

Research Article

Estimation of Vehicle Dynamic Response from Track Irregularity Using Deep Learning Techniques

Fei Yang,¹ Xiaoli Hao ,² Hang Zhang,³ Jie Jiang,² Zhixing Fan,² and Zilong Wei ¹

¹Infrastructure Inspection Research Institute, China Academy of Railway Sciences, Beijing, China

²School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China

³School of Civil Engineering, Beijing Jiaotong University, Beijing, China

Correspondence should be addressed to Zilong Wei; weizilong@rails.cn

Received 19 August 2022; Revised 9 November 2022; Accepted 29 November 2022; Published 20 December 2022

Academic Editor: Georges Kouroussis

Copyright © 2022 Fei Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To improve the quality of track maintenance work, it is a desire to estimate vehicle dynamic behavior from track geometry irregularities. This paper proposes a deep learning model to predict vehicle responses (e.g., vertical wheel-rail forces, wheel unloading rate, and car body vertical acceleration) using deep learning techniques. In the proposed CA-CNN-MUSE model, convolutional neural networks (CNNs) are used to learn features of track irregularities, and multiscale self-attention mechanisms (MUSE) are employed to capture the long-term and short-term trends of sequences. Coordinate attention (CA) is introduced into CNN to focus on important interchannel relationships and important spatial mileage points. The experiments were performed on a multibody simulation model of the vehicle system and the measured data of the actual high-speed line. The results show that the CA-CNN-MUSE has high prediction accuracy for vertical vehicle responses and fast computation speed. The predicted time-domain waveforms and power spectral densities (PSDs) agree well with the actual vehicle responses. The main features of the lateral vehicle responses can also be captured by the proposed method, yet the results are not as good as the vertical ones.

1. Introduction

Assessment of track quality for track maintenance work is critical to ensure train running safety and passenger ride comfort in high-speed railways. The current evaluation method of track quality is based on the amplitudes of track geometry irregularities. However, it is not enough to use the single-track irregularity index to evaluate the track quality without considering vehicle dynamic response. For example, some problems may occur in some track sections, where the amplitude of each track irregularity does not exceed the limit but a large vehicle vibration response appears. Conversely, there are also some track sections where the amplitudes of some track irregularities exceed the limits but do not cause poor vehicle response. These problems show that the vehicle vibration response is the result of the nonlinear coupling of multiple track irregularities.

To improve the track quality assessment standards and the track maintenance work, research has emerged to relate

track geometry irregularities to vehicle response. The key to these works is to find a model to accurately estimate the vehicle response from track irregularities. Then, the track geometry assessment can be performed by combining the indexes of actual track irregularities and the predicted vehicle responses. The models are not only helpful for the track maintenance department to find the track locations that cause undesirable vehicle responses and estimate the degradation trend of the track but also helpful for facilitating the upgrades of the vehicle design technology.

There are several ways to build the model to predict vehicle response from track geometry irregularities. Some researchers focus on building a mechanical model to simulate the nonlinear dynamic behavior of the vehicle [1–3]. Using commercial software such as SIMPACK, researchers can build a 3D vehicle-track dynamic model to investigate the relationship between track irregularities and the dynamics performance of railway vehicles [4–6]. However, the performance of the mechanical model depends on the

reliability of model parameters and is easier to be affected by variations in reality. Therefore, putting a theoretical model into track maintenance practice is difficult because the actual parameters of a vehicle-track system are difficult to obtain and change with time. Besides, the numerical iteration method used in solving a mechanism model is time-consuming.

Some researches characterize the behavior of the vehicle-track system as linear transfer functions and estimate parameters based on system identification theory [7–9]. However, system identification theory can be applied to only the linear system and only constant speed conditions.

The alternative approach is to carry out a prediction of pure data using machine learning techniques. The machine learning method has high computational efficiency, good performance in modeling nonlinear mapping relationships and can be designed for varying speeds. Machine learning methods can be further categorized into traditional methods and deep learning methods. Traditional machine learning approaches predict vehicle responses including multilayer perception (MLP [10]), a set of backpropagation (BP [11]), decision tree, support vector machines, and other regression algorithms and their comparisons in [12], and NARX neural networks [13].

In recent decades, deep learning has achieved great success in sequence modeling. The main advantages of deep learning are strong nonlinear modeling ability, end-to-end training, and eventually great improvement in model accuracy. The classical sequence modeling methods in deep learning are recurrent neural networks (RNN [14]), long term short memory (LSTM [15]), and gated recurrent (GRU [16]), and their combinations with convolutional neural networks (CNN). Li et al. used LSTM to estimate car body vertical and lateral acceleration based on the simulation model [17]. Ma et al. proposed a CNN-LSTM model to predict car body vibration acceleration from track irregularities [18]. The model uses two layers of CNN to learn different band features of sequences and feeds the extracted features into two layers of LSTM to learn the mapping relationship between the input and the output. The accuracy of the model is superior to BP and LSTM.

