

Received 7 November 2022, accepted 17 November 2022, date of publication 24 November 2022,
date of current version 30 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3224430

RESEARCH ARTICLE

Sound-Based Improved DenseNet Conveyor Belt Longitudinal Tear Detection

DI MIAO¹, YIMIN WANG^{2,3}, AND SHIXIN LI¹

¹School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300000, China

²Tianjin Electronic Information College, Tianjin 300350, China

³School of Mechanical Engineering, Tiangong University, Tianjin 300387, China

Corresponding author: Di Miao (09111042@bjtu.edu.cn)

This work was supported by the Science and Technology Think Tank Youth Talent Plan under Grant 20220615ZZ07110010.

ABSTRACT At present, conveyor belt tearing is the most serious fault of belt conveyor, causing the greatest loss. This paper aims to overcome the issues of poor accuracy, low precision, and poor real-time performance in the longitudinal tear detection of the conveyor belt of the belt conveyor. Specifically, this paper presents a method of longitudinal tear detection of conveyor belt based on the sound signal. According to the recognition of the tearing sound signal, the longitudinal tearing of conveyor belt can be detected. A dynamic MFCC feature extraction method is proposed to extract the sound signal feature. An improved DenseNet neural network model is designed, which is used to classify the longitudinal sound of the belt conveyor to realize the longitudinal tearing detection of the conveyor belt. The experimental results demonstrate that the method in this paper achieves the sound detection of the longitudinal tear of the conveyor belt, and the average accuracy of the longitudinal tear detection of the conveyor belt of the belt conveyor reaches 95.42%, which satisfies the requirements of the longitudinal tear detection of the conveyor belt of the belt conveyor. Applying this method to the longitudinal tearing detection of conveyor belt can solve the shortcomings of existing methods and realize the detection of the longitudinal tearing fault of conveyor belt.


INDEX TERMS Belt conveyor, longitudinal tear detection, sound, dynamic MFCC, improved DenseNet.

I. INTRODUCTION

Conveyor belts frequently fail during use, and their failures can be divided into two categories. One is the internal failure of the conveyor belt, mainly the failure of the wire rope core conveyor belt, involving the non-standard lap joint, stretching, corrosion, fracture, and scratches of the wire rope core. The wire rope core joint is the lowest tensile strength and the weakest link of the whole steel wire rope core conveyor belt, and the belt breakage accident caused by its failure often appears [1]. The other type is the surface failure of the conveyor belt, which mainly consists of deviation, surface damage, and longitudinal tearing. Misalignment and longitudinal tearing failures frequently occur and are the key failures to prevent [2], [3]. There are many reasons for conveyor belt failure. The major reasons are that the long-term use in harsh environments causes the conveyor belt to age, the load increases, the joints are poorly lapped and vulcanized, they

are scratched by foreign objects or obstacles such as scrap steel or coal gangue, the installation and adjustment of the belt conveyor are improper, and the incorrect blanking position at the transfer point leads to the uneven load. Due to the failure to detect and handle the failure in time, major safety accidents such as longitudinal tearing of the conveyor belt of the belt conveyor occur from time to time, resulting in the loss of coal transportation and production stoppage and thus severe economic losses [1].

At present, the main detection methods for longitudinal tearing of conveyor belts contain mechanical detection method [4], [5], ultrasonic detection method [6], and electromagnetic detection method [7], [8]. The mechanical detection method requires that foreign objects must be dropped to touch the detection device. It is prone to missed detection or a false alarm, with low detection efficiency. The ultrasonic detection method has high requirements on the detection environment, is sensitive to the interference caused by the vibration of the conveyor belt, and leads to severe false alarms in the system. Regarding the electromagnetic detection method, the

The associate editor coordinating the review of this manuscript and approving it for publication was Ines Domingues .

induction coil is first embedded into the conveyor belt; when a tear occurs, the tear is detected by detecting whether the induction coil loop is disconnected. However, it exhibits disadvantages of easy damage, high cost, and poor versatility.

With the development of machine vision technology, computer vision technology is widely applied in the fields such as security [9], [10], agricultural production [11], [12], and intelligent transportation [13]. Therefore, it has also become the mainstream direction of longitudinal tear detection [14]. Qiao et al. [1] proposed a comprehensive binocular vision detection method for longitudinal tearing of conveyor belts based on the fusion of infrared and visible light, while this method is relatively complicated. Yang et al. [15] adopted a linear CCD camera to obtain tearing images for detection. The method has complex structure and poor real-time performance. Li et al. [8] employed the linear laser emitted by the laser light source to irradiate the surface of the conveyor belt and acquired the image of the conveyor belt with a CMOS camera. This also has the disadvantage of complex structure. Yu et al. [16] obtained dual-band infrared images of conveyor belts using mid-infrared and long-infrared dual-band infrared detection sensors (DBID) to detect conveyor belt tearing. However, this method is greatly affected by the site environment. The above longitudinal tear detection method based on computer vision generally has the problems of complex detection systems and high installation costs. At the same time, due to the harsh working environment of the conveyor belt, there is a lot of dust and coal ash. As a result, the camera and the light source cannot work properly due to contamination, and the failure rate and maintenance cost are high.

