International Journal of Innovative Computing, Information and Control
Volume **21**, Number **2**, April **2025**

ICIC International ©2025 ISSN 1349-4198
pp. **307–322**

DEEPRIDGE: A HOP-LAYER DEEP LEARNING MODEL FOR DETECTING ENERGY RIDGES IN HOIST OPERATION NOISE SPECTROGRAMS

Yifan Meng[1,*], Heng Zhang[1] and Xiaoming Zhang[2]

[1]School of Artificial Intelligence
[2]School of Electrical and Information Engineering
Anhui University of Science & Technology
No. 168, Taifeng Street, Huainan 232000, P. R. China
{ 2023201674; 2021219 }@aust.edu.cn
[*]Corresponding author: myf@aust.edu.cn

Received June 2024; revised October 2024

Abstract. *The variation trend of energy ridge in the noise spectrum of mine hoist is an important index of its health state. Due to the interference of background noise, the energy ridges in the hoist spectrum have unclear boundaries and poor continuity, making them hard to separate and analyze. To address this challenge, we propose a new end-to-end training deep neural network that extracts and fuses multi-scale features of spectrographs using multiple encoder-attention module-decoder blocks. The multi-scale features enable the network to capture the energy ridges in the spectrum more accurately and robustly. The small-scale feature map is used for coarse localization, and the large-scale feature map is used for fine refinement. We design and train our network, named DeepRidge, on a noise spectrum dataset of mining hoist collected from a preliminary experiment. We compared DeepRidge with other state-of-the-art methods, and the test results show that DeepRidge achieves better accuracy at smaller parameter sizes, with an average precision (AP) of 0.826. We also conduct experiments to find the optimal network configuration for the energy ridge detection task in the hoist noise spectrum.*

Keywords: Ridge detection, Spectral analysis, Fault detection, Noise analysis, Deep learning

1. **Introduction.** The hoist is the main transport equipment of the mine, and its health status directly affects the production safety [1]. Therefore, it is an urgent problem for mining enterprises to develop an effective and comprehensive method to monitor the operating state of the hoist. As a large mechanical equipment, the hoist has many moving parts, such as the drive motor, gearbox, brake hydraulic pump, and drum. These components produce operating noise when they are running, which is related to the structure, material, lubrication, and other factors of the moving parts [2, 3, 4]. It is one of the inherent properties of the hoist's moving parts, containing a large amount of information about their motion conditions. Moreover, as shown in Figure 1, the frequency band changes of its inherent noise should be periodic in the spectrum image. The ridges of noise energy concentration are extracted from the spectrum texture, and finally the health status of the hoist is judged and predicted according to the trend of ridge steepness change [5, 6].

Noise spectrum can be seen as a three-dimensional structure. It can be divided into several segments to find the maximum points on the cross section. However, as shown in Figure 2, due to the huge difference in the energy distribution of different frequency bands, it is difficult to realize the filtering operation without retaining the detailed information.

DOI: 10.24507/ijicic.21.02.307

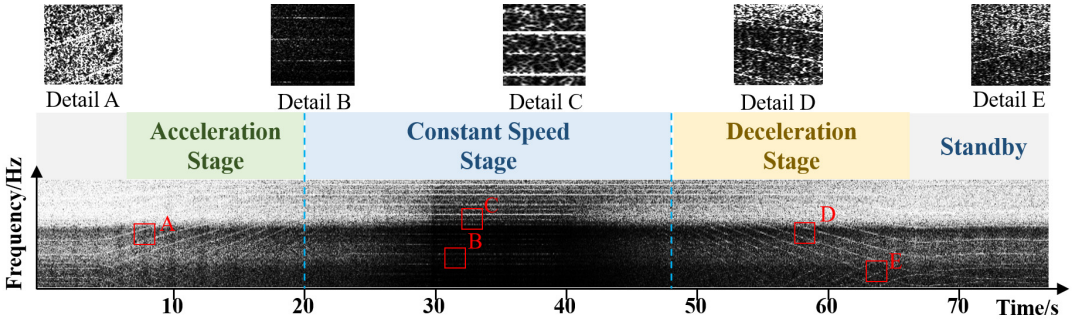
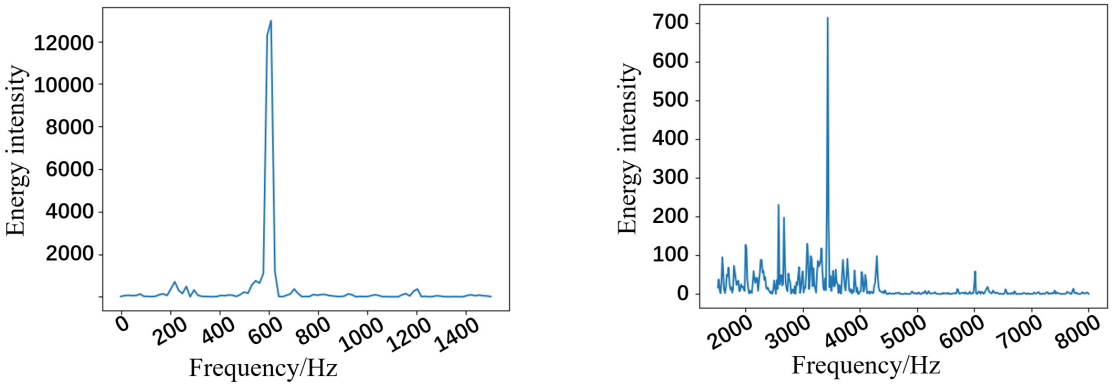


FIGURE 1. Noise spectrum diagram of hoist operation



(a) Low-frequency (b) High-frequency

FIGURE 2. Energy distribution of hoist in constant speed stage

An energy ridge is a linear or curvilinear structure, so ridge detection can be regarded as a line detection problem, which is fundamental in computer vision. The energy ridge in the spectrogram is a discontinuous edge from a local perspective, where the pixels/noise are unevenly distributed. From a global perspective, it is a continuous line structure with a certain pixel width. Therefore, it can be processed by image segmentation or edge detection methods. However, due to background noise [7], the spectral energy ridge has low visibility and clarity. For example, as shown in Figure 1, background noise particles reduce the contrast of ridge lines and break their continuity. This makes it difficult for traditional line detection models to effectively extract the ridge in the spectral image.

