	0	6	0	0	0.990	99%	97.7%
3	0	1	1082	0	0	69	0.951	93.9%	
4	0	0	0	1152	0	0	0.846	100%	
5	0	11	1	317	785	38	0.810	68.1%	85.4%
6	0	1	40	96	1	1014	0.892	88%	
SE-CNN (B = 4, S = 3, K = 5; The accuracy of 6 categories = 94.3%)									
1	1152	0	0	0	0	0	1.000	100%	
2	0	1152	0	0	0	0	1.000	100%	99.5%
3	0	1	1134	0	0	17	0.966	98.4%	
4	0	1	0	1095	47	10	0.888	95.1%	
5	0	13	0	154	976	9	0.891	84.7%	89.1%
6	0	0	63	65	15	1009	0.919	87.3%	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Some scholars have used a support vector machine (SVM) for pipeline leakage detection [43, 44]. The classification results of SVM, CNN and SE-CNN with the optimal parameters are shown in Tab.3. The results in Tab.3 are obtained on a dataset with sampling distance of 0.8m-1.2m. Compared SVM, CNN with SE-CNN, the accuracy of leak detection under the weak background noise increases from 96.1%, 97.7% to 99.5%, which achieves a 3.4%, 1.8% improvement, respectively. In the industrial environment, strong background noise makes it difficult to detect the leak in the gas pipeline, but the improvement of leak detection under strong background noise is significant. The accuracy of label 4, 5, 6 increases from 56.1%, 85.4% to 89.1%, which achieves a 33%, 3.7% improvement.

Table 4 illustrates the computing cost of CNN and the SE-CNN. Compared with CNN, the training time, testing time and predict time of the SE-CNN are reduced by 90.6%, 62.5% and 15.4%, respectively. While using two-dimensional discrete convolution operation to reduce noise and enhance the striations of spectrums, increasing the stride (S) to speed up the convolution operation process, so that the accuracy and efficiency of the SE-CNN are significantly improved.

According to the parameters determined above ($B = 4$, $S = 3$, $K = 5$), these audio data collected at 6m away from the gas pipeline were input into SE-CNN as a test set to confirm its practicality in the industrial environment. Table 5 lists the results of these data on CNN and SE-CNN.

9
10 **Table 4**
11 Comparison of computing cost between SE-CNN and CNN.
12

Algorithm	Training time(s) 16128 samples	Test time(s) 6912 samples	Predict time(s) 1 sample
CNN	5746.12	70.75	1.23
SE-CNN (B = 4, S = 3, K = 5)	539.08	26.56	1.04

24 **Table 5**
25 Comparison results of CNN and SE-CNN at sampling distance of 6m.
26

Model	Metrics	Label					
		1	2	3	4	5	6
CNN	f_1 -score	0.997	0.615	0.764	0.639	0.277	0.658
	Accuracy	99.5%	47.1%	100%	95.5%	16.9%	50%
	Avg Acc			68.2%			
SE-CNN	f_1 -score	1	0.813	0.900	0.907	0.734	0.858
	Accuracy	100%	68.5%	83.1%	98.9%	82.3%	87.8%
	Avg Acc			86.8%			

44 As the sampling distance increases, the average accuracy of CNN and
45 SE-CNN for 6 categories both decreases. However, the average accuracy
46 of SE-CNN can reach 86.8%, which is 18.6% higher than CNN when the
47 sampling distance is 6m. The experimental results indicate that SE-CNN
48 can still effectively detect leaks even at long sampling distances.
49
50
51
52
53

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

5. Conclusions

The intelligent architecture (SE-CNN) for leak detection of the gas pipeline has been presented. Experiments were conducted to train the SE-CNN and test the performance of SE-CNN. In this experiment, the diameters of the main pipeline and the branch pipeline are 125 mm and 25 mm. The critical features of the study can be summarized as follow

- The spectrum enhancement method can effectively enhance the stationary leak signals and reduce noise in the aliasing signals. Therefore, the SE-CNN can achieve an average accuracy of 94.3%, which is 18.2%, 2.8% higher than SVM, CNN. Especially, the accuracy is 33%, 3.7% higher than SVM, CNN under strong background noise (SNR = -25 dB).
- The spectrum enhancement can also be regarded as a data compression method, it can significantly reduce the size of the original signal, which is good for speeding up the training process of the neural network and reducing the computing cost. The training time of SE-CNN is 539 seconds, and the predict time is 1.04 seconds, reducing 90.6% and 15.4% computing cost, respectively, compared with CNN.
- The SE-CNN keeps all the advantages of CNN. One important characteristic of CNN is to avoid the challenging feature engineering problem and achieve end-to-end processing of the raw data. On the

other hand, The SE-CNN uses the average-pooling instead of fully connected layers in the end, which dramatically reduces the risk of overfitting.