Recently, with the emergence of the attention mechanism (AM [19]), deep learning has stepped into a new stage. The main idea of attention is to tell a model to attend “what” and “where” to focus on the most relevant information instead of the entire sequence. The traditional RNN and LSTM models are replaced by a multihead self-attention mechanism, which achieves better accuracy and allows more parallel computation in machine translation tasks [20]. With its success in sequence-to-sequence tasks, the self-attention mechanism has become a standard component for capturing long-term dependence. However, the self-attention mechanism also has shortcomings. With the deepening of self-attention levels, a certain input vector will be paid too much attention, resulting in insufficient use of local information. Therefore, Zhao et al. replaced multihead self-attention with multiscale attention (MUSE) to encode global and local relations in parallel.

Attention mechanisms are widely deployed for boosting the performance of modern deep neural networks [21–26]. Unlike channel attention, it only considers interchannel information but neglects the importance of positional information. Hou et al. proposed a novel lightweight coordinate attention for mobile networks [27], which can capture important interchannel relationships and precise positional information meanwhile.

In this paper, we established a CA-CNN-MUSE model, which combines CNN with MUSE and introduces coordination attention into CNN to estimate vehicle response. The experiments show that the model improves estimation accuracy, compared with LSTM and CNN-LSTM models and has good computation speed.

2. The Proposed Model

2.1. Dataset. In this paper, we use two datasets to investigate the proposed model. The first dataset is the “Inspection-Simulation” dataset, which means inspected track irregularities by high-speed comprehensive inspection train (see Figure 1) in China and simulated vehicle responses by a multibody model of the vehicle. The multibody dynamics (MBD) software SIMPACK is employed for the simulation of vehicle-track interaction. The CRH380B EMU trailer is modeled, in which the car body, bogie frames, and wheel sets are simplified as a rigid body and are connected using the primary and secondary suspensions. The second dataset is the “Inspection” dataset, which means that all the track irregularities and vehicle responses are inspected by track comprehensive inspection trains in actual high-speed railway lines. The items of track irregularities and vehicle responses of the two datasets are different.

The inspection-simulation dataset includes 4 track irregularities, i.e., left and right longitudinal levels and left and right alignment, coming from 3 Chinese high-speed railways, of which the mileage of line 1, line 2, and line 3 is 320 km, 120 km, and 80 km, respectively. The vehicle response data includes 14 quantities: wheel-rail force (left and right vertical forces of 1-axle, 2-axle, 3-axle, and 4-axle), wheel unloading rate (1-axle, 2-axle, 3-axle, and 4-axle), and vertical car body accelerations (front and rear).

The inspection dataset was extracted from the database of track comprehensive inspection trains on a high-speed line in China. The whole mileage is 17.5 km, the spatial sampling interval is 0.25 m, and the vehicle speed range is from 215 km/h to 245 km/h. The track irregularity includes 11 quantities: left and right longitudinal levels, left and right alignment, left and right long-wave longitudinal level, left and right long-wave alignment, gauge, cross-level, and twist. The vehicle response includes 6 quantities left and right vertical forces, left and right lateral forces, and vertical and lateral car body accelerations.

2.2. Model Structure. The prediction framework of vehicle dynamic response is established as shown in Figure 2. The network mainly consists of three main modules: CNN,

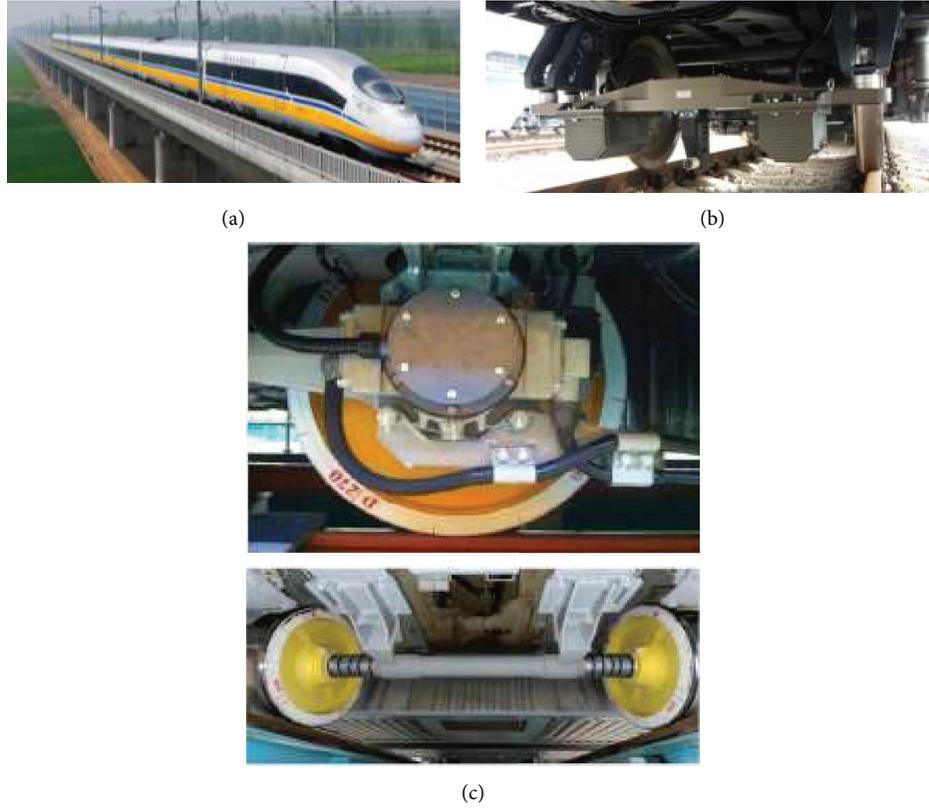


FIGURE 1: High-speed comprehensive inspection train. (a) CRH380BJ EMU, (b) track geometry inspection system, and (c) wheel-rail force inspection system.