With the development of sound detection technology, multi-channel sound acquisition equipment has been widely applied in various industries owing to its stable and mature technology. Sound detection technology has been employed to detect foreign objects and faults. The current mainstream sound detection methods are based on feature parameter extraction and classifier combination. Therefore, the selection of feature parameter extraction methods and classifiers directly influences the recognition rate of foreign body sounds. Cemal Hanilca [17] effectively identified the source mobile device using Mel frequency and linear frequency cepstral coefficients and support vector machine (SVM) detection method, and the identification accuracy rate reached 98%. Hu et al. [18] utilized reptile calls and vocalizations to identify animal species. Lim et al. [19] used the FBANK method for feature extraction, based on the CNN model for sound classification, and achieved high results in urban sound classification. Jung et al. [20] used the MFCC method for feature extraction and the CNN model for cattle sound classification to realize a real-time livestock monitoring system. Duan et al. [21] used the PW-MFCC method for feature extraction, and designed an automatic recognition system for diseased sheep based on LSTM networks for sheep foraging sound classification.

The belt conveyor emits a sound when working, and the working sound has unique changes when it is not working normally. Therefore, the change in the sound at the work site

can well reflect whether there is an abnormality in the event in the environment. The method of detecting faults by sound has been widely applied in the industrial field. For example, the sound inside the machine (including the motor) was adopted to detect whether there is an abnormal failure in the mechanical equipment. In the actual work site, the abnormal sounds of the longitudinal tear of the belt conveyor belt are generally realized by nearby workers. Therefore, the idea of detecting the longitudinal tear of the conveyor belt by sound is feasible. Hou Chengcheng et al. [22] designed a combined audio-visual detection method (sounds and images) to detect conveyor belt tearing. However, there are also shortcomings in using machine vision technology to detect longitudinal tearing of conveyor belts. Moreover, the detection system is too complicated, and the accuracy of sound detection is poor. In summary, the study of a high recognition rate and relatively simple structure method for conveyor belt longitudinal tear identification has become the focus of research. The main contributions of this paper are as follows:

1) In this paper, a longitudinal tearing detection method for conveyor belt based on sound is proposed. This method detects the longitudinal tearing of conveyor belt by identifying the tearing sound of conveyor belt during operation. It has the advantages of a simple structure and low cost.

2) This paper uses the dynamic MFCC method for sound sample feature extraction. Sound signals are classified using an improved DenseNet network. This combination improves the detection accuracy of the conveyor belt tearing sound signal.

II. A METHOD BASED ON THE IMPROVED DENSENET NETWORK

The conveyor belt longitudinal tear detection method of improved DenseNet sound is mainly divided into four parts: sound signal acquisition, sound signal feature extraction, detection model training, and tear state judgment. The structure diagram is shown in Figure 1.

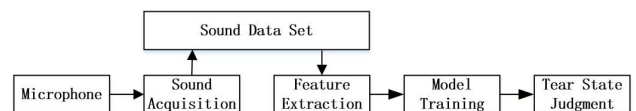


FIGURE 1. The structure diagram.

Firstly, the microphone array is used to collect the sound signal of the conveyor belt working site. Then the dynamic MFCC method is used to extract the feature of the sound signal. Finally, the feature map is input into the trained improved DenseNet network model to determine the tearing state of the conveyor belt.

III. ACOUSTIC SAMPLE COLLECTION

In this paper, the sound collection is conducted in the working site of the belt conveyor. The sampling frequency of sound collection is set to 48kHz, and the length of each sound collection is about 2s. Considering that the working sound of the belt conveyor is different at different belt speeds, the sound is collected at the belt speed of 3m/s

and 5m/s. The collected sound data sets are normal operation sound signal of belt conveyor (no-load operation sound, load operation sound, running mixed human sound), conveyor belt tearing sound signal (no-load operation tearing sound, load operation tearing sound). Sampling frequency selection 48k, normal operation sound acquisition 840 segments, tear sound 360 segments, a total of 1200 segments of sound data. Then, the data set is expanded by the method of time extension, pitch change, pitch offset and background noise, and 4800 audio segments are obtained as the data set. The data is divided into training set and test set, training set 4400, and test set 400. With a sound file for each sound type as an example, the amplitude waveform diagram of the sound file is drawn in Figure 2. The figure reveals that the loud noise of the belt conveyor covers up the human voice and the sound of the conveyor belt tearing. Therefore, it cannot be distinguished by the amplitude waveform of each sound.

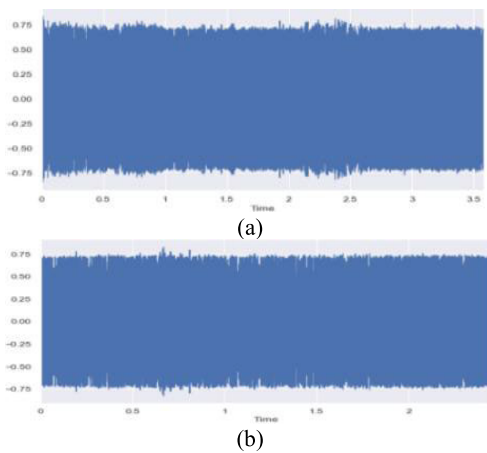


FIGURE 2. Mel frequency cepstral coefficient sound sample feature extraction flow chart.