The proposed SE-CNN can satisfy two major requirements of gas pipeline leak detection, which are accuracy and efficiency. The good performance of SE-CNN brings prospects to industrial applications of gas pipeline leak detection. However, due to the leak in the industrial environments are diverse, the six categories samples in this study cannot fully cover all leak types. Therefore, more experiments need to be conducted in the future to obtain more types of leak signals such as crack leak, weld defect leak and so on. The SE-CNN can perform the real-time detection of gas pipelines to ensure the process safety of transportation. Besides, timely response to pipeline leaks is also important to risk engineering for the gas pipeline. Therefore, SE-CNN provides potential applications in industrial environments for both process safety and risk engineering.

Conflict of interest statement

The authors declared that there is no conflict of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 51675425, 52075441), Shaanxi Key Research Program

1
2
3
4
5
6
7
8
9 Project (Grant No. 2020ZDLGY06-09), Dongguan Social Science and
10 Technology Development(key) Project (Grant No. 20185071021600),
11 Science and Technology on Micro-system Laboratory Foundation (Grant
12 No. 6142804200405).
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

- [1] S. Datta, S. Sarkar, A review on different pipeline fault detection methods, *Journal of Loss Prevention in the Process Industries* 41 (2016) 97–106.
- [2] S. Bonvicini, G. Antonioni, P. Morra, V. Cozzani, Quantitative assessment of environmental risk due to accidental spills from onshore pipelines, *Process Safety and Environmental Protection* 93 (2015) 31 – 49.
- [3] Q. Wang, L. Han, X. Fan, J. Zhu, Distributed fiber optic vibration sensor based on polarization fading model for gas pipeline leakage testing experiment, *Journal of Low Frequency Noise Vibration and Active Control* 37 (3) (2018) 468–476.
- [4] Q. Xu, L. Zhang, W. Liang, Acoustic detection technology for gas pipeline leakage, *Process Safety and Environmental Protection* 91 (4) (2013) 253 – 261.
- [5] Y. An, X. Wang, B. Yue, S. Jin, L. Wu, Z. Qu, A novel method for natural gas pipeline safety online monitoring based on acoustic pulse

1
2
3
4
5
6
7
8
9 compression, Process Safety and Environmental Protection 130 (2019)
10
11 174 – 181.
12
13

- 14 [6] J. Shi, Y. Chang, C. Xu, F. Khan, G. Chen, C. Li, Real-time
15 leak detection using an infrared camera and faster r-cnn technique,
16 Computers & Chemical Engineering 135 (2020) 106780.
17
18 [7] W. Lu, W. Liang, L. Zhang, W. Liu, A novel noise reduction method
19 applied in negative pressure wave for pipeline leakage localization,
20 Process Safety and Environmental Protection 104 (2016) 142 – 149.
21
22 [8] S. Boll, Suppression of acoustic noise in speech using spectral
23 subtraction, IEEE Transactions on acoustics, speech, and signal
24 processing 27 (2) (1979) 113–120.
25
26 [9] I. Cohen, Noise spectrum estimation in adverse environments: improved
27 minima controlled recursive averaging, IEEE Transactions on Speech
28 and Audio Processing 11 (5) (2003) 466–475.
29
30 [10] J. Chen, J. Benesty, Y. Huang, S. Doclo, New insights into the
31 noise reduction wiener filter, IEEE Transactions on audio, speech, and
32 language processing 14 (4) (2006) 1218–1234.
33
34 [11] P. Yu, J. Cao, V. Jegatheesan, L. Shu, Activated sludge process faults
35 diagnosis based on an improved particle filter algorithm, Process Safety
36 and Environmental Protection 127 (2019) 66 – 72.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- [12] Q. Yuan, L. Zhang, H. Shen, Hyperspectral image denoising employing a spectral-spatial adaptive total variation model, *IEEE Transactions on Geoscience and Remote Sensing* 50 (10) (2012) 3660–3677.
- [13] T. Xu, S. Chen, S. Guo, X. Huang, J. Li, Z. Zeng, A small leakage detection approach for oil pipeline using an inner spherical ball, *Process Safety and Environmental Protection* 124 (2019) 279 – 289.
- [14] L. Zheng, Y. Yang, Q. Tian, Sift meets cnn: A decade survey of instance retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (5) (2018) 1224–1244.
- [15] B. Tan, Q. Peng, X. Yao, C. Hu, Z. Xu, Z. Zhang, Character recognition based on corner detection and convolution neural network (2017) 503–507.
- [16] F. Li, W. Wang, J. Xu, J. Yi, Q. Wang, Comparative study on vulnerability assessment for urban buried gas pipeline network based on svm and ann methods, *Process Safety and Environmental Protection* 122 (2019) 23 – 32.
- [17] P. Arpaia, U. Cesaro, M. Chadli, H. Coppier, L. De Vito, A. Esposito, F. Gargiulo, M. Pezzetti, Fault detection on fluid machinery using hidden markov models, *Measurement* 151 (2020) 107126.
- [18] X. Li, G. Chen, H. Zhu, Quantitative risk analysis on leakage failure of

1
2
3
4
5
6
7
8
9 submarine oil and gas pipelines using bayesian network, Process Safety
10 and Environmental Protection 103 (2016) 163 – 173.
11
12