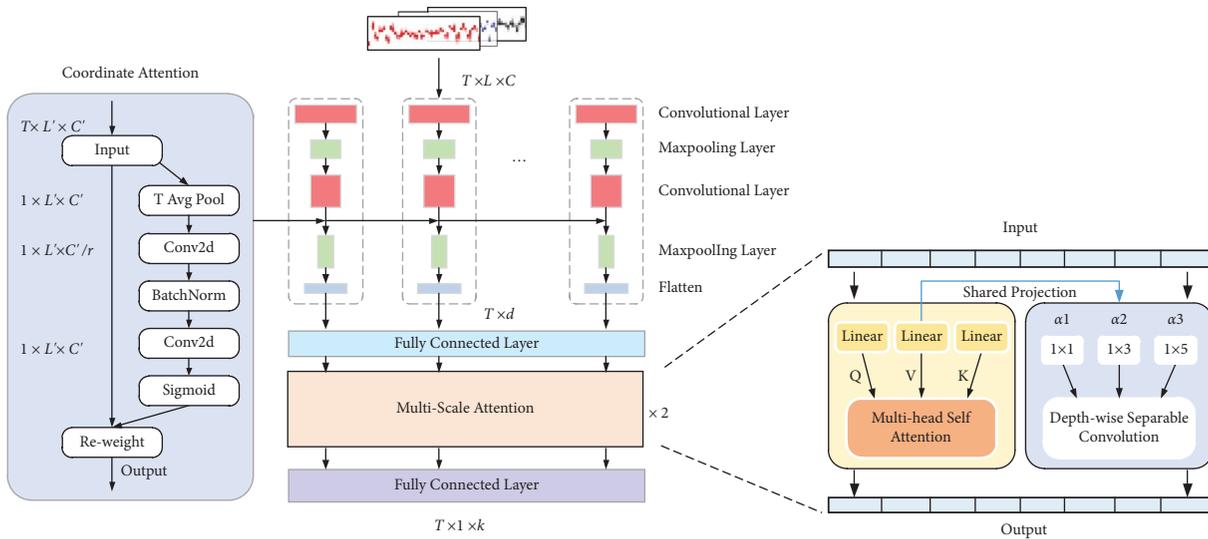


FIGURE 2: The CA-CNN-MUSE structure.

coordination attention (CA), and multiscale attention (MUSE). CNN is composed of alternately stacked convolution layers and pooling layers, which are used to extract different features of track irregularities. The results of CNN are fed into two stacking multiscale attention layers, each of which is composed of a multihead self-attention mechanism and depth-wise separable convolution to encode global and

local relations in parallel. Coordinate attention [27] is added to CNN to focus on important feature channels and important mileage points. There are fully connected layers before and after the multiscale self-attention module to perform nonlinear mapping and change the dimensions, and the last fully connected layer outputs the estimated vehicle responses.

2.2.1. CNN Module. Since the maximum management wavelength of track irregularity is 120 m, to fully capture long-distance wavelength information, we take all track irregularity data within 120 m as input. We use CNN to learn the features of input vectors and fed the features into the multiscale attention layer to estimate the vehicle response at the current mileage point. Assuming the current mileage

point is t , let us estimate the vehicle response at T mileage points at a time

$$Y = \{\hat{y}^{(t)}, \hat{y}^{(t+1)}, \dots, \hat{y}^{(t+T-1)}\}. \quad (1)$$

The input should be

$$X = \{[x^{(t-L+1)}, \dots, x^{(t-1)}, x^{(t)}], [x^{(t-L+2)}, \dots, x^{(t)}, x^{(t+1)}], \dots, [x^{(t-L+T)}, \dots, x^{(t+T)}, x^{(t+T-1)}]\}, \quad (2)$$

where $x^{(t)} \in R^{1 \times C}$, $y^{(t)} \in R^{1 \times K}$, $x^{(t)}$ is the C -dimensional vector, $\hat{y}^{(t)}$ is the K -dimensional vector, L is the number of mileage points within 120 m. The size of the input sequence is $T \times L \times C$, and the size of the output sequence is $T \times 1 \times K$. For the inspection-simulation dataset, C is 4 and K is 14. For the inspection dataset, C is 8 and K is 6.

The structure and parameters of CNN are as same as that in reference [27]. It includes two convolution layers (Conv1D), where the numbers of convolution cores are 4 and 8 respectively, and the size of the convolution core is 1×5 , and the step is 1. Two max-pooling layers are used with 1×2 pool core size and step 2. After two stacking operations of convolution and max-pooling, the multidimensional features of track irregularities are extracted. The flattened layers compress the multidimensional feature vectors into one-dimensional feature vectors to get global features.