In this study, we elected SegNet as the foundational architecture for the DeepRidge model due to its proven efficacy and efficiency in image segmentation tasks. SegNet employs an encoder-decoder structure with skip connections that preserve spatial information which allows for high-resolution feature retention and significantly reduces network parameters. This is a critical advantage for our task, as it enables the model to operate more efficiently on resource-constrained embedded platforms.

Moreover, we integrated an attention mechanism to enhance the model's performance. This mechanism enables the network to focus on critical image information during the fusion of multi-scale feature maps while disregarding irrelevant details. This approach emulates the human visual system's behavior when processing ambiguous spectra: initially searching for potential targets in a broad scope, followed by local detail confirmation. By doing so, the DeepRidge model can more accurately capture spectral energy ridges across various scales, thereby improving its performance in detecting energy ridges in the noise spectrum of mining hoists. The main contributions of this work are as follows.

- Our major contribution is the design of a neural network architecture for spectral energy ridge detection. The new network constructs multiple search fields of different sizes through multi-level encoders and decoders.
- In the proposed network, the convolution features of each pair of encoders and decoders are fused by introducing attention module, so as to avoid information overload and improve model processing efficiency.
- A data set was constructed for performance evaluation, and the ground truth was labelled by human. A lot of experiments show that the method is effective. All dataset is open sourced to the community to facilitate research.

The rest of this paper is organized as follows. Section 2 introduces the technical research work about line detection and ridge detection. Section 3 describes the network architecture and details of DeepRidge. Three data sets of different sizes were constructed for performance evaluation, and the ground truth were labeled by human. A lot of experiments show that the method is effective. All datasets are open sourced to the community to facilitate research. Finally, Section 5 concludes the paper.

2. **Related Work.** Line segment detection (LSD) is a fundamental task in computer vision, which aims to extract meaningful linear structures, also known as semantic lines, from natural scenes [8]. Semantic lines can be used for scene understanding, image segmentation, 3D reconstruction, and other applications. Many previous methods treat this problem as a special case of object detection and adapt existing object detectors to perform semantic line detection. However, these methods ignore the intrinsic characteristics of line segments, resulting in suboptimal performance. Line segments have simpler geometric properties than complex objects, and thus can be compactly represented by a few parameters. To better exploit the features of line segments, some novel methods combine the classical Hough transform [9] technique with deep learning, and propose an end-to-end line segment detection framework [10].

Traditional gradient-based LSD methods are very fast and accurate, such as Canny edge detector [11], but lack robustness in noisy images and challenging conditions. They usually require post-processing steps, such as non-maximum suppression (NMS) [12] or clustering, to eliminate redundant and erroneous line segments. In recent years, some deep learning-based or machine learning based LSD methods are proposed to handle challenging images, which use neural networks to learn image features and line segment parameters. Dollár and Zitnick proposed a general structured learning approach applied to edge detection, which formulates the edge detection problem as predicting a segmentation mask patch for an input image [13]. Feng et al. artificially suppressed image noise to improve the accuracy of edge detection, and proposed an edge detection algorithm combined with information fusion. The algorithm is studied and improved from two aspects of fuzzy radial basis and fusion discrimination [14]. Some researchers have proposed that edge detection and contour grouping can be performed by energy minimization methods and mid-level feature learning algorithms. Martin et al. proposed to detect image boundaries by learning to use local brightness, color, and texture, which is a probabilistic boundary (Pb)-based edge detection method [15]. However, the ridge is broken due to the large number of background noise points in the elevator noise map. It is difficult to extract ridges efficiently by traditional universal methods.

A ridge is a convex landform formed by the meeting of two slopes of opposite slope and different slope gradients. The line connecting the highest point of the ridge is the intersection of the two slopes, called the ridge line. Ridges are useful geometric features because of their wide range of applications, mainly for image analysis problems such as object detection, image segmentation, and scene understanding [16]. Therefore, ridge

detection can be divided into two methods. One is to find the point with zero gradient through mathematical method, that is, the extremum points of mountain profile curve. The other is to find the point through the method approximating line detection, edge detection or crack detection in computer vision [17, 18, 19].

Sun and Mao detected rivers from remote sensing images by using edge extraction based on wavelet domain and then ridge tracking to merge water areas [20]. Many other researchers have also proposed to use wavelet transform for edge detection [21, 22, 23, 24]. However, due to the large amount of noise in the frequency spectrum of hoist noise, it is easy to entangle with the ridge line. Machine learning-based methods are also used in ridge or edge detection. In [25], a novel road crack detection framework based on random structure forest is proposed. By introducing random structure forest to generate a high-performance crack detector, arbitrary complex cracks can be identified. In [26], the SegNet model was used to classify topographic data, so as to detect different types of topography such as ridge line and hillside.

3. DeepRidge Network.

In this section, we will cover the details of the DeepRidge network, including the basic architecture of the network, the structure of the encoder and decoder, the construction of the loss function, and the optimization for the embedded platform.

3.1. Network architecture.

The DeepRidge network is built on SegNet network [27] which is designed for semantic segmentation on pixel level. SegNet is a typical encode-decode structure. Each convolution encoder has a corresponding decoder. The encoder network extracts feature maps by convolution, enlarges the perception field by pooling, and reduces the image size. The SegNet encoder uses the first 13 layers of the VGG16 convolution network [28], and removing the fully connected layer preserves higher resolution feature maps and significantly reduces the parameters of the network. The decoder uses upsamping to double the size of the image, and then uses the index information of the encoder to directly put the data back to the corresponding position, and finally fuses the features through the convolution layer.

In the encoder, the relative position of the max weight in the 2×2 filter is saved every time max-pooling is performed; in the decoder, upsampling is performed according to the saved indices: first, the input feature map is enlarged by two times, and then the data of the input feature map is placed according to the index position of the pooling layer in the encoder, and other positions are 0. Using the pooling index to perform nonlinear upsampling can retain some important boundary information. In this way, we can improve the network model's description of the boundary, connect the encoder and decoder networks with a simple structure, and reduce the parameters that need to be trained for upsampling.

When processing the noise spectrum of the hoist, there are random noise points in each frequency band of the spectrum, which make the edge of the energy ridge line blurred. When humans deal with such images, they usually search for possible targets in a large range first, and then check the details to confirm. Therefore, when constructing the DeepRidge network, a search method combining large and small fields of view is adopted, and the ridge line is searched at different scales and then fused to judge. This search method is realized by increasing the number of encoder-decoder pairs. As the number of encoder-decoders increases, the number of convolution layers also increases, and the search range of convolution also expands. At the same time, fusing convolution features at different levels can effectively improve the performance of the model [27]. Based on this idea, we designed the DeepRidge network.