IV. SOUND SAMPLE FEATURE EXTRACTION

Although the amplitude waveform diagram of the sound signal can effectively express the characteristics of the sound signal, the time domain analysis for the identification and processing of the sound signal has the same amount of information for all the information. Particularly, the working environment of the belt conveyor is complex, and the extracted sound signal is formed by the superposition of various signals. The conveyor belt tearing sound stressed in our study is mixed in the complex sound signal. Therefore, we need to directly pay attention to the signal frequency of the tearing sound of the conveyor belt, convert the audio signal to the frequency domain, and intercept the torn sound signal of the conveyor belt by filtering out other frequency information from the frequency domain, so as to effectively extract the signal of the tearing sound of the conveyor belt.

A. MEL FREQUENCY

The amplitude waveform diagram can intuitively exhibit the change in the collected sound signal. Nevertheless, experiments on human auditory perception have demonstrated that human hearing is focused on a specific frequency range rather

than the entire spectrum of sound signals. Mel frequency analysis is designed for human auditory perception and simulates the human ear to design a series of filter banks to identify sounds of interest. It first maps the linear spectrogram of the original sound to the Mel nonlinear spectrum based on human auditory perception, and the formula for converting the original frequency to the Mel frequency, as in

$$f_{mel} = 2595 * \lg(1 + f/700) \quad (1)$$

B. FEATURE EXTRACTION OF SOUND SAMPLES BASED ON MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

The feature extraction of sound samples based on Mel Frequency Cepstral Coefficients (MFCC) can target the collected sound signals to extract interesting features. This is an essential process for speech recognition, as illustrated in Figure 3.

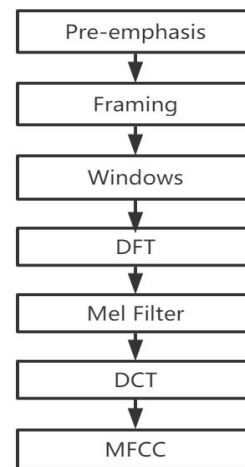


FIGURE 3. Amplitude waveform diagram:(a)Normal operation sound;(b) Tearing sound.

1) PRE-EMPHASIS

Pre-emphasis is the use of a high-pass filter to boost the high-frequency part of the signal, making the spectrum flatter. This ensures that the spectrum can be obtained with the same signal-to-noise ratio in the entire frequency band from low frequency to high frequency. The high-pass filter is expressed in formula (2), where the value of u is a constant of 0.9-1.0.

$$H(z) = 1 - uz^{-1} \quad (2)$$

2) FRAMING

Framing is the division of a longer sound signal into smaller segments. The time covered by one frame is usually 20-30ms. An overlapping area (usually half the frame length) between two adjacent frames should be set to avoid excessive changes between two adjacent frames. In this paper, every 1200 sampling signals is regarded as a frame, and the data duration of each frame is 25ms.

3) WINDOWING

Windowing is to avoid a signal discontinuity at both ends of each frame of signal. Commonly used window functions

are square window, Hamming window, and Hanning window. Since the Hamming window is suitable for processing signals with complex spectral representations and multiple spectral components, Hamming is selected for data windowing in this paper.

Each frame signal is $S(i), i = 0, 1, \dots, I-1$ where I denotes the frame size. The data after windowing is $S'(i) : S'(i) = S(i) \times W(i)$. Where $W(i)$ indicates Hamming window: $W(i) = (1 - a) - a \times \cos[\frac{2\pi i}{I-1}], 0 \leq i \leq I - 1$

4) FAST FOURIER TRANSFORM (FFT)

Since it is difficult to analyze the signal characteristics in the time domain, it is necessary to convert the signal to the frequency domain for analysis using a fast Fourier transform (FFT). In this paper, FFT is performed on the frame-by-frame windowed data, and the FFT of the i -th frame signal is expressed in formula (3). The frequency-domain signal after each frame of DFT is superimposed in time to obtain the spectrogram of the audio data. The spectrogram of the collected sound sample data obtained after processing the signal data in Figure 1 is exhibited in Figure 4.

$$X_i(k) = \sum_{i=0}^{I-1} x(i)e^{-j2\pi k/I}, \quad 0 \leq k \leq I \quad (3)$$

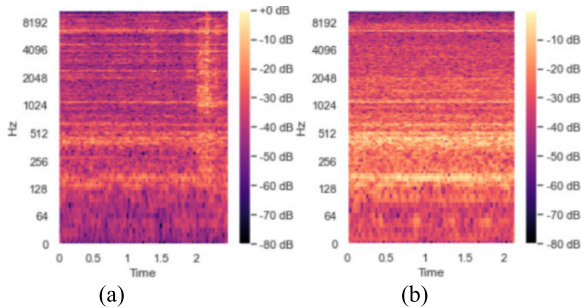


FIGURE 4. The spectrogram of the collected sound data:(a)Normal operation sound;(b) Tearing sound.