2.2.2. Multiscale Attention. The diagram of multiscale attention is shown in Figure 2. This module consists of two parts in parallel, a multihead self-attention mechanism for capturing global features and a depth-wise convolution mechanism for capturing local features. For the input sequence x , the output y passing through the multiscale attention layer can be expressed as follows:

$$Y = \text{Attention}(X) + \text{Conv}(X). \quad (3)$$

The multihead self-attention mechanism is responsible for learning representations of long-term dependencies. In this module, the input sequence X is projected into three representations, query Q , key K , and value V [20].

$$Q, K, V = \text{Linear}_1(X), \text{Linear}_2(X), \text{Linear}_3(X). \quad (4)$$

Then, the output representation is calculated as follows:

$$\sigma(Q_1, K_1, V_1) = \text{softmax}\left(\frac{Q_1 K_1^T}{\sqrt{d_k}}\right) V_1, \quad (5)$$

$$\text{Attention}(X) = \text{Attention}(Q, K, V) = \sigma(QW^Q, KW^K, VW^V)W^O,$$

where W^Q , W^K , W^V , and W^O are projection parameters, $V = XW^V$.

The convolution module is used to capture the local contextual sequence representations in the same mapping space. Based on the depth-wise convolution operation, the module uses three convolution submodules, which contain multiple cells with different kernel sizes of 1, 3, and 5 to capture the different range features. A gating mechanism is introduced to automatically select the weights of different convolution cells to converge the information of different convolution submodules. The depth-wise convolution will first perform independent convolution on each channel and then perform ordinary convolution. The calculation process of the convolution module can be expressed as follows:

$$\text{Conv}(X) = \sum_{i=1}^n \frac{\exp(\alpha_i)}{\sum_{j=1}^n \exp(\alpha_j)} \text{Conv}_{k_i}(X), \quad (6)$$

where α is the weight coefficient; the output of the convolution cell with kernel size is as follows:

$$\text{Conv}_k(X) = \text{Depth_conv}_k(V)W^{\text{out}}. \quad (7)$$

2.2.3. CA Module. Coordinate attention is added in CNN to focus on the important positions in the spatial dimension L' and the important channels in the channel dimension C' , which have important impacts on the outputs. The input dimension of the coordinate attention is $T \times L' \times C'$. The size after 1D average pooling in the T dimension is $1 \times L' \times C'$. 1×1 convolution is used to reduce and increase the dimension of the channel, where r is the reduction ratio. At last, the coordinate attention weight generates through the sigmoid function.

3. Evaluation and Analysis

3.1. Experimental Setup. For each dataset, the division ratio of training and test data is 7 : 3, which is a common setup in deep learning. In the training process, the loss function is the mean square error of the actual and the predicted vehicle

response, adding the L_1 -norm and L_2 -norm regularization of model parameters.

$$\text{loss} = \frac{1}{T} \sum_{t=1}^T (y^{(t)} - \hat{y}^{(t)})^2 + \lambda_1 \sum |W| + \lambda_2 \sum |W|^2, \quad (8)$$

where T is the sequence length, W is the all-trainable model parameters, λ_1 and λ_2 are regularization coefficients. The

learning rate is 0.001 and the optimizer is the Adam algorithm.

3.2. Evaluation Indices. Four evaluation indices for model performance are employed, including the mean absolute error (MAE), root mean square error (RMSE), their inequality coefficient (TIC), and the correlative coefficient (ρ). The indices are defined as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{M} \sum_{k=1}^M |y_k - \hat{y}_k|, \\ \text{RMSE} &= \sqrt{\frac{1}{M} \sum_{k=1}^M (y_k - \hat{y}_k)^2}, \\ \text{TIC} &= \frac{\sqrt{\frac{1}{M} \sum_{k=1}^M (y_k - \hat{y}_k)^2}}{\sqrt{\frac{1}{M} \sum_{k=1}^M y_k^2} + \sqrt{\frac{1}{M} \sum_{k=1}^M \hat{y}_k^2}}, \\ \rho &= \frac{\sum_{k=1}^M (\hat{y}_k - \bar{\hat{y}})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^M (\hat{y}_k - \bar{\hat{y}})^2} * \sqrt{\sum_{k=1}^M (y_k - \bar{y})^2}}, \left(\bar{y} = \frac{1}{M} \sum_{k=1}^M y_k, \bar{\hat{y}} = \frac{1}{M} \sum_{k=1}^M \hat{y}_k \right), \end{aligned} \quad (9)$$

where M is the length of test data; y_k and \hat{y}_k are the k -th actual and predictive values; \bar{y} and $\bar{\hat{y}}$ are the expectation of the actual value and the predicted value of the model, respectively. MAE and RMSE reflect the absolute accuracy of the prediction, and the smaller their values, the better the performance of the model. TIC and ρ represent relative accuracy indices. Smaller TIC (ranging from 0 to 1) means higher accuracy. ρ ranges from -1 to 1 , and the greater the absolute value, the higher the accuracy.

