Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/42/20240669

Road Pothole Detection Model Based on Local Attention Resnet18-CNN-LSTM

Jiahe Guang[1,3,*], Xingrui He[1,4], Zeng Li[2,5], Shiyu He[2,6]

[1]School of Electronics and Information Engineering, Liaoning Technical University, Xingcheng District, Huludao City, Liaoning Province, China
[2]School of Safety Science and Engineering, Liaoning Technical University, Xingcheng District, Huludao City, Liaoning Province, China
[*]corresponding author

[3]guangjiahe@qq.com
[4]1014699735@qq.com
[5]2513598481@qq.com
[6]3128124522@qq.com

Abstract. In response to the low detection accuracy and slow speed of existing road pothole detection methods, a road pothole classification detection model based on local attention Resnet18-CNN-LSTM (Long Short-Term Memory network) is proposed. On the basis of Resnet18, a local attention mechanism and a CNN-LSTM combined model are added to propose a road pothole detection model based on local attention Resnet18-CNN-LSTM. The local attention mechanism is used to accurately extract specific target feature values, CNN is used to extract the spatial features of the input data, and LSTM enhances the detection model's extraction of sequential features and performs classification, thereby improving the accuracy of the road and pothole model. Experimental results show that on the training set, the accuracy of the local attention mechanism-based ResNet18-CNN-LSTM model reached 99.2188%, which is an increase of 0.7813% and 2.3438% compared to the ResNet34-CNN-LSTM and ResNet50-CNN-LSTM models under the same conditions, respectively. On the test set, the model's accuracy was 93.4437%, an increase of 0.5437% and 1.9867% compared to the ResNet34-CNN-LSTM and ResNet50-CNN-LSTM models, respectively. After dealing with overfitting issues through early stopping, the detection accuracy of this model has significantly improved compared to the detection models based on ResNet34 and ResNet50, with an increase of 1.2% and 1.49% respectively. The model shows faster processing speed in identification time, effectively retains the correlation and sequence features of the data, overcomes the problem of gradient disappearance in deep networks, and thereby enhances the extraction capability of local target features of road pothole images. The above results indicate that the local attention mechanism-based ResNet18-CNN-LSTM model shows superior performance in road pothole detection.

Keywords: Classification Detection, Resnet, CNN, Long Short-Term Memory Network, Road Pothole.

DOI: 10.54254/2753-8818/42/20240669

1. Introduction

Road potholes refer to the uneven depressions on the road surface. If detected improperly during vehicle travel, they can not only cause damage to the vehicle but also have adverse effects on the driver, leading to traffic accidents. Therefore, it is crucial to make highly accurate and short-recognition-time judgments on road potholes. In recent years, many scholars have conducted extensive research on road pothole detection. Gao Min [1] et al. proposed an improved YOLOv7 road pothole detection algorithm, which can significantly enhance the model training speed and detection accuracy, with an increase of 6.18% in the average precision mean compared to the original YOLOv7; Bai Rui [2] et al. proposed an improved YOLOv5s algorithm, which has significantly improved the accuracy and speed of road pothole detection compared to traditional detection algorithms; Chen Peng [3] et al. used the inception_v1 algorithm architecture to train the data, obtained a multi-scenario classification model, and detected the specific location of potholes through the pothole detection model, which performed better than the direct mixing of pothole model detection results; Zhang Zirui [4] et al. proposed a design for pothole road detection based on PQCR-PSL sensors, combining a piezoelectric quartz crystal resonator (PQCR) with a planar spiral inductor (PSL) in series as a sensor probe fixed on the vehicle, reflecting road potholes by sensing the different vibrations generated when the vehicle travels on smooth and potholed roads. Although there have been many studies in the field of road pothole detection, the detection methods, accuracy, and time still need to be improved.