In SegNet, there are five encoders, each with a pooling layer, corresponding to five scales. We designed a connection structure to fuse the coefficients and continuous feature maps of the five pairs of encoder-decoder, and obtain the final output result. As shown in Figure 3, it is the architecture diagram of DeepRidge network. The last convolution layer in each encoder network and the last convolution layer in the corresponding decoder network are connected by an attention module, and then a fused feature map is generated by a fuse module. The fused feature maps of each scale are summarized by the output module to generate the final result. We introduced attention modules in the connection structure to fuse the feature maps of each pair of encoder and decoder, selectively process the corresponding positions, extract the key information in the image while ignoring irrelevant information, as well as improve the efficiency of feature fusion.

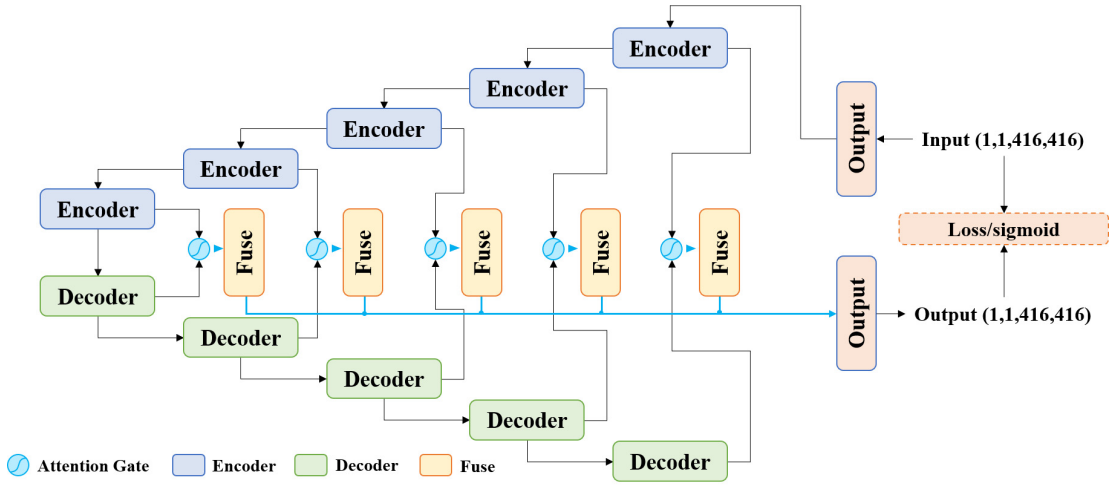


FIGURE 3. An example of DeepRidge model with five layers

Encoder-decoder attention connect structure. As shown in Figure 4, the encoder network and the decoder network are built based on the inverted residual structure. First, 1×1 convolution is used to achieve dimensionality reduction. Then, the attention module is used to connect the feature maps of the encoder network and the decoder network. After that, a 1×1 Conv layer is used to reduce the multi-channel feature maps into one channel. And finally, a deconvolution layer is employed to obtain the fused feature map of the n-th scale.

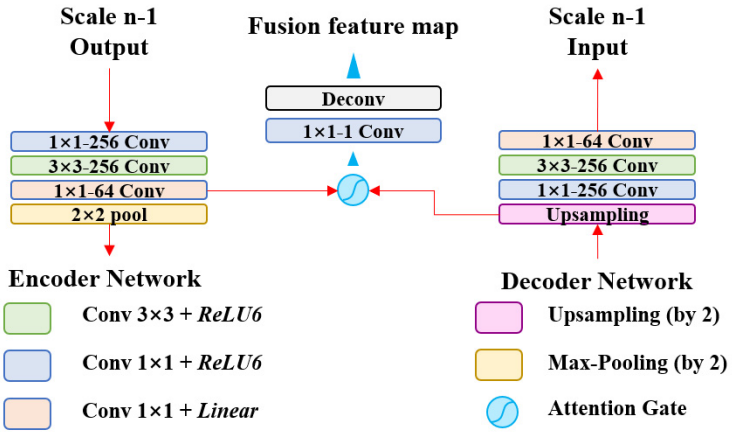


FIGURE 4. Encoder-decoder structure

Among them, the 1×1 dimensionality reduction convolution uses a linear activation function to avoid information loss during the dimensionality reduction process. This is because during the transformation process, low-dimensional information needs to be mapped to high-dimensional space, and then remapped back to low-dimensional space through ReLU. If the dimension of the output is relatively high, the information loss in the transformation process is small; if the dimension of the output is relatively low, the information loss in the transformation process is large. Therefore, linear functions are used instead of ReLU when performing dimensionality reduction transformation. For other convolution layers, ReLU6 is used instead of ReLU as the activation function. The purpose of replacing ReLU with ReLU6 is to ensure that the low-precision float16 of the embedded platform can also maintain a good numerical resolution. If there is no limit to the output value of ReLU, the output range is from 0 to positive infinity, and the low-precision float16 cannot accurately describe its value, which will cause a loss of precision. Therefore, in order to ensure the running effect on the embedded platform, ReLU6 is used instead of ReLU as the activation function.

In addition, the encoder and decoder need to go through a 2×2 max pooling layer after all convolution transformations. The main purpose of introducing the maximum pooling layer in the encoder-decoder network is to achieve downsampling, and ultimately achieve scale and receptive field changes. At the same time, it can also achieve the purpose of reducing dimensionality, removing redundant information, compressing features, simplifying network complexity, reducing calculation load, and reducing memory consumption, so that the model is more suitable for embedded platforms with limited resources.

The Attention Gate is introduced to suppress the irrelevant regions in the input spectrogram and enhance the salient features of a specific local region. When concatenating the features from each resolution level of the encoder with the corresponding ones in the decoder, an attention module is employed by the DeepRidge network to generate a gating signal that modulates the importance of the features at different spatial locations.

Figure 5 illustrates the structure of the Attention Gate. The outputs of the encoder e and the decoder d, F_e and F_d, are first convolved to halve their channel numbers. Then, they are added together and activated by a ReLU6 function, followed by a 1×1 convolution to obtain a single-channel feature map. A sigmoid function is then applied to producing a 1-dimensional attention map with the same size as x. Finally, the attention map and the encoder output are element-wise multiplied to yield the weighted vector.

