The efforts for the interpretability of deep models

Investigating the interpretability of deep models is a meaningful research field. To explore the learning ability of the model as thoroughly as possible, we conduct the following works.

1. The vibration signals collected in the experiment often are masked by electrical noise and coupling interference from other parts. To better understand the signal representation learned by the deep model at different speeds, we follow the literature [*Yunjia Dong. Investigation of dynamic simulation and transfer learning based fault diagnosis method for rolling element bearings. MS thesis. Harbin Institute of Technology, 2019.*] to establish a dynamic model for the bearing, because the simulation signal can significantly reflect the bearing fault characteristics, which is conducive to reducing the interference on the learning process visualization. The essential parameters for establishing the differential equations of bearing dynamics are listed as follows.

Table 1.1 Bearing geometric parameters and fault frequency.

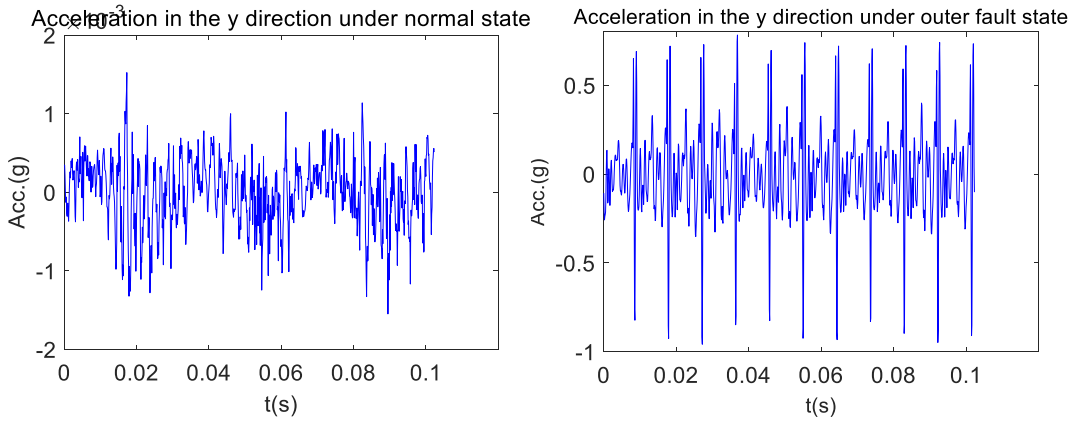
Parameter	Value
Pitch diameter D_m *(mm)*	39.04
Number of rolling elements N_b	9
Ball diameter d_b (mm)	7.94
Contact angle α (°)	0
Fault frequency of the inner race f_{in}	5.42 f_r
Fault frequency of the inner race f_{out}	3.58 f_r
Fault frequency of the ball f_{ball}	4.71 f_r

Note: f_r denotes the rotating frequency of the shaft and the bearing inner race.

Table 1.2 Dynamic simulation parameters of the bearing.

Parameter	Value	Parameter	Value
Mass of inner race-shaft m_{in} (kg)	50	Contact stiffness k_b (N/m)	8.753×10^9
Shaft-connection stiffness k_{im} (N/m)	7.42×10^7	Acceleration of gravity g (m/s^2)	9.8
Shaft-connection damping c_{im} (Ns/m)	1376.8	Fault width L (mm)	0.5
Mass of outer race-bearing seat m_{out} (kg)	5	Fault depth h (mm)	0.5

Bearing seat foundation connection stiffness k_{out} (N/m)	1.51×10^{7}	Radial clearance c_r (m)	2×10^{-6}
Bearing seat foundation connection damping c_{out} (Ns/m)	2210.7	Eccentricity e (m)	15×10^{-6}



(a) (b)

Fig. 1.1 Vibration signal of (a) normal bearing; (b) bearing with outer race fault.

Fig. 1.1(a) presents the vibration signal of a normal bearing at 1800 rpm, and Fig. 1.1(b) gives the vibration signal of a bearing with an outer race defect at 1800 rpm. Since the environmental noise is not incorporated into the simulation signal, the periodic impulse shocks induced by the outer race defect are evident in Fig. 1.1(b). The bearing dynamics simulation code is available at *https://github.com/LinBo-Team/Support-code/tree/eb5c27e71db6ba9d95d356e1da63c0977c306acf/Bearing%20dynamics%20simulation* (MATLAB version).

2. It is worth noting that the raw vibration samples are not preprocessed or transformed here in order to visually analyze the processing of cross-speed vibration signals by the deep model. The output of the convolution layer is extracted during the training process, as displayed in Fig. 1.2. At the initial training stage of the deep model, the extracted state representations of the convolution layer are irregular due to the random initialization of network parameters such as the convolution kernel. With the increase of training epoch, the network gradually learns the characteristics of periodic pulses buried in vibration signals. The envelope spectrum suggests that the learned

representations are characterized by medium and low-frequency bands, and the irrelevant frequency components gradually decrease as the deep model trains.