This paper proposes a road pothole detection method based on local attention Resnet18-CNN-LSTM. Many scholars have already conducted research on model methods in various fields. Cai Shuyu [5] et al. proposed an aviation engine performance anomaly detection method based on ResNet-LSTM in the aviation field, which significantly improved the anomaly detection accuracy compared to the anomaly detection models based on ResNet18, ResNet34, and ResNet50 networks, with increases of 11.81%, 9.45%, and 3.78% respectively, and the loss values were also significantly reduced. It has a good effect on the abnormal detection of aviation engines; Tang Qingwei [6] et al. constructed a wind power prediction model based on the CNN-LSTM hybrid neural network in the wind power field, effectively combining the clustering results of wind farms with the CNN-LSTM prediction model, constructing a transfer prediction model to achieve the prediction of wind farm power generation during the planning phase.

In response to the accuracy and time issues of road pothole detection, this paper combines the local attention mechanism, Resnet model, and CNN-LSTM model to propose a road pothole detection method based on local attention Resnet18-CNN-LSTM. It is tested with real performance data and compared with the Resnet34 and Resnet50 models under the same conditions to verify the effectiveness of the model.

2. Road Pothole Detection Model Based on Local Attention Resnet18-CNN-LSTM

2.1. Resnet18 Model

The deep residual network (Resnet) can well solve the problem of gradient disappearance and make its features better expressed. The key to the Resnet network is the residual unit in its network structure. Resnet18 is a deep residual network (Residual Network) model proposed by Microsoft Research, which has 18 layers of depth based on the Resnet architecture, including multiple residual blocks, each consisting of two convolutional layers (including skip connections), improving the training efficiency and convergence speed of the network, while reducing the risk of overfitting.

2.2. Local Attention Mechanism

The local attention mechanism is a local attention mechanism that considers only the local subsequences in the input sequence when calculating the attention weights, rather than globally considering all positions. This approach can reduce computational complexity and make the model more focused on the information related to the current position in the input sequence. By calculating

DOI: 10.54254/2753-8818/42/20240669

$$a^{(L)}{}_{ij} = \frac{exp\{\beta \cdot cos(Wh_i, Wh_j)\}}{\sum_{j \in N_i} exp\{\beta \cdot cos(Wh_i, Wh_j)\}} \quad (1)$$

In the formula, cos（·） represents the standard deviation, which is used to calculate the cosine similarity to aggregate information from the node neighborhood. Equation (2) can be expressed as:

$$h^{(L)}{}_i{}' = \sigma\Big(\sum_{j \in N(vi)} a^{(L)}{}_{ij} W h_j\Big) \quad (2)$$

2.3. CNN-LSTM Model

Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Network (RNN) that can solve the problems of gradient vanishing and gradient explosion during the training of long sequences. It is a typical neural network with a structure that includes an input layer, hidden layer, and output layer. The input to the hidden layer not only consists of the inputs from the input layer but also the output from the hidden layer of the previous moment. This structure allows the RNN to remember not only the current moment's information but also the previous information, which is used as input for the current computation. It successfully addresses the shortcomings of the original recurrent neural networks and has excellent performance in fitting nonlinear sequence data. The principle structure is shown in Figure1.

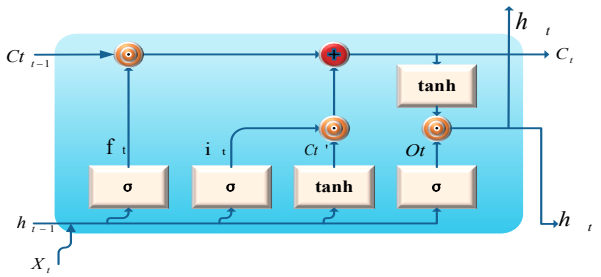


Figure 1. LSTM Model Principle Structure Diagram

LSTM controls the discarding or addition of information through "gates," achieving the functions of forgetting or remembering. A "gate" is a structure that allows selective passage of information, composed of a sigmoid function and a point multiplication operation. The output value of the sigmoid function is within the [0,1] interval, where 0 represents complete discarding, and 1 represents complete passage. An LSTM unit has three such gates: the forget gate, the input gate, and the output gate.