5) MEL FILTER

Sensors have different sensitivities to different frequencies of sound. Thus, a filter bank should be designed to process the spectrum signal for making the collected sound signal not affected by the pitch during the recognition and processing. This paper defines a filter bank with M filters as a Mel filter bank, where the response function of each filter bank is:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) < k < f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad (4)$$

In the feature extraction of the sound signal, it is necessary to distinguish the features of the sound by the energy of the signal. Therefore, this paper adopts the method of calculating

the short-term energy of sound. The energy of the i -th frame signal is generally $E_i(k) : E_i(k) = |X_i(k)|^2$

The energy of each frame signal passes through the Mel filter to obtain the final frame energy, expressed as:

$$S_i(m) = \ln\left(\sum_{k=0}^{N-1} E_i(k)H_m(k)\right), \quad 0 \leq m < M \quad (5)$$

6) DISCRETE COSINE TRANSFORM (DCT)

DCT is mainly used for compressing data or compressing images with good decorrelation. The discrete cosine transform is performed on the i -th frame signal to obtain the cepstral parameters, and the formula is:

$$C_i(n) = \sum_{m=0}^{M-1} S_i(m) \cos\left(\frac{\pi n(m-0.5)}{M}\right), \quad n = 0, 1, \dots, L \quad (6)$$

The MFCC image of the sound sample is obtained by subjecting the example sound file in Figure 1 to the feature extraction process of the Mel-frequency cepstral coefficient sound sample, as presented in Figure 5.

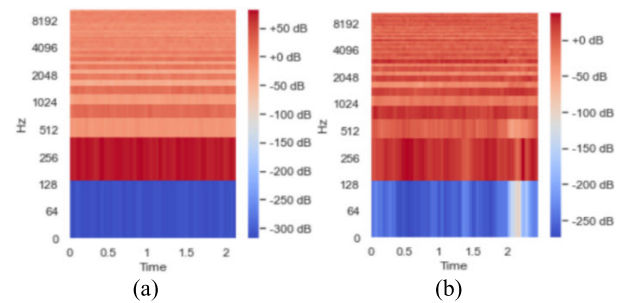


FIGURE 5. The MFCC image of the sound sample:(a)Normal operation sound;(b) Tearing sound.

C. FEATURE EXTRACTION OF DYNAMIC MFCC SOUND SAMPLES

The dynamic MFCC feature extraction adopts the scramble frequency parameter combined with dynamic and static features to perform sound recognition to improve the performance of recognition. Dynamic sound parameters are obtained by differencing static feature parameters. The expression of the difference calculation is:

$$d_t = \begin{cases} C_{t+1} - C_t, & t < K \\ \frac{\sum_{k=1}^K k(C_{t+k} - C_{t-k})}{\sqrt{2 \sum_{k=1}^K k^2}}, & \text{other} \\ C_t - C_{t-1}, & t \geq Q - K \end{cases} \quad (7)$$

where, d_t represents the t -th first-order difference, C_t denotes the t -th cepstral coefficient, Q refers to the order of the cepstral coefficient, K designates the time difference of the first derivative, generally from 1 to 2. It is taken as 2 in this paper. The data calculated by MFCC is recorded as MFCC; the result obtained by the first-order difference calculation of formula (10) is recorded as DMFCC; the result of the

second-order difference obtained by performing a differential calculation on DMFCC is recorded as DDMFCC.

Then, the whole composition of dynamic MFCC is: $\nabla\text{MFCC} = \text{MFCC} + \text{DMFCC} + \text{DDMFCC}$

In the calculation, MFCC takes the first 40-dimensional features with significant features. Both DMFCC and DDMFCC are 10-dimensional features. The final dynamic MFCC is a 60-dimensional feature parameter, as illustrated in Figure 6.

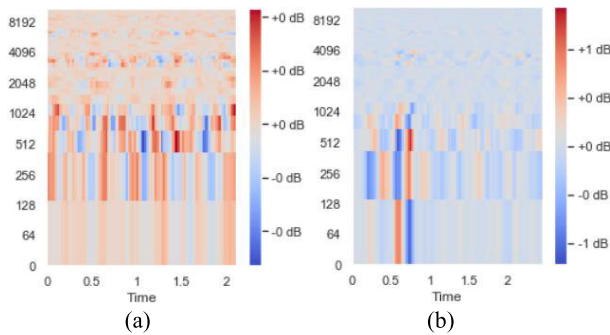


FIGURE 6. Dynamic MFCC feature map:(a)Normal operation sound; (b) Tearing sound.

V. IMPROVEMENT OF DENSENET NETWORK MODEL DESIGN

Generally, the network structure such as Highway Networks, Residual Networks (ResNets), and GoogLeNet is deepened to better improve the classification results. The gradient disappearance problem should be first addressed to deepen the network structure. However, the emergence of DenseNet jumps out of the stereotyped thinking of deepening the number of network layers and widening the network structure to improve network performance. From the perspective of features, feature reuse and bypass settings not only significantly reduce the number of network parameters but also alleviate gradient disappearance to a certain extent. Its structure is presented in Figure 7. The DenseNet network is mainly composed of (dense block) dense blocks and (transition layer) transition layers. In the network structure of DenseNet, the input of each layer comes from the output of all previous layers, and each layer in the Dense Block is densely connected to all subsequent layers [23], [24]. The dense block is adopted to define the connection relationship between the input and output, and the transition layer is employed to control the number of channels. Dimensionality reduction is performed by adding 1*1 convolution to the transition layer. This effectively solves the problem of the large output dimension of Dense Block. For an i -layer DenseNet network structure (Figure 6), the input and output of the i -th layer are denoted as x_i and y_i respectively, then, $y_i = H(x_i) = H([x_0, x_1, \dots, x_{i-1}])$.