3.3. Comparison with Other Models. To evaluate the proposed model CA-CNN-MUSE, we build multiple models and show the test results of the inspection-simulation dataset of line 1, as shown in Table 1. There, the LSTM model has two stacking LSTM layers, which all have 64 hidden neuron nodes. The size of the input sequence is $B \times 480 \times C$, and that of the output is $B \times 480 \times k$, $B = 32$ is the batch-size. CNN-LSTM network has the same CNN module as that of the proposed CA-CNN-MUSE and two stacking LSTM layers. CA-CNN-LSTM is the network when adding the CA module into the CNN-LSTM. CNN-MUSE network replaces the LSTM with MUSE.

For comparing the performance of different models, the accuracy indicators of each vehicle response are averaged to evaluate the model accuracy as a whole and more indices are used, including params, FLOPs, and inference time on the test set. The following can be seen from the table:

- (1) When CA is added to the CNN-LSTM model, all the accuracy indices of RMSE, MAE, ρ , and TIC improve. Replacing LSTM with MUSE in CNN-LSTM, all the accuracy indices also become better. Using CA and MUSE at the same time, CA-CNN-MUSE reaches the best RMSE, MAE, ρ , and TIC.
- (2) Replacing LSTM with MUSE in CNN-LSTM, although the params and FLOPs increase, the inference time is reduced. The reason is that the multihead attention module in muse can be computed in parallel mode. When the CA module is added to CNN-MUSE, the inference time only increases by 0.04 s.

To sum up, the evaluation indices of CA-CNN-MUSE are superior to that of other models.

3.3.1. Results on Different Lines. Based on the inspection-simulation dataset, we investigate the model estimation performance on different lines, as shown in Table 2. As can be seen, compared with CNN-LSTM, CA-CNN-MUSE also has better performance online 2 and line 3.

3.4. Results of Specific Vehicle Responses

3.4.1. Results on the Inspection-Simulation Dataset. The accuracy indices of 14 vehicle responses of line 1 in the

TABLE 1: Performance of different models.

Model	RMSE	MAE	ρ	TIC	params	FLOPs (MB)	Time (s)
LSTM	1361.8107	933.4527	0.7492	0.1129	119.054 KB	57.876	2.23
CNN-LSTM	1177.3689	818.7498	0.8062	0.1031	364.0 K	178.381	5.67
CA-CNN-LSTM	1116.7266	762.0530	0.8287	0.1023	364.058 K	179.155	5.95
CNN-MUSE	1062.8726	737.1695	0.8418	0.0959	1.222 M	613.581	4.67
CA-CNN-MUSE	1019.6666	705.9139	0.8531	0.0925	1.226 M	615.993	4.71

TABLE 2: Performance on different lines.

Line	Model	RMSE	MAE	ρ	TIC
Line 2	CNN-LSTM	1074.7724	746.7088	0.7920	0.1083
	CA-CNN-MUSE	916.3021	630.4018	0.8570	0.1010
Line 3	CNN-LSTM	1711.3334	1240.1423	0.7147	0.1609
	CA-CNN-MUSE	1486.9768	1035.1907	0.8079	0.1585

TABLE 3: Accuracy indices of the vehicle response of line 1 in the inspection-simulation dataset.

Item	Output	RMSE	MAE	ρ	TIC
Vertical forces	Left force of axle 1	1745.1141	1211.5344	0.8810	0.0155
	Left force of axle 2	1796.8741	1239.4564	0.8825	0.0160
	Left force of axle 3	1770.2314	1238.6849	0.8775	0.0158
	Left force of axle 4	1836.0664	1275.5680	0.8773	0.0163
	Right force of axle 1	1746.3233	1204.2282	0.8819	0.0156
	Right force of axle 2	1783.9973	1221.8113	0.8831	0.0159
	Right force of axle 3	1775.3752	1232.1890	0.8770	0.0158
	Right force of axle 4	1821.2450	1259.2483	0.8783	0.0162
Wheel unloading rate	Axle 1	0.0196	0.0140	0.7506	0.2237
	Axle 2	0.0197	0.0141	0.7380	0.2309
	Axle 3	0.0197	0.0142	0.7468	0.2261
	Axle 4	0.0199	0.0144	0.7330	0.2331
Vertical car body acceleration	Front	0.0132	0.0085	0.9689	0.1258
	Rear	0.0135	0.0087	0.9678	0.1283

inspection-simulation dataset are shown in Table 3. The following can be seen:

- (1) CA-CNN-MUSE model can effectively estimate vertical wheel-rail forces, wheel unloading rate, and vertical car body accelerations, and the corresponding ρ are 0.87, 0.73, and 0.96, respectively. Among these, the estimation accuracy of vertical car body acceleration is the highest.
- (2) In vertical wheel-rail forces, the right vertical force of axle 1 has the highest estimation accuracy. The estimation accuracy of the wheel unloading rate of the four wheel sets is the same.