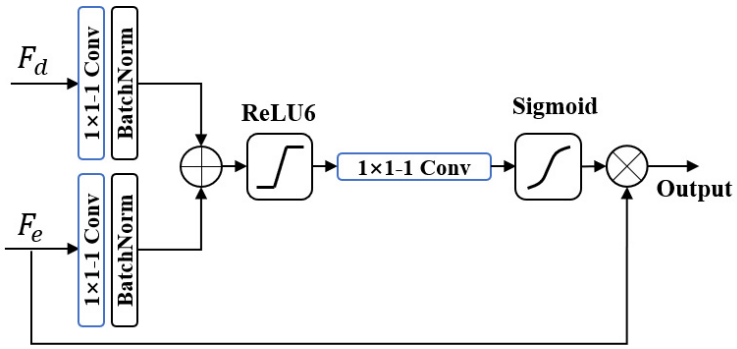


FIGURE 5. Schematic of Attention Gate

3.2. **Loss function.** To analyze the changes of the kurtosis of the energy ridges in the spectrogram, the ridge positions need to be extracted first, and the specific energy values of the ridges can be obtained from the corresponding original data. Therefore, ridge

detection can be regarded as a binary classification problem, that is, whether a pixel is on a ridge or not. Hence, we define a pixel-level loss function based on binary cross-entropy.

For a given training dataset containing N spectrograms, D is

$$D = \{(X_n, Y_n), n = 1, \ldots, N\} \tag{1}$$

$$X_n = \{x_{n,i}, i = 1, \ldots, I\} \tag{2}$$

$$Y_n = \{y_{n,i}, i = 1, \ldots, I\} \tag{3}$$

where X_n is the original spectrogram input in the training set, Y_n is the corresponding ground-truth ridge map in the training set, and I is denoted as the number of pixels in each spectrogram. Our goal is to train a DeepRidge model that takes the raw spectrogram as input and outputs the corresponding ridge map. The output of the DeepRidge model is the fusion of the outputs of the encoders of each scale, so the loss function is defined as follows:

$$loss(F_i; W) = \begin{cases} \log[1 - S(F_i; W)], & \text{if } y_i = 0 \\ \log(S(F_i; W)), & \text{otherwise} \end{cases} \tag{4}$$

$$L(W) = \sum_{i=1}^{I} \left(\sum_{f=1}^{K} loss(F_{f,i}; W) + loss(F_{o,i}; W) \right) \tag{5}$$

where F_i is the feature map output by the network at piexl i position, W is a series of weights of the network, $S()$ is the standard sigmoid function, K is the number of encoder-decoder pairs, and $F_{o,i}$ is the output result of the model.

4. **Experiments and Results.** In this section, we first introduce the experimental set-up and then analyze the ridge detection effect of DeepRidge compared to other networks. Finally, we studied further optimization method of the embedded platform and a performance test was carried out.

4.1. **Experimental settings.** In this section, we describe our experiments in detail in terms of implementation details, datasets, evaluation options, and comparison of state-of-the-art algorithms.

1) Implementations Details

We built the DeepRidge network based on the PyTorch framework. The weights of all convolution layers in the network are initialized using the "kaiming" method. In training, the initial global learning rate is set to 1e-6 and divided by 10 after every 10k iterations. Momentum decay and weight decay are set to 0.9 and 1e-8, respectively. Batch size is 10, epoch is 1000, using the root mean square propagation (RMSProp) to update network parameters. The training platform and testing platform parameters are shown in Table 1.

Table 1. Training platform and inference platform parameters

Platform	CPU	RAM	GPU
Training	Intel Xeon Gold 6133	40GB	NVIDIA Tesla V100 32GB
Inference	Intel Core i7-12700F	32GB	NVIDIA GeForce RTX 3060 12GB

2) Dataset

The hoist noise data were collected from the main well of Dingji Coal Mine of Huainan Mining Group and saved as a file every 60 seconds in the format of .wav. The audio timing data were processed by short-time Fourier transform to generate a spectrum map with

a longitudinal resolution of 416. For the convenience of model processing, the spectrum map was divided into 416×416 segments. Then manual marking was used to generate labels for each segment. The training set data contained 396 segments and the test set contained 180 segments.

3) Evaluation Metrics

For each spectral image segment, precision and recall can be calculated by comparing the detected ridges to the ground-truth annotated by humans. Then, the F-measure can be calculated as an overall measure of performance evaluation. To balance precision and recall, β is set as 1.

$$F\text{-}measure = \left(1 + \beta^2\right) \times \frac{(Precision \bullet Recall)}{(\beta^2 \times Precision + Recall)} \tag{6}$$

Specifically, three different F-measure-based metrics are used in the evaluation.

ODS (optimal dataset scale): ODS is to choose a fixed threshold applied to all images, and makes the whole data set on the F-score the largest.

OIS (optimal image scale): OIS is a different threshold that maximizes the F-score for each image.

AP (average precision): AP is the integral of the PR curve (that is, the area under the PR curve).

Considering that the energy ridge has a certain width, if the detected energy ridge is not more than 2 pixels away from the human-labeled curve, it is still considered to be correctly predicted.

4) Comparison Methods

We compared the performance of DeepRidge with other state-of-the-art methods.

SegNet [27]: SegNet is a deep neural network model for image segmentation. It consists of an encoder-decoder architecture with skip connections that preserve spatial information. SegNet can achieve high accuracy on various segmentation tasks with limited training data.

Attention U-Net [31]: Attention U-Net is a deep neural network model for image segmentation that incorporates attention mechanisms. It is based on the U-Net architecture, but with an additional Attention Gate module that selectively focuses on the most relevant features. Attention U-Net can improve the segmentation performance on various medical and natural images.

DeepCrack [29]: DeepCrack is a deep learning model for crack detection and segmentation in images. It is based on the fully convolution network (FCN) architecture, but with a novel crack refinement module that enhances the crack features and suppresses the background noise.

U-Net [30]: A network model composed of encoder-decoder with skip connections. It is widely used in the field of image segmentation and can train a high-precision model with a limited data set.

4.2. **Overall performance.** Figure 6 shows the precision-recall curves of the five deep learning methods. It can be seen that the precision-recall curve of DeepRidge has a curve closer to the upper right corner. U-Net performs the worst, which indicates that the method proposed in this paper to comprehensively judge the position of the ridge line by fusing the feature maps at different scales is effective.