- 13
14 [19] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553)
15 (2015) 436–444.
16
17
18 [20] Z. Li, H. Zhang, D. Tan, X. Chen, H. Lei, A novel acoustic emission
19 detection module for leakage recognition in a gas pipeline valve, Process
20 Safety and Environmental Protection 105 (2017) 32–40.
21
22
23
24 [21] R. P. D. Cruz, F. V. D. Silva, A. M. F. Fileti, Machine learning and
25 acoustic method applied to leak detection and location in low-pressure
26 gas pipelines, Clean Technologies and Environmental Policy (2020) 1–12.
27
28
29
30
31 [22] J. Sun, Q. Xiao, J. Wen, F. Wang, Natural gas pipeline small leakage
32 feature extraction and recognition based on lmd envelope spectrum
33 entropy and svm, Measurement 55 (2014) 434–443.
34
35
36
37 [23] M. Zadkarami, M. Shahbazian, K. Salahshoor, Pipeline leak diagnosis
38 based on wavelet and statistical features using dempster–shafer classifier
39 fusion technique, Process safety and environmental protection 105
40 (2017) 156–163.
41
42
43
44
45
46
47
48
49
50 [24] J. Wang, L. P. Tchapmi, A. P. Ravikumar, M. McGuire, C. S. Bell,
51 D. Zimmerle, S. Savarese, A. R. Brandt, Machine vision for natural gas
52 methane emissions detection using an infrared camera, Applied Energy
53 257 (2020) 113998.
54
55
56
57
58
59
60
61
62
63
64
65

- [25] J. Bae, D. Yeo, D. Yoon, S. W. Oh, G. J. Kim, N. Kim, C. Pyo, Deep-learning-based pipe leak detection using image-based leak features, in: 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 2361–2365.
- [26] J. SUN, Y. QIAO, J. WEN, Intelligent aperture identification combining compressed data acquisition with sparse filtering-based deep learning towards natural gas pipeline leak, Structural Health Monitoring 2017 (shm) (2017).
- [27] A. Fukane, S. Sahare, Noise estimation algorithms for speech enhancement in highly non-stationary environments, International Journal of Computer Science Issues 8 (03 2011).
- [28] M. Rastegari, V. Ordonez, J. Redmon, A. Farhadi, Xnor-net: Imagenet classification using binary convolutional neural networks, in: European Conference on Computer Vision, Springer, 2016, pp. 525–542.
- [29] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu, Convolutional neural networks for speech recognition, IEEE/ACM Transactions on audio, speech, and language processing 22 (10) (2014) 1533–1545.
- [30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets,

1
2
3
4
5
6
7
8
9 atrous convolution, and fully connected crfs, IEEE transactions on
10 pattern analysis and machine intelligence 40 (4) (2017) 834–848.
11
12

- 13 [31] A. El-Sawy, E.-B. Hazem, M. Loey, Cnn for handwritten arabic digits
14 recognition based on lenet-5, in: International Conference on Advanced
15 Intelligent Systems and Informatics, Springer, 2016, pp. 566–575.
16
17 [32] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with
18 deep convolutional neural networks, in: Advances in neural information
19 processing systems, 2012, pp. 1097–1105.
20
21 [33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-
22 scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
23
24 [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov,
25 D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with
26 convolutions, in: Proceedings of the IEEE conference on computer vision
27 and pattern recognition, 2015, pp. 1–9.
28
29 [35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image
30 recognition, in: Proceedings of the IEEE conference on computer vision
31 and pattern recognition, 2016, pp. 770–778.
32
33 [36] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.
34
35 [37] J. Han, C. Moraga, The influence of the sigmoid function parameters on
36 the speed of backpropagation learning, in: International Workshop on
37 Artificial Neural Networks, Springer, 1995, pp. 195–201.

- [38] B. L. Kalman, S. C. Kwasny, Why tanh: choosing a sigmoidal function, in: International Joint Conference on Neural Networks, 1992.
- [39] A. F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375 (2018).
- [40] F. Demir, M. Turkoglu, M. Aslan, A. Sengur, A new pyramidal concatenated cnn approach for environmental sound classification, Applied Acoustics 170 (2020) 107520.
- [41] K. Zhao, H. Jiang, Z. Wang, P. Chen, B. Zhu, X. Duan, Long-term bowel sound monitoring and segmentation by wearable devices and convolutional neural networks, IEEE Transactions on Biomedical Circuits and Systems (2020).
- [42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014) 1929–1958.
- [43] Z. Jia, L. Ren, H. Li, T. Jiang, W. Wu, Pipeline leakage identification and localization based on the fiber bragg grating hoop strain measurements and particle swarm optimization and support vector machine, Structural Control and Health Monitoring 26 (2) (2019) e2290.
- [44] Z. Qu, H. Feng, Z. Zeng, J. Zhuge, S. Jin, A svm-based pipeline leakage detection and pre-warning system, Measurement 43 (4) (2010) 513–519.

Declaration of interests

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

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