2.4. Road Pothole Detection Model Based on Local Attention Resnet18-CNN-LSTM

The road pothole detection model based on local attention Resnet18-CNN-LSTM consists of four parts: the image processing module, the local feature processing module, the sequence extraction module, and the output module. The image processing module is based on the Resnet18 model as the core and is equipped with image rotation, size determination, clarity adjustment, and channel adjustment to meet the requirements of the Resnet18 model and better extract features. The local feature processing module adds a local attention mechanism to the Resnet module to deeply explore target features. The sequence extraction module uses LSTM as a post-treatment to transform data to meet the requirements of the LSTM series. The output module integrates the network output through the LSTM model with a CNN deep neural network and uses a Softmax classifier to identify the output information, achieving the classification and recognition of road potholes. The composition of each module is as follows:

Image Processing Module includes:

- Resize: Adjust the size of the input image to 224x224 pixels, which is the expected input size for the ResNet18 model.

- Random Horizontal Flip: Randomly flip the image horizontally along the horizontal axis with a probability of 0.5, increasing the diversity of data during model training.

DOI: 10.54254/2753-8818/42/20240669

- Random Rotation: Randomly rotate the image by a certain angle (here ±10 degrees) with a uniform distribution, also used to increase data diversity.
- To Tensor: Convert the PIL image format to a tensor format that PyTorch can process.
- Normalize: Normalize each channel of the image so that the mean and standard deviation of each channel are 0.5, which helps with the stability of model training.

Output Module

The output module consists of a Fully connected layer and a Softmax layer. The Fully connected layer is a fully connected layer, quantified as: (The original text was cut off and does not specify the details of the quantification.)

$$f(x) = \sum_{i=1}^{n} \omega_i x_i + b \tag{3}$$

In the formula: x_i and ω_i represent the input value and weight of the ith input, respectively; b is the bias.

The flowchart of the road pothole detection model based on local attention Resnet18-CNN-LSTM is shown in Figure 2.

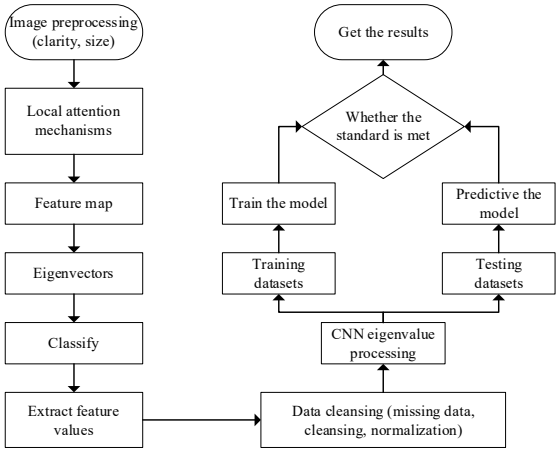


Figure 2. Flowchart of the Road Pothole Detection Model Based on Local Attention Resnet18-CNN-LSTM

2.5. Model Evaluation Metrics

To evaluate the road pothole identification detection model, this paper adopts accuracy as the main evaluation metric for the model, and loss as the auxiliary evaluation metric.

The anomaly detection rate is:

$$accuracy = \frac{x_i}{N_i} \tag{4}$$

where x_i is the number of correct samples in the ith detection, and N_i is the total number of samples in the ith detection.

The loss value is used to calculate the degree of difference between the predicted value and the actual value, using the binary cross-entropy function to calculate the model's loss value:

$$Loss = -\sum_{i=1}^{n} (y_i \, log(\hat{y}_i) + (1 - y_i) \, log(1 - \hat{y}_i)) \tag{5}$$

where: $\hat{y}_i$ is the probability that the ith sample is a correct sample, and it is the result predicted by the model.