The quantitative evaluation results of the five deep learning models are shown in Table 2. The ODS measured by DeepRidge is 0.87, which performs well. DeepCrack has the highest ODS, but its AP is lower than DeepRidge. By introducing inverted residual structure and attention mechanism in the architecture, DeepRidge effectively controls the number

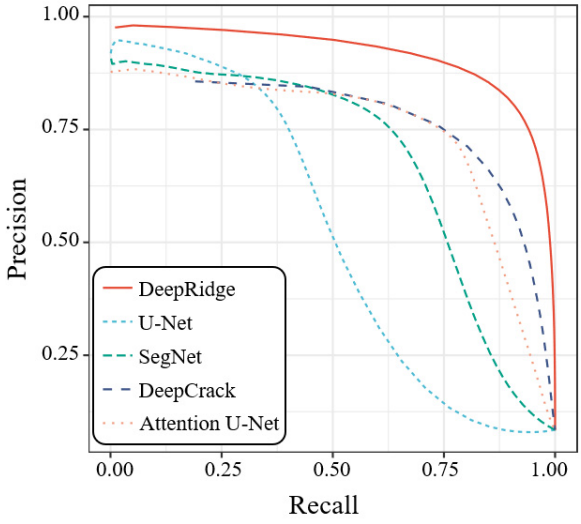


FIGURE 6. Precision-recall curves on the test dataset

TABLE 2. Quantitative evaluation of five deep learning models

Methods	OIS	ODS	AP	FLOPs/million	Parameters/million
DeepRidge	0.88	0.87	0.82577	105,452	11.66
U-Net [30]	0.46	0.45	0.43279	81,974	13.39
Attention U-Net [31]	0.44	0.42	0.65501	241,004	42.88
SegNet [27]	0.70	0.68	0.59462	77,027	18.81
DeepCrack [29]	0.91	0.93	0.68546	361,049	30.90

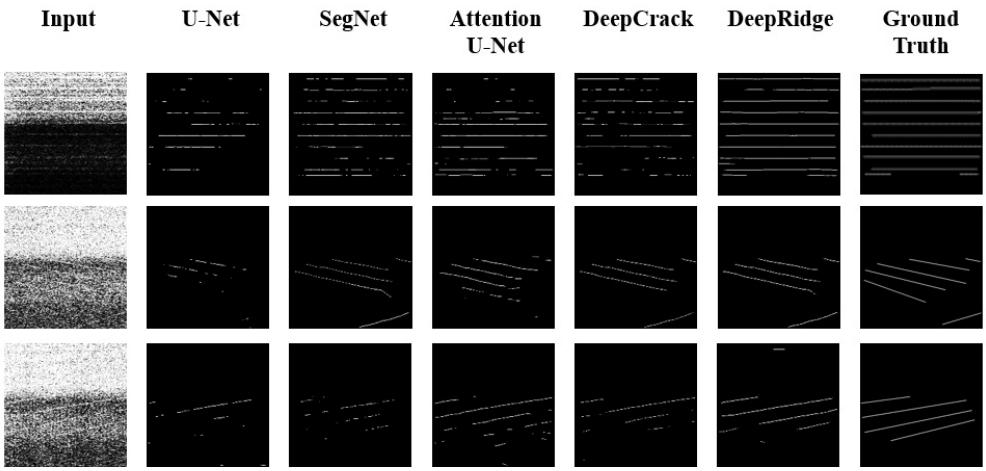


FIGURE 7. Visual comparison of five algorithms in typical scenarios, the threshold for each method is set to the OSD value of that method.

of parameters and calculation of the model, which are reduced by 62.3% and 70.8% respectively compared with DeepCrack, providing help for subsequent deployment on resource-constrained embedded platforms.

We also compared the five methods in terms of vision. In Figure 7, we show the spectrograms of four typical inputs and the results of the five methods for the outputs, with the threshold of each method set to the OSD value of the method. It can be seen that

DeepRidge performs best in various typical scenes. This is because DeepRidge fuses feature maps of different scales, which is closer to the human perception. Although Attention U-Net also adopts layer-hopping connection and attention mechanism, it fails to effectively fuse multi-scale features, which makes it difficult to effectively detect ridges in some scenes.

In summary, the DeepRidge model effectively improves the ability of the model to detect ridges by fusing multi-scale convolution features in the encoder and decoder networks through the attention mechanism, and significantly reduces the model's size by introducing an inverted residual structure to construct the encoder-decoder.

4.3. **DeepRidge of different sizes and depths.** The previous experiments have shown that the DeepRidge model has a good effect in detecting the energy ridge of the hoist noise spectrum. In this section, we will modify the "Standard Edition" DeepRidge model shown in Figure 8 to study the difference in detection effect of models with different configurations by changing the scale or number of encoder-decoder pairs. Finally, we obtain four models including the "standard version". The model configuration parameters are shown in Table 3. Finally, we test these models on the test dataset.