3.4.2. Results on the Inspection Dataset. We also train and test the proposed CA-CNN-MUSE on the track inspection dataset. The 11 track irregularity items and the vehicle running speed are used as the input of the network. We first remove the wavelength components below 2 m from the signals using wavelet decomposition and reconstruction, then feed them into the network to estimate the vehicle responses. The following results are obtained on the test set:

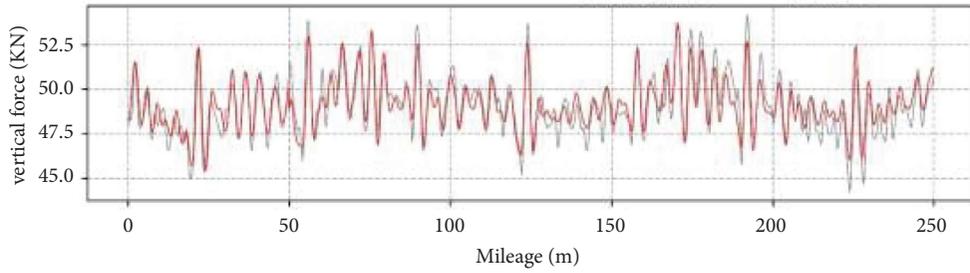
Table 4 summarizes the accuracy indices of the CA-CNN-MUSE model. As can be seen, the CA-CNN-MUSE model

TABLE 4: The accuracy index of various vehicle response parameters in measured data.

Item	RMSE	MAE	ρ	TIC
Left vertical force	1259.4704	936.4605	0.7313	0.0096
Left lateral force	740.8014	581.6191	0.4938	0.3606
Right vertical force	1214.3663	908.8109	0.7258	0.0123
Right lateral force	658.9214	475.4641	0.5138	0.2201
Vertical car body acceleration	0.0076	0.0061	0.8993	0.2433
Lateral car body acceleration	0.0050	0.0040	0.5358	0.5508

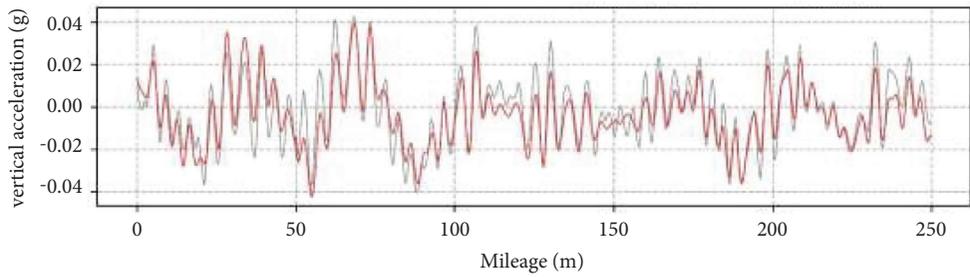
can effectively estimate vertical wheel-rail force and acceleration. The estimation accuracy of vertical vehicle response is good. Compared with the vertical responses, the estimation of the lateral vehicle responses is not good enough.

To further analyze the model performance on different wavelengths, the waveforms of the right vertical force and vertical car body acceleration are illustrated in Figures 3(a)-3(b), the waveforms of the right lateral forces and lateral car body acceleration are illustrated in Figures 3(c)-3(d), where the length of the track section is 250 m. Similarly, the proposed model succeeds in predicting the vertical forces and vertical car body acceleration. Meanwhile, the main features of the lateral



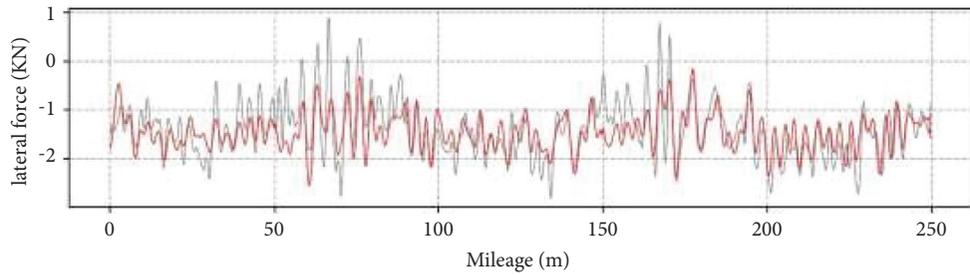
— Inspection data
— CA-CNN-MUSE

(a)



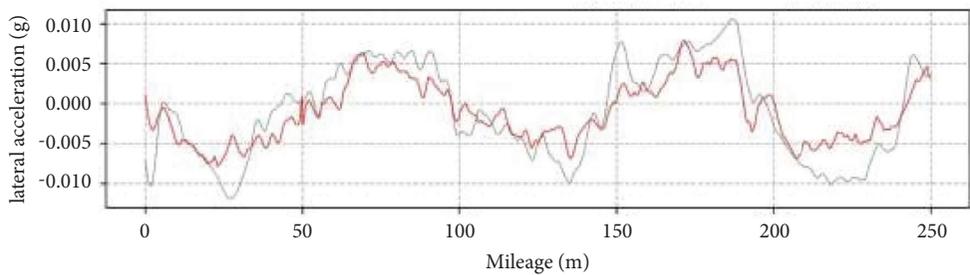
— Inspection data
— CA-CNN-MUSE

(b)



— Inspection data
— CA-CNN-MUSE

(c)



— Inspection data
— CA-CNN-MUSE

(d)

FIGURE 3: Waveforms of vehicle responses. (a) Right vertical force, (b) vertical car body acceleration, (c) right lateral force, and (d) lateral car body acceleration.