DOI: 10.54254/2753-8818/42/20240669

3. Example Verification and Analysis

3.1. Road Pothole Image Dataset

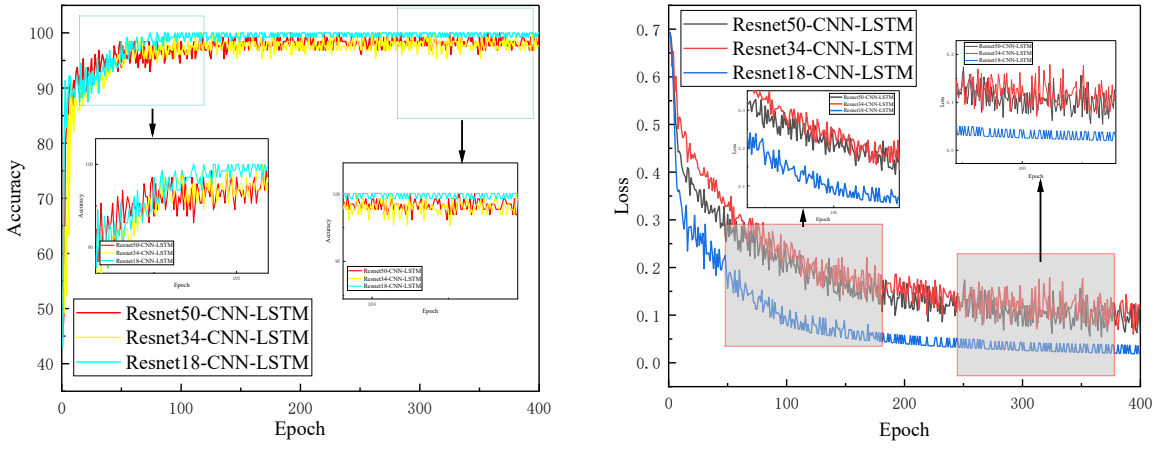
To verify the road pothole detection method based on local attention Resnet18-CNN-LSTM, images from a U.S. database were selected, collecting 501 images, including 252 normal road images and 249 road pothole images. The ratio of the training set to the test set is set to 70%:30%. First, the image size is adjusted to 224x224 pixels through image processing and normalized. Then, through the detection method model's feature value collection, the image dataset is transformed into a digital dataset, with 1000 features collected for each image. See Table 1 for specific information. (Note: Table 1 is referenced but not provided in the original text.)

Table 1. Feature Value Collection Information of the Resnet18 Model Based on Local Attention Mechanism

Type	Feature1	Feature2	Feature3	Feature4	...	Feature1000
1	1.0021	-1.7528	-5.1128	-1.7239		1.3091
2	-0.6458	-1.1740	-1.1802	-1.0693		2.5408
1	1.6875	-2.1638	-1.4650	-0.3545		1.6950
2	0.7401	0.0154	-0.5042	-1.5840		-1.7595
1	1.6875	-2.1638	-1.4650	-0.3545		-3.2115
2	-0.9452	1.9716	0.6680	1.9264		-0.2219
1	-1.0628	-1.7537	-2.8963	-0.5570		-0.5175
2	-2.4406	-1.7855	-0.6365	-0.0874		3.1530
1	-1.0218	0.9023	-1.8376	-2.6181		2.0532
2	-0.2431	-0.9385	-0.7012	0.6339		0.7437

3.2. Validation and Analysis

Accuracy and loss values on the training set. The comparison of accuracy and loss values among the Resnet18-CNN-LSTM, Resnet34-CNN-LSTM, and Resnet50-CNN-LSTM models in road pothole detection and recognition is shown in Figure 3.



(a) (b)

Figure 3. (a) Comparison of Accuracy Rate Changes. (b) Comparison of Loss Value Changes

As can be seen from Figure 5(a), in the early stages of training, the accuracy rates of all three models increased rapidly. As the number of training epochs increased, the rate of convergence slowed down, the accuracy fluctuated within a certain range, then the fluctuations decreased and gradually stabilized,

DOI: 10.54254/2753-8818/42/20240669

eventually converging. From Figure 5(b), in the early stages of training, the loss values of all three models decreased quickly. As the number of training sessions increased, the rate of convergence began to slow down, the loss values fluctuated within a certain range, and finally converged. By comparing the three models, it can be observed that the ResNet18-CNN-LSTM road pothole detection model based on the local attention mechanism has a higher accuracy rate on the training set than other anomaly detection models, and its loss value is lower, indicating the best model performance. This result fully demonstrates that the Resnet18 model performs better in road pothole detection and recognition under the same conditions compared to other Resnet models. It not only improves the model's accuracy rate but also slightly increases the speed of convergence while reducing the loss value. This helps to enhance the application effect of the road pothole detection model on the test set and in practical scenarios.