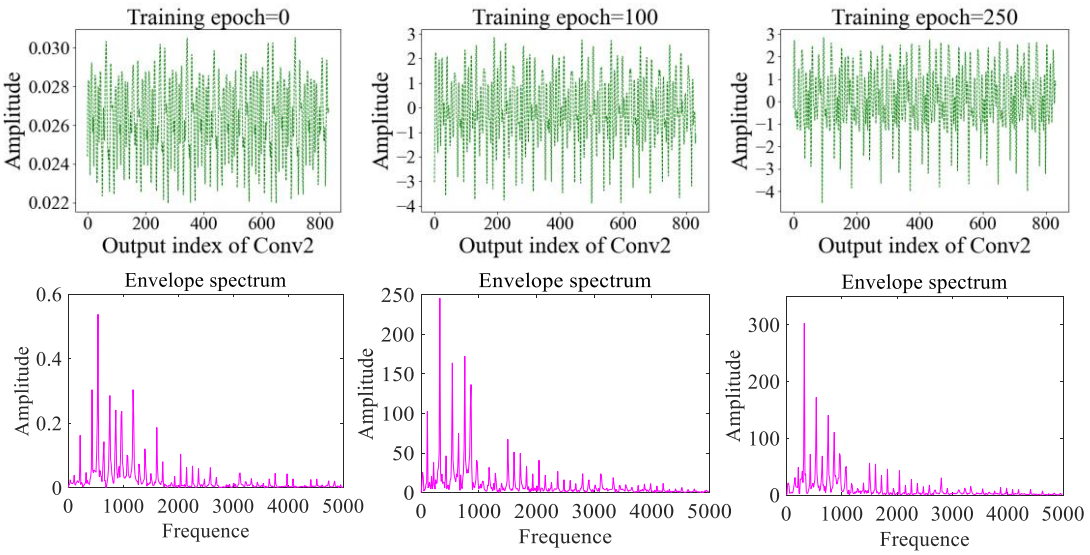
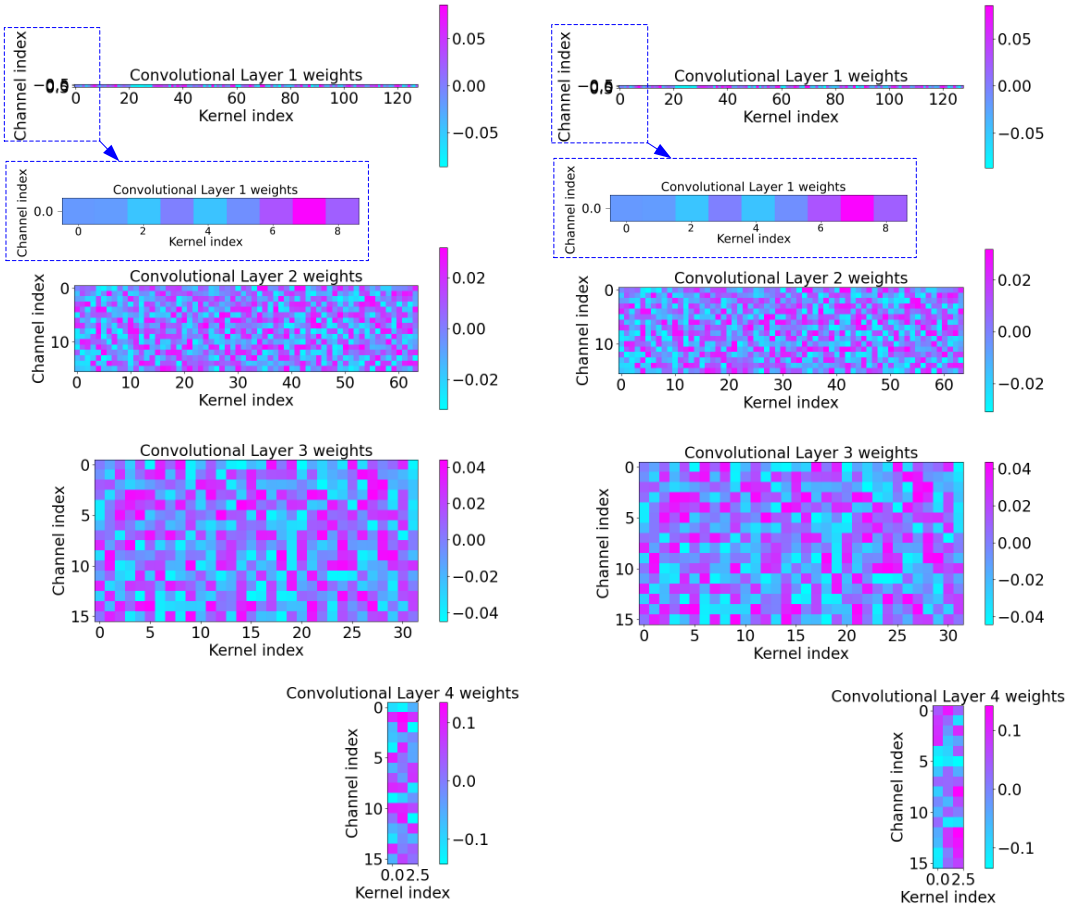


Fig. 1.2 The state representation visualization of the convolution layer 2.

3. Fig. 1.3 visualizes the convolution weights of the deep network. Fig. 1.3(a) shows the convolution kernel when the deep model is used to identify the health state of the outer race fault bearing and the normal bearing under 1800 rpm, and Fig. 1.3(b) depicts the convolution kernel when identifying the health state of the two bearings under 1200 rpm.

The well-trained model deals with objects that are very similar, such as fault characteristics, and the rotational speed mainly affects the pulse impact period. Fig. 1.3 shows that the convolution kernels of the first layer and the second layer of the network are similar for the fault diagnosis tasks at two different rotational speeds, especially the convolution kernel of the first layer. This phenomenon implies that the shallow convolution layer focuses on the general characteristics of learning tasks. On the contrary, the convolution kernel weights of the third layer and the fourth layer are quite distinct, especially in the last convolution layer, the weight distribution difference is the most significant. According to the above analysis, the regularized adaptive weight optimization strategy determines the weight of each working condition adaptively based on the domain loss of the final output of the model.



(a) (b)

Fig. 1.3 Convolution kernel visualization of the deep model under (a) 1200 rpm (b) and 1800 rpm.