The DenseNet network designed in this paper adopts 3 dense blocks. Each dense block consists of 6 groups (BN-ReLU-Conv(1*1)+BN-ReLU-Conv(3*3)). The transition layers are taken to adjust the size of the output features. Each Transition Layer includes a Conv (1*1) and a 2*2 mean pooling. Among them, 1 × 1 convolution realizes feature

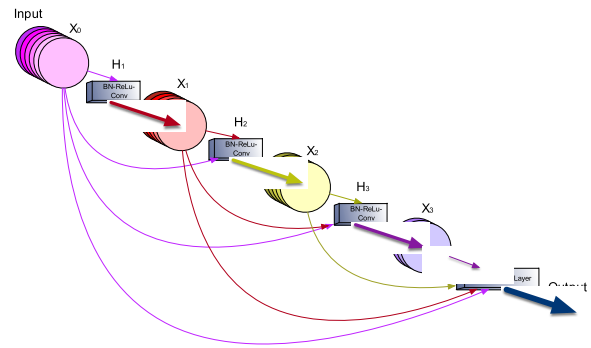


FIGURE 7. Diagram of dense block.

TABLE 1. The structure of DenseNet.

Layer type	Output size	DenseNet
Convolution layer	39,173	Conv(7*7)
Maximum pool layer	19,86	Pool(3*3)
Dense Block1	19,86	$\left[\begin{array}{l} \text{BN} + \text{Re lu} + \text{Conv}(1 * 1) \\ \text{BN} + \text{Re lu} + \text{Conv}(3 * 3) \end{array} \right] * 6$
Transition Layer	18,85	Conv(1*1)
	9,44	Average Pool(2*2)
Dense Block2	9,44	$\left[\begin{array}{l} \text{BN} + \text{Re lu} + \text{Conv}(1 * 1) \\ \text{BN} + \text{Re lu} + \text{Conv}(3 * 3) \end{array} \right] * 6$
Transition Layer	9,44	Conv(1*1)
	5,22	Average Pool(2*2)
Dense Block3	5,22	$\left[\begin{array}{l} \text{BN} + \text{Re lu} + \text{Conv}(1 * 1) \\ \text{BN} + \text{Re lu} + \text{Conv}(3 * 3) \end{array} \right] * 6$
Transition Layer	5,22	Conv(1*1)
	2,11	Average Pool(2*2)
Pool Layer	1,5	Average Pool(2*2)
Output		softmax

compression, and 2 × 2 mean pooling realizes feature dimension reduction. The structure of the network is presented in the following TABLE 1.

The main feature of this structure of DenseNet is that the input of the current layer comes from the output of all previous layers. This study improves the original DenseNet structure and changes the input of each layer, allowing it to be only related to the output of the first two layers. This can not only ensure enough feature information but also reduce the depth of the model, making it more suitable for the application of sound classification. Its structure diagram is illustrated in Figure 8. The input and output of the i -th layer of the DenseNet structure are denoted as x_i and y_i , respectively, then, $y_i = H(x_i) = H([x_{i-2}, x_{i-1}])$.

Compared with the original network structure, the improved DenseNet structure reduces the reuse of feature information to weaken the over-fitting phenomenon and further lowers the number of network parameters and saves memory, making it more suitable for application scenarios of sound classification.

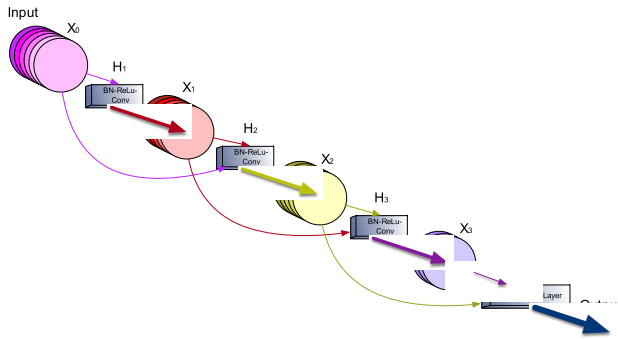


FIGURE 8. Improved dense block diagram in DenseNet structure.



FIGURE 9. Belt conveyor experimental platform.

VI. EXPERIMENT AND RESULT ANALYSIS

A. EXPERIMENT ENVIRONMENT

The team established a belt conveyor experimental platform, and the photos of the platform are shown in Figure 9. The experimental environment is the Windows10 operating system with I7-6800K@3.4 GHz CPU, GTX1080 graphics card, 16 G memory. Based on the jupyter platform, an open-source deep-learning Tensorflow2.0 framework is established. Program writing is performed using Python 3.7.

The experimental process is divided into data generation, dynamic MFCC feature extraction, generation of training data sets and test data sets, training with the improved DenseNet network, and model testing. Its flow chart is displayed in Figure 10.

- 1) Preprocess the sound signal;
- 2) Perform dynamic MFCC feature extraction on the new data obtained to obtain a data set;
- 3) Use Improved DenseNet neural network for model training;
- 4) Employ the trained model for testing.

B. ANALYSIS OF RESULTS

The experiment uses the above collection and expansion data set for experimental verification. The data structure is exhibited in Figure 11.