The training results of the ResNet18-CNN-LSTM, Resnet34-CNN-LSTM, and Resnet50-CNN-LSTM models based on local attention on the training set are presented in Table 2. (Note: Table 2 is referenced but not provided in the original text.)

Table 2. Training Results of the Three Models on the Training Set

Model	Accuracy/%	Loss Value	Training Time/s
Resnet50-CNN-LSTM	**96.875**	**0.1739**	**292.24**
Resnet34-CNN-LSTM	**98.4375**	**0.1601**	**292.75**
Resnet18-CNN-LSTM	**99.2188**	**0.0315**	**273.61**

As can be seen from Figures 5, due to the excessive number of iterations of the algorithm and the small number of samples in the training set, the model has exhibited the problem of overfitting. To address this issue, we have optimized the model against overfitting by applying EarlyStopping.

3.3. EarlyStopping

In most introductory papers on the training of supervised neural networks, as the number of training rounds increases, the loss value of the training set first decreases and then tends to flatten. In contrast, the loss value of the validation set first decreases and then increases. Our goal in training the model is to obtain a model with good generalization performance, that is, a smaller error (loss value) on the validation set. We select the training round (epoch) at which the minimum validation set error is achieved, terminate the training in advance, which not only saves the best model but also saves training time.

3.4. The First Type of Stopping Criterion

The value of Eopt(t) is defined as the minimum validation set error obtained before iteration t.

GL(t) is called the generalization loss, which is represented as the increase in validation error relative to the minimum value (in percentage terms).

A high generalization loss is one of the candidate criteria for stopping training because it directly indicates overfitting (when the loss value on the validation set begins to rise, i.e., the generalization performance begins to decline). We choose an appropriate threshold and stop the training when the generalization loss exceeds this threshold.

Table 3. EarlyStopping training results on the training set.

Model	Accuracy/%	Loss Value	Training Time/s
Resnet50-CNN-LSTM	**90.821**	**04921**	**16.93**
Resnet34-CNN-LSTM	**91.111**	**0.2811**	**14.64**
Resnet18-CNN-LSTM	**92.311**	**0.1991**	**11.23**

Note: The specific results are not provided in the original text.

DOI: 10.54254/2753-8818/42/20240669

Table 4. EarlyStopping Test Set Training Results

Model	Accuracy/%	Loss Value	Training Time/s
Resnet50-CNN-LSTM	**91.457**	**0.3821**	**18.83**
Resnet34-CNN-LSTM	**92.9**	**0.2346**	**16.77**
Resnet18-CNN-LSTM	**93.4437**	**0.1583**	**12.39**

3.5. Model Comparison Analysis

Shallow networks may already be sufficient to capture the complexity in the data. In such cases, increasing the network depth may not lead to significant performance improvements and could potentially result in a decline in performance. Therefore, it is concluded that for the road pothole problem, a shallower network is adequate to meet the classification and recognition requirements. The test set accuracy of the Resnet18-CNN-LSTM model based on local attention is 92.311%, as shown in Figure4.