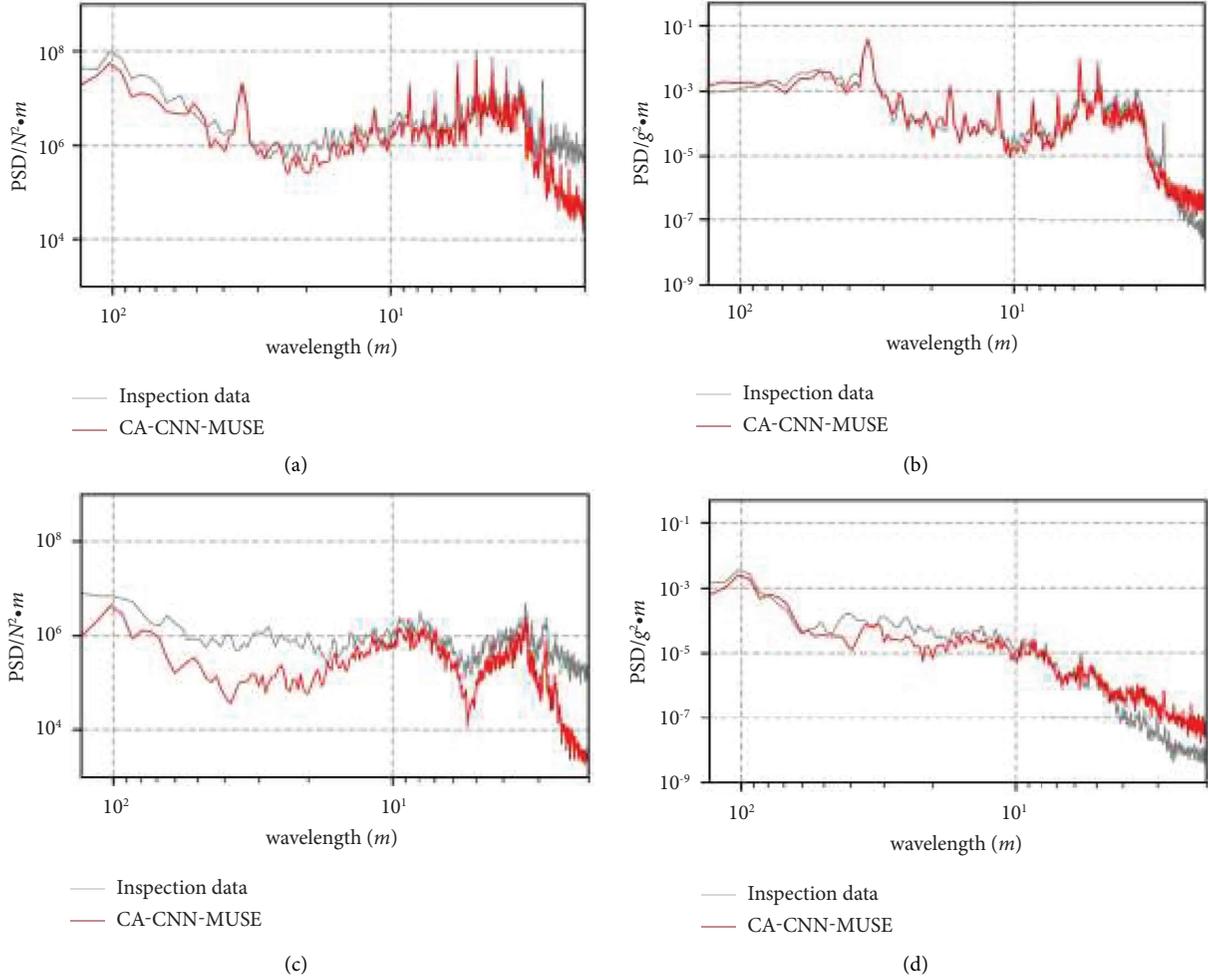


FIGURE 4: PSD of vehicle responses. (a) Right vertical force, (b) vertical car body acceleration, (c) right lateral force, and (d) lateral car body acceleration.

vehicle responses can also be captured by the proposed method, yet the results are not as good as the vertical ones.

The power spectral density (PSD) of the right vertical forces and car body acceleration are illustrated in Figures 4(a)-4(b), and the power spectral density of the right lateral forces and car body acceleration are illustrated in Figures 4(c)-4(d). As can be seen, the predicted PSD of right vertical forces and car body acceleration has good fitting effect at the wavelength above 3 m.

4. Conclusion

To relate track geometry irregularities with vehicle responses to improve track quality assessment standards and track maintenance work, this paper proposed a CA-CNN-MUSE model to predict vehicle response. We use a multiscale self-attention mechanism to replace the LSTM structure, which is dominant in this kind of task. Besides, we introduced the light weight coordinate attention mechanism into CNN to focus on important interchannel relationships and important mileage information. The results of this paper show the following:

- (1) CA-CNN-MUSE has higher prediction accuracy than LSTM and CNN-LSTM and a faster inference speed than CNN-LSTM. The estimated waveforms and PSDs of vertical wheel-rail forces and car body acceleration by CA-CNN-MUSE agree well with the actual items;
- (2) CA-CNN-MUSE is applicable to the multibody simulation model of a vehicle system and the measured data of actual high-speed lines.