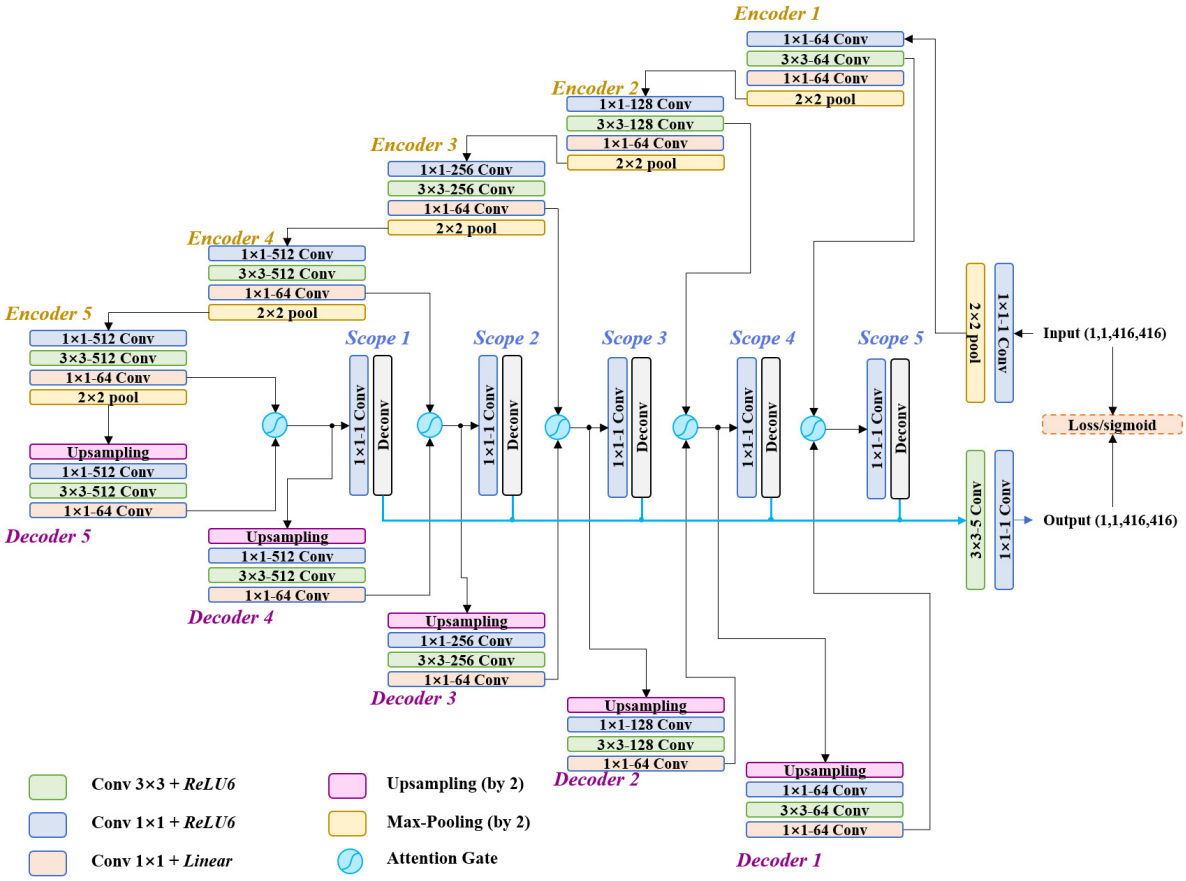


FIGURE 8. "Standard Edition" DeepRidge model structure

The DeepRidge standard encoder consists of a 1×1 dimensional raising convolution, a 3×3 feature detection convolution, a 1×1 dimensional reduction convolution, and a pooling layer. The corresponding DeepRidge standard decoder consists of a 1×1 dimensional raising convolution, a 3×3 feature detection convolution, a 1×1 dimensionality reduction convolution, and an upsampling layer. In order to explore the effect of the depth of the convolution layer in the encoder-decoder on the performance of DeepRidge, we construct a

TABLE 3. DeepRidge models in different configurations

Model	Category	Number
DeepRidge-Shoal	Standard	4
DeepRidge-Standard	Standard	5
DeepRidge-Heavy	Heavy	5
DeepRidge-Deep	Standard	6

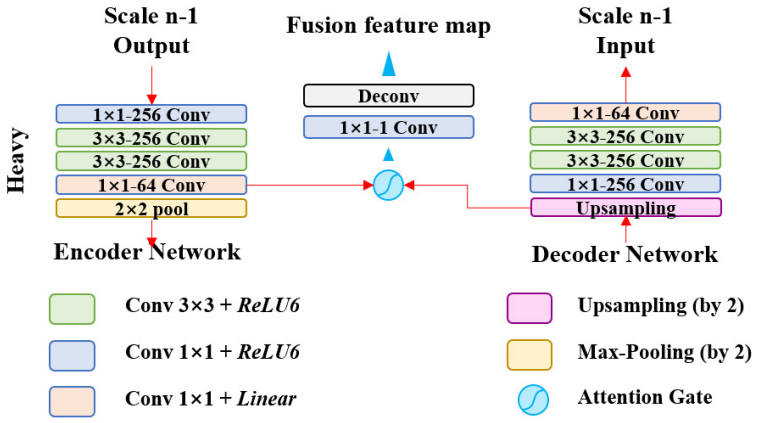


FIGURE 9. Heavy edition encoder-decoder

heavy encoder and a corresponding decoder, as shown in Figure 9. The attention module is also used to connect the encoder and decoder networks.

To test the effect of the number of different scale feature fusions on DeepRidge's performance, we constructed "Deep Edition" and "Shoal Edition" by increasing or decreasing the number of encoder-decoder pairs based on "Standard Edition" DeepRidge.

Figure 10 shows the precision-recall curves of four different configurations of DeepRidge models, and it can be seen that the "Standard Edition" DeepRidge still occupies the optimal position. Table 4 shows the quantitative test results of four different configurations of DeepRidge, which also confirms this point.

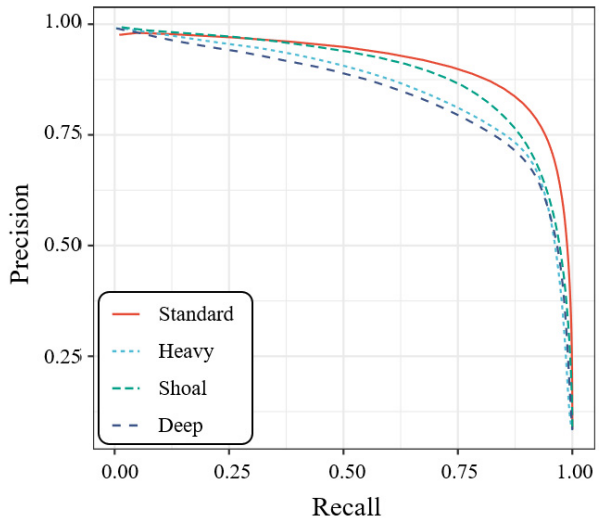


FIGURE 10. Different configurations DeepRidge models precision-recall curves on the test dataset

TABLE 4. Quantitative evaluation of DeepRidge models in different configurations

Methods	OIS	ODS	AP	FLOPs/ million	Parameters/ million	Recall	Precision
Standard	0.88	0.87	0.82577	105,452	11.66	0.8500	0.8546
Heavy	0.64	0.638	0.77194	197,955	22.87	0.7611	0.8055
Shoal	0.84	0.85	0.80034	77,027	18.81	0.8373	0.8047
Deep	0.66	0.67	0.76120	361,049	30.90	0.8504	0.7343

The test results show that the "standard version" DeepRidge is in the optimal configuration in terms of the number of scale fusion. When the number of encoder-decoder pairs is increased, the maximum precision achieved by the model at ODS is reduced. Through analysis, it is believed that this is because the size of the feature map of the sixth scale is only 1/32 of the original input image, which contains too little information, and the negative impact on the final result is greater than the contribution, resulting in a decrease in precision. As shown in Figure 11, the visual display of the output feature map of each scale of the "standard version" DeepRidge model. When the number of encoder-decoder pairs is reduced, the final result lacks the contribution of the fifth scale feature map, which also leads to a decrease in precision. The fifth scale feature map is 1/16 of the size of the original image, which still contains a certain amount of information. When the number of encoder-decoder convolution layers is increased on the basis of the "standard version", the test results show that the recall rate of the DeepRidge-Heavy model at ODS decreases. This is caused by the overfitting of the model.