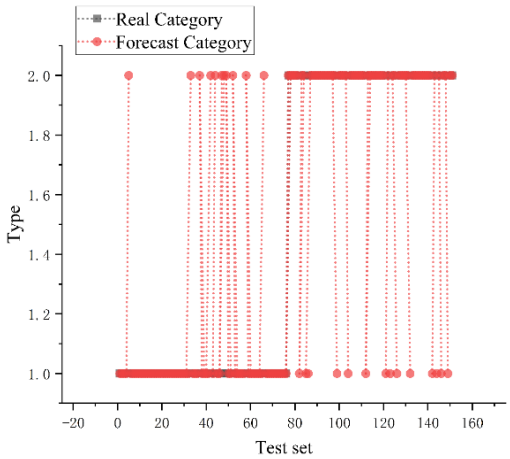


Figure 4. Comparison of Test Set Performance for the Resnet18-CNN-LSTM Model Based on Local Attention

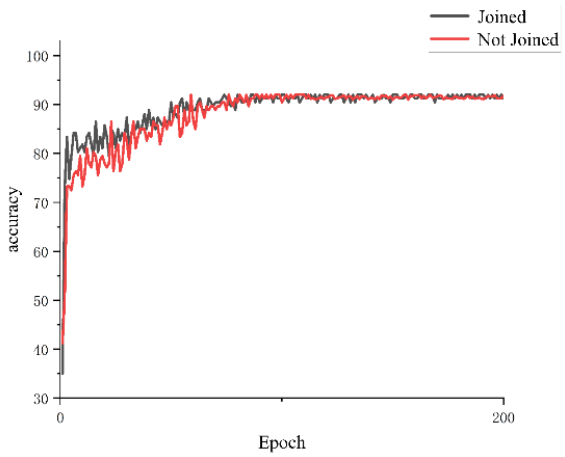


Figure 5. Accuracy Comparison Chart of the Resnet18-CNN-LSTM Model After Removing the Local Attention Mechanism

Under the premise of the same model parameters, using the same image dataset, after removing the local attention mechanism from the Resnet18-CNN-LSTM model, the accuracy comparison chart of the two models is shown in Figure 5. As can be seen from the figure, during the training process of the model with the addition, the convergence speed is almost the same as it increases with the epoch, but the accuracy is significantly higher than when it is not added. This indicates that the addition of the local attention mechanism is beneficial to the optimization of this model.

4. Conclusion

(1) A road pothole detection model based on local attention Resnet18-CNN-LSTM was established. Compared with the detection models based on the ResNet34 and ResNet50 networks under the EarlyStopping method, the detection accuracy was significantly improved, increasing by 1.2% and 1.49%, respectively. At the same time, the recognition time was shorter, the model fully retained the correlation and sequence features of the data, eliminated the problem of gradient disappearance in deep networks, and improved the extraction capability of local target features in road pothole images.

(2) The local attention mechanism was introduced, proving that this mechanism has good performance in extracting feature values for images with obvious local features such as road potholes, compared to overall extraction.

DOI: 10.54254/2753-8818/42/20240669

References

[1] Gao Min, Li Yuan. Road surface defect detection based on YOLOv7-CA-BiFPN [J/OL]. Computer Measurement & Control: 1-12 [2024-03-11]. http://kns.cnki.net/kcms/detail/11.4762.TP.20240119.1708.020.html.

[2] Bai Rui, Xu Yang, Wang Bin, et al. Road pothole detection algorithm based on improved YOLOv5s [J]. Computer & Modernization, 2023(06): 69-75.

[3] Chen Peng, Ying Jun. Multi-scene road pothole image detection based on convolutional neural networks [J]. Journal of Shanghai Normal University (Natural Science Edition), 2020, 49(01): 96-101.

[4] Zhang Zirui, Chen Xiangdong, Ding Xing. Design of road surface pothole detection based on PQCR-PSL sensors [J]. Electronic Design Engineering, 2019, 27(19): 90-94. DOI: 10.14022/j.cnki.dzsjgc.2019.19.020.

[5] Cai Shuyu, Yin Hang, Shi Tao, et al. Aero-engine performance anomaly detection method based on ResNet-LSTM [J]. Aero-Engine, 2024, 50(01): 135-142. DOI: 10.13477/j.cnki.aeroengine.2024.01.019.

[6] Tang Qingwei, Xiang Yue, Dai Jia Kun, et al. Migration prediction method for wind farm power generation based on CNN–LSTM [J/OL]. Engineering Science and Technology, 1-9 [2024-03-11]. https://doi.org/10.15961/j.jsuese.202201165.