- 1)Comparative experiment of feature extraction of MFCC and dynamic MFCC sound samples

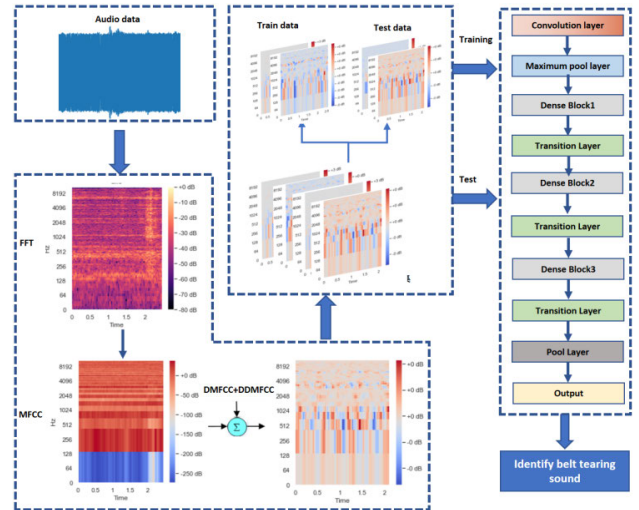


FIGURE 10. Flow chart of experiment.

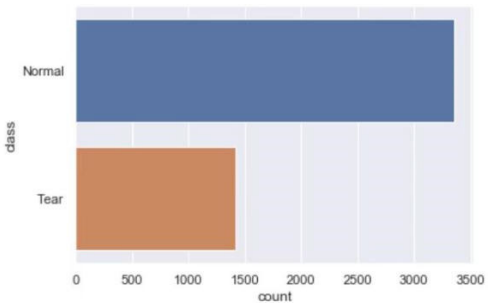


FIGURE 11. Experimental data structure.

Based on the jupyter platform, the sklearn library is adopted to establish an SVM classifier, and the feature extraction results of MFCC and dynamic MFCC sound samples are verified. The identified confusion matrices are illustrated in Figure 12.

It can be observed from Figure 12 that the recognition rate of MFCC feature extraction using SVM classifier is 90.8%, while the recognition rate of dynamic MFCC feature extraction using SVM classifier is 93.22%. After averaging this verification, the recognition rate of MFCC feature extraction is stabilized at 90%. However, the recognition rate using dynamic MFCC feature extraction can reach 93%. Therefore, the recognition rate of sound sample feature extraction using dynamic MFCC is better than the result of MFCC feature extraction.

2)Improved DenseNet network model experiment

The improved DenseNet network designed above is used for training, with 128 data batches and 2000 iteration cycles. The obtained recognition rate and loss rate are provided in Figure 13. As can be seen from the figure, the recognition rate tends to be flat after training 600 times.

A comparative experiment is designed to verify the recognition effect of the improved DenseNet model on the longitudinal tearing sound of the belt conveyor. Design comparison test based on jupyter platform. The random forest classifier is created based on the ‘RandomForestClassifier’

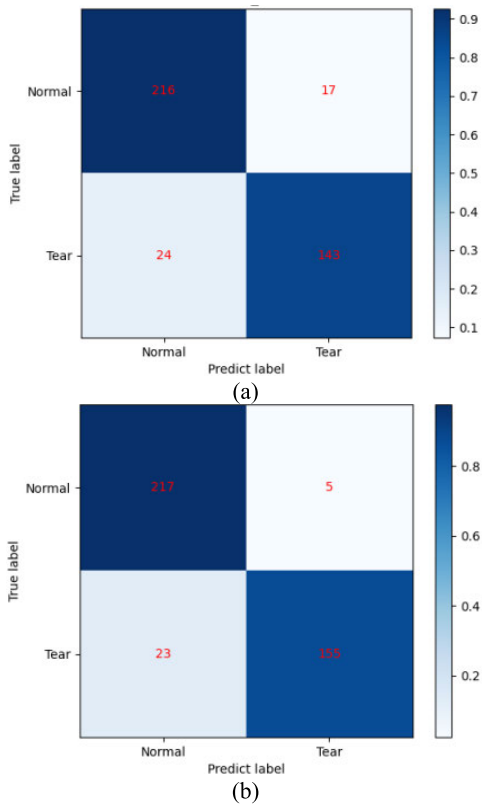


FIGURE 12. Confusion matrix:(a)MFCC feature recognition confusion matrix;(b)Dynamic MFCC Feature Extraction Recognition Confusion Matrix.

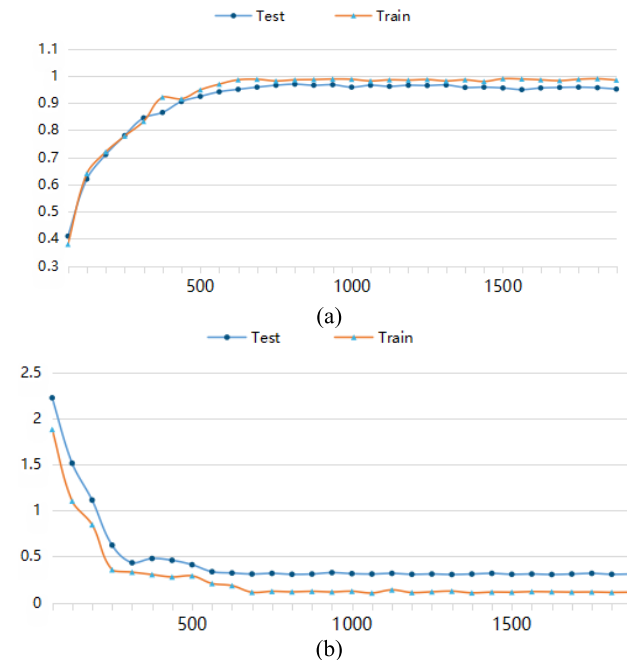


FIGURE 13. Training results of improved DenseNet network:(a) accuracy; (b) loss rate.

function built in the sklearn library. The main parameters are, number of estimators:100, criterion:gini, max_depth:4, max_features:auto, the other parameters use default values.