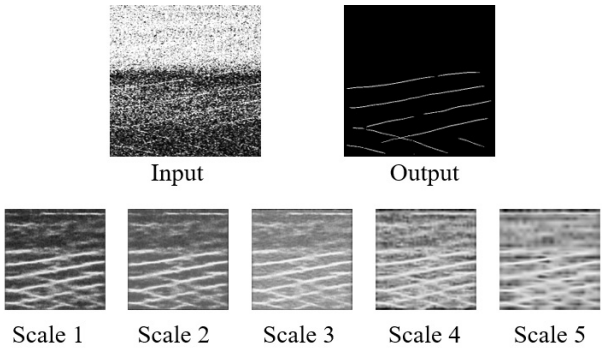


FIGURE 11. A real example of ridge detection using DeepRidge-Standard

By exploring the experiment of the model parameter configuration, we choose a more appropriate model. The actual effect and computing power requirements of the target model are relatively balanced. From the perspective of engineering practice, the DeepRidge model is best deployed on edge devices to improve the response speed of the system. Therefore, it is necessary to further verify the operation effect of DeepRidge on embedded platforms.

4.4. **Experimental settings and model optimization.** Our ultimate goal is to deploy the DeepRidge model in embedded devices to achieve noise data collection and in-situ analysis near the hoist. Therefore, we need to optimize the model specifically to reduce the model size and improve the operational efficiency. The target platform we selected is the RK3588S processor produced by Rockchip. It is equipped with 4 core ARM Cortex-A76 and 4 core ARM Cortex-A55, and integrates NPU with 6TOPS computing power. It

TABLE 5. Parameters of embedded platform

Parameter	Value
CPU	Rockchip RK3588S Quad-core ARM Cortex-A76 up to 2.4GHz Quad-core Cortex-A55 CPU up to 1.8GHz
RAM	64-bit 8 GB LPDDR4X at 2133MHz
Disk	32 GB eMMC
OS	Ubuntu 22.04 Desktop Kernel version Linux-5.10-LTS U-boot-2017.09
Toolkit	rknn-toolkit2 v1.5.0

adopts a three-core architecture design and supports int4/int8/int16/FP16/BF16/TF32. The device configuration used in this experiment is shown in Table 5.

When converting the model to rknn format, we reconstructed MaxUnpool2d using Pytorch tensor. sactter_operation, as rknn does not support MaxUnpool2d operator. As shown in Table 6, the quantitative evaluation data of DeepRidge running on the RK3588S platform are shown. On the RK3588S platform, we test the performance of the model after float32, int8 and mixed quantization. It can be seen that the OIS, ODS and AP of the model without quantization are higher than those of the model quantized to int8, but the inference speed is lower. Mixed quantization can change the specified quantized layer to a non-quantized layer, and its test effect is between float32 and int8.

TABLE 6. Quantitative evaluation of DeepRidge on RK3588S platform

Parameter	Float32	Int8	Hybrid quantization
OIS	0.862	0.81	0.854
ODS	0.847	0.79	0.851
AP	0.8026	0.6471	0.7675
Time/frame	1.196743 s	0.776506 s	0.85417 s

5. **Conclusion.** In this paper, we have successfully developed DeepRidge, a novel deep learning approach for detecting the trend changes of energy ridges in the operational noise spectrum of mining hoists. Compared to state-of-the-art methods, DeepRidge demonstrated superior performance on the test dataset, particularly in accurately identifying energy ridges.

The main contribution of DeepRidge lies in its ability to precisely capture subtle changes in spectral energy ridges through multi-scale feature fusion and attention mechanisms. This capability is crucial for real-time monitoring of hoist operating conditions, as anomalies in energy ridges can indicate early signs of equipment failure. Detecting these changes early can prevent potential breakdowns, reduce downtime, and enhance the safety and efficiency of mining operations.

While DeepRidge has shown promise in laboratory settings, we acknowledge that its deployment and application in actual mining environments present challenges. Future work will focus on further optimizing the model to adapt to varying operational conditions and ambient noise levels, as well as exploring integration methods with existing mining monitoring systems.

Acknowledgment. This work is supported by the Scientific Research Foundation for High-level Talents of Anhui University of Science & Technology (Grant No.2022yjrc46); the National Natural Science Foundation of China project: "Research on Cloud-Fog Resource Intelligent Scheduling Method for Mine Safety Control Linkage System" (No. 52374 154). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

Data Availability Statement. The dataset was established for the author collection and does not contain sensitive information. Dataset and key codes are publicly available on GitHub to facilitate related research efforts. Open source address is: https://github.com/AlphaSkyate/deepRidge.