The proposed model succeeds in representing the vertical wheel-rail force and car body acceleration, yet the estimations on the lateral wheel-rail force and car body acceleration are not as good as the vertical results. In fact, the estimation of lateral wheel-rail force and car body acceleration is indeed a difficult problem, and we will continue to investigate the challenge.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (No. 52278465) and the Scientific Research Plan of China Railway (No. P2021T013).

References

- [1] C. S. Bonaventura, J. W. Palese, and A. M. Zarembski, "Intelligent system for real-time prediction of railway vehicle response to the interaction with track geometry," in *Proceedings of the 2000 ASME/IEEE Joint Railroad Conference (Cat. No. 00CH37110)*, pp. 31–45, IEEE, Newark, NJ, USA, April 2000.
- [2] W. Zhai, K. Wang, and C. Cai, "Fundamentals of vehicle-track coupled dynamics," *Vehicle System Dynamics*, vol. 47, no. 11, pp. 1349–1376, 2009.
- [3] I. I. Y. Choi, J. H. Um, J. S. Lee, and H. H. Choi, "The influence of track irregularities on the running behavior of high-speed trains," *Proceedings of the Institution of Mechanical Engineers - Part F: Journal of Rail and Rapid Transit*, vol. 227, no. 1, pp. 94–102, 2013.
- [4] L. Ling, Y. Deng, Q. Guan, and X. Jin, "Effect of track irregularities on the dynamic behavior of a tram vehicle," *Journal of Advances in Vehicle Engineering*, vol. 3, no. 1, pp. 29–39, 2017.
- [5] J. Sun, E. Meli, W. Cai et al., "A signal analysis based hunting instability detection methodology for high-speed railway vehicles," *Vehicle System Dynamics*, vol. 59, no. 10, pp. 1461–1483, 2021.
- [6] R. Gadhavane and N. S. Vyas, "Rail-wheel contact forces and track irregularity estimation from on-board accelerometer data," *Vehicle System Dynamics*, vol. 60, no. 6, pp. 2145–2166, 2022.
- [7] A. Furukawa, "A method to predict vertical vehicle motion caused by track irregularities," *Proceedings of the Institution of Mechanical Engineers - Part F: Journal of Rail and Rapid Transit*, vol. 231, no. 4, pp. 431–443, 2017.
- [8] M. X. D. Li, E. G. Berggren, M. Berg, and I. Persson, "Assessing track geometry quality based on wavelength spectra and track-vehicle dynamic interaction," *Vehicle System Dynamics*, vol. 46, no. 1, pp. 261–276, 2008.
- [9] B. Lubner, "Railway track quality assessment method based on vehicle system identification," *E & I: Elektrotechnik und Informationstechnik*, vol. 126, no. 5, pp. 180–185, 2009.
- [10] S. Nefti and M. Oussalah, "A neural network approach for railway safety prediction," in *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, pp. 3915–3920, IEEE, The Hague, Netherlands, March 2004.
- [11] D. Li, A. Meddah, K. Hass, and S. Kalay, "Relating track geometry to vehicle performance using neural network approach," *Proceedings of the Institution of Mechanical Engineers - Part F: Journal of Rail and Rapid Transit*, vol. 220, no. 3, pp. 273–281, 2006.
- [12] G. M. Shafiullah, A. B. M. S. Ali, A. Thompson, and P. J. Wolfs, "Predicting vertical acceleration of railway wagons using regression algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 290–299, 2010.
- [13] S. Zheng, Q. Zhong, X. Chai, X. Chen, and L. Peng, "A novel prediction model for car body vibration acceleration based on correlation analysis and neural networks," *Journal of Advanced Transportation*, vol. 2018, no. 8, Article ID 1752070, 13 pages, 2018.
- [14] R. J. Williams and D. Zipser, "Experimental analysis of the real-time recurrent learning algorithm," *Connection Science*, vol. 1, no. 1, pp. 87–111, 1989.
- [15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [16] K. Cho, B. Van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014.
- [17] H. Li, T. Wang, and G. Wu, "Dynamic response prediction of vehicle-bridge interaction system using feedforward neural network and deep long short-term memory network," *Structures*, vol. 34, pp. 2415–2431, 2021.
- [18] S. Ma, L. Gao, X. Liu, and J. Lin, "Deep learning for track quality evaluation of high-speed railway based on vehicle-body vibration prediction," *IEEE Access*, vol. 7, pp. 185099–185107, 2019.
- [19] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014.
- [20] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [21] G. Zhao, X. Sun, J. Xu, Z. Zhang, and L. Luo, "Muse: parallel multi-scale attention for sequence to sequence learning," 2019.
- [22] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, IEEE, Salt Lake City, UT, USA, June 2018.
- [23] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: efficient channel attention for deep convolutional neural networks," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11531–11539, IEEE, Seattle, WA, USA, June 2020.
- [24] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, ECCV, Munich, Germany, October 2018.
- [25] H. Liu, F. Liu, X. Fan, and D. Huang, "Polarized self-attention: towards high-quality pixel-wise regression," 2021.
- [26] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou, "Rotate to attend: convolutional triplet attention module," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3139–3148, IEEE, Waikoloa, HI, USA, January 2021.
- [27] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13713–13722, IEEE, Nashville, TN, USA, June 2021.