TABLE 2. Comparison of recognition rate of random forest, SVM classifier, LSTM model, original DenseNet model and proposed model.

Model	Accuracy
Random forest	86.06%
SVM	93.22%
LSTM	92.41%
This Method	95.42%

TABLE 3. Comparison results of this method with existing sound processing methods.

Model	Feature	Accuracy
CNN	FBANK	92.06%
CNN	MFCC	94.45%
LSTM	PW-MFCC	92.65%
This Method	Dynamic MFCC	95.42%

The main parameters are, penalty parameter of error term (C) :10, kernel type:rbf, gamma:0.001, the other parameters use default values.Design Long Short Term Memory (LSTM) to create a sound classification model with, input:40, the cell number of LSTM:300, learning rate:0.05, batch_size:50, dropout:0.5. The experimental results are listed in TABLE 2.

As revealed from TABLE 2, the detection method for longitudinal tearing of conveyor belt based on the sound of improved DenseNet designed in this paper can achieve a recognition rate of 95.42% in the detection of longitudinal tearing of conveyor belt. This is 9.36%, 2.2%, and 3.01% higher than that obtained using random forest, SVM classifier, and LSTM model, respectively. Therefore, the method in this paper is better for the recognition of longitudinal tearing sound.

3)Comparison between the proposed method and the existing sound processing methods

After using the dynamic MFCC feature extraction method in this paper for feature extraction, the improved DenseNet model is used to compare the tearing sound recognition method with the literature [18], [19], [20] sound classification and recognition method. Training and testing are performed on the data set of this paper. The results are shown in TABLE 3.

The average detection time of each audio segment is 29.76ms when the method proposed in this paper is used for testing. It suggests that when longitudinal tearing occurs in the belt conveyor, the longitudinal tearing fault can be accurately detected according to the sound in a very short time, and the loss caused by the longitudinal tearing of the belt conveyor can be reduced. Therefore, the method for longitudinal tear detection of conveyor belts based on the sound of improved DenseNet proposed in this paper has good real-time performance, satisfying the requirements of production.

VII. CONCLUSION

Given the issues of poor accuracy and low precision in the longitudinal tear detection of conveyor belts in belt conveyors, a method for longitudinal tear detection of conveyor belts is proposed in this paper based on the sound of improved DenseNet. Besides, the dynamic MFCC feature extraction method is adopted to identify the sound signal features. The dataset is divided into the training set and test set.

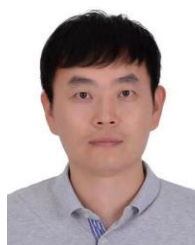
An improved DenseNet network model is designed to train and test the training set and test set. Finally, the processing and classification of the sound signal of the belt conveyor are realized, as well as the detection of the longitudinal tear of the conveyor belt. The experimental results demonstrate that the detection method can accurately detect the longitudinal tear of the conveyor belt of the belt conveyor by using the characteristics of the longitudinal tearing sound of the conveyor belt. The recognition rate can reach 95.42%. Moreover, its average time is 29.67ms. Compared with the traditional longitudinal tear detection method, this detection method has the characteristics of a simple structure and high recognition efficiency. It can solve the problems of existing methods. Improving the effectiveness of belt conveyor belt longitudinal tear detection system, in mining, transportation, and other fields have broad application prospects.