REFERENCES

[1] J. Li, J. Xie, Z. Yang and J. Li, Fault diagnosis method for a mine hoist in the Internet of Things environment, *Sensors*, vol.18, no.6, 2018.
[2] X. Liu, D. Yang, Q. Li and J. Liu, Fast average prediction method based on the upper-lower limit theory for ship's mechanical noise, *Applied Acoustics*, vol.197, 108894, 2022.
[3] G. Wu, N. Yan, K.-N. Choi, H. Jung and K. Cao, A two-step vibration-sound signal fusion method for weak fault feature detection in rolling bearing systems, *Advances in Mechanical Engineering*, vol.13, no.12, 16878140211067155, 2021.
[4] A. Rahman, M. Khan and A. Mushtaq, Predictive modeling of surface wear in mechanical contacts under lubricated and non-lubricated conditions, *Sensors*, vol.21, no.4, 1160, 2021.
[5] Y. Li, Y. Yang, Y. Chen and Z. Chen, Iterative characteristic ridge extraction for bearing fault detection under variable rotational speed conditions, *ISA Transactions*, vol.119, pp.172-183, 2022.
[6] N. Laurent and S. Meignen, A novel ridge detector for nonstationary multicomponent signals: Development and application to robust mode retrieval, *IEEE Transactions on Signal Processing*, vol.69, pp.3325-3336, 2021.
[7] Z. Lv, Y. Geng, H. Hou, T. Lin, X. Dai and Y. Hou, BM3D with bi-hard thresholding for image denoising in high-level noise situations, *International Journal of Innovative Computing, Information and Control*, vol.19, no.4, pp.1103-1116, 2023.
[8] Y. Liu, Z. Xie and H. Liu, LB-LSD: A length-based line segment detector for real-time applications, *Pattern Recognition Letters*, vol.128, pp.247-254, 2019.
[9] Q. Huang and J. Liu, Practical limitations of lane detection algorithm based on Hough transform in challenging scenarios, *International Journal of Advanced Robotic Systems*, vol.18, no.2, 17298814211008752, 2021.
[10] E. Oğuz, A. Küçükmanisa, R. Duvar and O. Urhan, A deep learning based fast lane detection approach, *Chaos, Solitons & Fractals*, vol.155, 111722, 2022.
[11] L. Ding and A. Goshtasby, On the Canny edge detector, *Pattern Recognition*, vol.34, no.3, pp.721-725, 2001.
[12] X. S. Tang, X. Xie, K. Hao, D. Li and M. Zhao, A line-segment-based non-maximum suppression method for accurate object detection, *Knowledge-Based Systems*, vol.251, 108885, 2022.
[13] P. Dollár and C. L. Zitnick, Fast edge detection using structured forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.37, no.8, pp.1558-1570, 2015.
[14] L. Feng, J. Wang and C. Ding, Image edge detection algorithm based on fuzzy radial basis neural network, *Advances in Mathematical Physics*, DOI: 10.1155/2021/4405657, 2021.
[15] D. Martin, C. Fowlkes and J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.5, pp.530-549, 2004.
[16] G.-S. Shokouh, B. Magnier, B. Xu and P. Montesinos, Ridge detection by image filtering techniques: A review and an objective analysis, *Pattern Recognition and Image Analysis*, vol.31, pp.551-570, 2021.
[17] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, *International Journal of Computer Vision*, vol.30, pp.117-156, 1998.
[18] S. Berlemont and J.-C. Olivo-Marin, Combining local filtering and multiscale analysis for edge, ridge, and curvilinear objects detection, *IEEE Transactions on Image Processing*, vol.19, no.1, pp.74-84, 2010.

[19] X. Yang and W. Wang, Road identification in aerial images on fractional differential and one-pass ridge edge detection, *Journal of Applied Remote Sensing*, vol.8, no.1, 083597, 2014.
[20] J. Sun and S. Mao, River detection algorithm in sar images based on edge extraction and ridge tracing techniques, *International Journal of Remote Sensing*, vol.32, no.12, pp.3485-3494, 2011.
[21] P. Subirats, J. Dumoulin, V. Legeay and D. Barba, Automation of pavement surface crack detection using the continuous wavelet transform, *2006 International Conference on Image Processing*, pp.3037-3040, 2006.
[22] F. Faghih and M. Smith, Combining spatial and scale-space techniques for edge detection to provide a spatially adaptive wavelet-based noise filtering algorithm, *IEEE Transactions on Image Processing*, vol.11, no.9, pp.1062-1071, 2002.
[23] Z. Zhang, S. Ma, H. Liu and Y. Gong, An edge detection approach based on directional wavelet transform, *Computers & Mathematics with Applications*, vol.57, no.8, pp.1265-1271, 2009.
[24] A. Khorram, F. Bakhtiari-Nejad and M. Rezaeian, Comparison studies between two wavelet based crack detection methods of a beam subjected to a moving load, *International Journal of Engineering Science*, vol.51, pp.204-215, 2012.
[25] Y. Shi, L. Cui, Z. Qi, F. Meng and Z. Chen, Automatic road crack detection using random structured forests, *IEEE Transactions on Intelligent Transportation Systems*, vol.17, no.12, pp.3434-3445, 2016.
[26] D. G. Lee, Y. H. Shin and D.-C. Lee, Land cover classification using SegNet with slope, aspect, and multidirectional shaded relief images derived from digital surface model, *Journal of Sensors*, vol.2020, pp.1-21, 2020.
[27] V. Badrinarayanan, A. Kendall and R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.39, no.12, pp.2481-2495, 2017.
[28] X. Zan, X. Zhang, Z. Xing, W. Liu, X. Zhang, W. Su, Z. Liu, Y. Zhao and S. Li, Automatic detection of maize tassels from UAV images by combining random forest classifier and VGG16, *Remote Sensing*, vol.12, no.18, 3049, 2020.
[29] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang and S. Wang, DeepCrack: Learning hierarchical convolutional features for crack detection, *IEEE Transactions on Image Processing*, vol.28, no.3, pp.1498-1512, 2019.
[30] O. Ronneberger, P. Fischer and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, N. Navab, J. Hornegger, W. Wells and A. Frangi (eds.), Cham, Springer, 2015.
[31] Z. Xu, S. Wang, L. V. Stanislawski, Z. Jiang, N. Jaroenchai, A. M. Sainju, E. Shavers, E. L. Usery, L. Chen, Z. Li et al., An attention U-Net model for detection of fine-scale hydrologic streamlines, *Environmental Modelling & Software*, vol.140, 104992, 2021.

Author Biography

Yifan Meng is a lecturer and master's supervisor in the Department of Intelligent Science at the School of Artificial Intelligence, Anhui University of Science & Technology. He received his Doctorate in Mechanical Engineering from Anhui University of Science & Technology in 2022. His research focuses on artificial intelligence and the Industrial Internet of Things, with a particular emphasis on teaching and research in the application of artificial intelligence, machine learning, and related technologies to the Industrial Internet of Things, industrial production process control, and intelligent information processing. He has also contributed to a project funded by the National Natural Science Foundation of China.

Heng Zhang graduated from Anhui Polytechnic University with a Bachelor's degree in Automation in 2022. He is currently pursuing a Master's degree in Artificial Intelligence at the School of Artificial Intelligence, Anhui University of Science & Technology. His primary research interests include intelligent information processing and industrial equipment fault perception.

Xiaoming Zhang graduated from Anhui University of Science & Technology in 2021 with a Doctor of Engineering degree. He is currently a lecturer at the School of Electrical and Information Engineering at Anhui University of Science & Technology. His main research areas include mining Internet of Things and intelligent information processing. He is the principal investigator of a university-level research project at Anhui University of Science & Technology and has also participated in a project funded by the National Natural Science Foundation of China.