REFERENCES

- [1] T. Qiao, L. Chen, Y. Pang, G. Yan, and C. Miao, "Integrative binocular vision detection method based on infrared and visible light fusion for conveyor belts longitudinal tear," *Measurement*, vol. 110, pp. 192–201, Nov. 2017.
- [2] J. Li and C. Miao, "The conveyor belt longitudinal tear on-line detection based on improved SSR algorithm," *Optik*, vol. 127, no. 19, pp. 8002–8010, Oct. 2016.
- [3] J. Wang, C. Miao, W. Wei, and X. Lu, "Research of X-ray nondestructive detector for high-speed running conveyor belt with steel wire ropes," *Proc. SPIE*, vol. 6833, Nov. 2008, Art. no. 68331X.
- [4] X. Peng, "A novel image-based method for conveyor belt rip detection," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2013, pp. 1–4.
- [5] C. Wang, "Design and research of ultrasonic nondestructive testing system for conveyor belt," *Mech. Manag. Develop.*, vol. 177, pp. 98–100, Mar. 2018.
- [6] G. Davies, "Modern concepts in belt rip detection for steel-cord reinforced conveyor belts," in *Proc. Conf. Int. Mater. Handling Conf. (BELTCON)*, vol. 4, 1987, pp. 1–9.
- [7] Y. Pang, "A novel embedded conductive detection system for intelligent conveyor belt monitoring," in *Proc. IEEE Int. Conf. Service Operations Logistics, Informat.*, Nov. 2006, pp. 803–808.
- [8] L. Xianguo, S. Lifang, M. Zixu, Z. Can, and J. Hangqi, "Laser-based on-line machine vision detection for longitudinal rip of conveyor belt," *Optik*, vol. 168, pp. 360–369, Sep. 2018.
- [9] C. Costa, F. Antonucci, F. Pallottino, J. Aguzzi, D.-W. Sun, and P. Menesatti, "Shape analysis of agricultural products: A review of recent research advances and potential application to computer vision," *Food Bioprocess Technol.*, vol. 4, no. 5, pp. 673–692, Jul. 2011.
- [10] J. P. Dandois, M. Olano, and E. C. Ellis, "Optimal altitude, overlap, and weather conditions for computer vision UAV estimates of forest structure," *Remote Sens.*, vol. 7, no. 10, pp. 13895–13920, 2015.
- [11] Y. Su, Z. Guo, and Q. Zhang, "An improved intelligent transportation algorithm based on image processing," in *Proc. 8th Int. Conf. Intell. Syst., Model. Simul. (ISMS)*, May 2018, doi: 10.1109/ISMS.2018.00014.
- [12] A. Vu, A. Ramanandan, A. Chen, J. A. Farrell, and M. Barth, "Real-time computer vision/DGPS-aided inertial navigation system for lane-level vehicle navigation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 899–913, Jun. 2012.
- [13] R. Shanmugamani, M. Sadique, B. Ramamoorthy, R. Shanmugamani, and M. Sadique, "Detection and classification of surface defects of gun barrels using computer vision and machine learning," *Measurement*, vol. 60, pp. 222–230, Jan. 2014.
- [14] D. Xiao, X. Li, and K. He, "Power balance of starting process for pipe belt conveyor based on master-slave control," *IEEE Access*, vol. 6, pp. 16924–16931, 2018.
- [15] Y. Yang, C. Miao, X. Li, and X. Mei, "On-line conveyor belts inspection based on machine vision," *Optik*, vol. 125, no. 19, pp. 5803–5807, Oct. 2014.
- [16] B. Yu, T. Qiao, H. Zhang, and G. Yan, "Dual band infrared detection method based on mid-infrared and long infrared vision for conveyor belts longitudinal tear," *Measurement*, vol. 120, pp. 140–149, May 2018, doi: 10.1016/j.measurement.2018.02.029.
- [17] C. Haniilçi and T. Kinnunen, "Source cell-phone recognition from recorded speech using non-speech segments," *Digit. Signal Process.*, vol. 35, pp. 75–85, Dec. 2014.
- [18] J. J. Noda and C. M. Travieso-González, "Acoustic classification of singing insects based on MFCC/LFCC fusion," *Appl. Sci.*, vol. 9, no. 19, p. 4097, 2019, doi: 10.3390/app9194097.2019.
- [19] M. Lim, D. Lee, H. Park, Y. Kang, and J. Oh, "Convolutional neural network based audio event classification," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 6, pp. 2748–2759, 2018.
- [20] D.-H. Jung, N. Y. Kim, S. H. Moon, C. Jhin, H.-J. Kim, J.-S. Yang, H. S. Kim, T. S. Lee, J. Y. Lee, and S. H. Park, "Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering," *Animals*, vol. 11, no. 2, p. 357, Feb. 2021, doi: 10.3390/ani11020357.
- [21] G. Duan, S. Zhang, M. Lu, C. Okinda, M. Shen, and T. Norton, "Short-term feeding behaviour sound classification method for sheep using LSTM networks," *Int. J. Agricult. Biol. Eng.*, vol. 14, no. 2, pp. 43–54, 2021, doi: 10.25165/ijabe.20211402.6081.
- [22] C. Hou, T. Qiao, M. Qiao, X. Xiong, Y. Yang, and H. Zhang, "Research on audio-visual detection method for conveyor belt longitudinal tear," *IEEE Access*, vol. 7, pp. 120202–120213, 2019, doi: 10.1109/access.2019.2937660.
- [23] K. Zhang, Y. Guo, X. Wang, J. Yuan, and Q. Ding, "Multiple feature reweight DenseNet for image classification," *IEEE Access*, vol. 7, pp. 9872–9880, 2019, doi: 10.1109/ACCESS.2018.2890127.
- [24] L. Huang and C.-M. Pun, "Audio replay spoof attack detection using segment-based hybrid feature and DenseNet-LSTM network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2567–2571, doi: 10.1109/ICASSP.2019.8682573.



DI MIAO received the B.S. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2009 and 2016, respectively. She is currently a Professor with the Tianjin University of Technology and Education, Tianjin, China. Her main research interests include digital image processing and networks.



YIMIN WANG received the B.S. and M.S. degrees from Tianqong University, Tianjin, China, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the School of Mechanical Engineering. His main research interests include fault detection and digital image processing.



SHIXIN LI received the M.S. and Ph.D. degrees from Tianjin University, Tianjin, China, in 2000 and 2003, respectively. He is currently a Professor with the Tianjin University of Technology and Education, Tianjin. His main research interests include digital image processing and inertial technology.

...