

PAPER • OPEN ACCESS

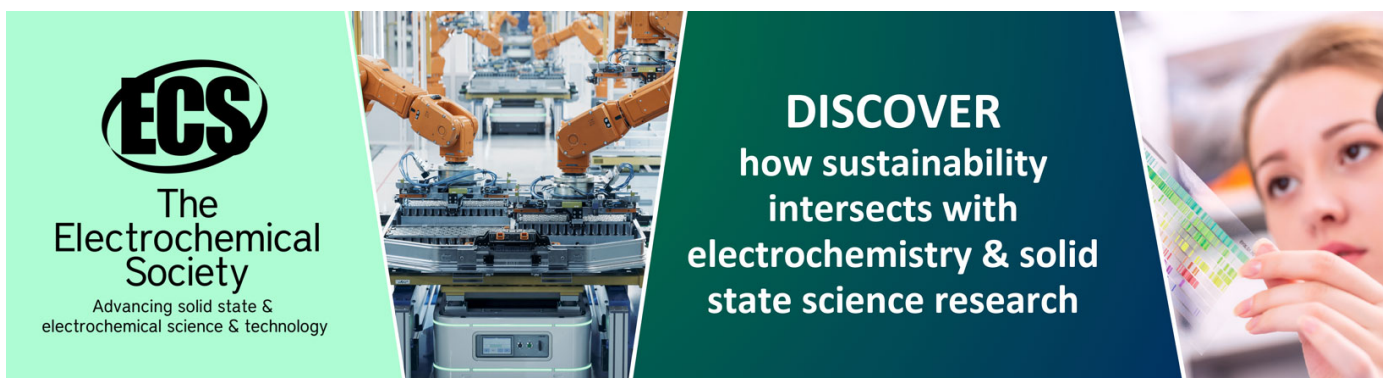
Comparison of Convolutional Sparse Coding Network and Convolutional Neural Network for Pavement Crack Classification: A Validation Study

To cite this article: Haitong Tang *et al* 2020 *J. Phys.: Conf. Ser.* **1682** 012016

View the [article online](#) for updates and enhancements.

You may also like

- [Terahertz wave imaging using time domain iterative shrink threshold method](#)
Tatsuya Kaneoya, Ryo Toh and Koji Suizu
- [Eigenvalue-free iterative shrinkage-thresholding algorithm for solving the linear inverse problems](#)
Can Tong, Yueyang Teng, Yudong Yao et al.
- [A projected gradient method for \$\ell_2\$ sparsity regularization](#)
Liang Ding and Weimin Han



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Comparison of Convolutional Sparse Coding Network and Convolutional Neural Network for Pavement Crack Classification: A Validation Study

Haitong Tang^{1,2}, Jun Shi¹, Xia Lu², Zhichao Yin¹, Lixue Huang¹, Dongbao Jia¹, and Nizhuan Wang^{1*}

¹Artificial Intelligence & Neuro-Informatics Engineering (ARINE) Laboratory of School of Computer Engineering, Jiangsu Ocean University, Lianyungang, Jiangsu, 222023, China

² School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang, Jiangsu, 222023, China

*2018000013@jou.edu.cn

*Corresponding author's e-mail: wangnizhuan1120@gmail.com

Abstract. Accurate and effective identification of pavement cracks can provide a reference for pavement performance evaluation and prediction. Inspired by recent outstanding performance of convolutional sparse coding theory in various research fields, we envision whether convolutional sparse coding can be beneficial for crack detection in complex environments. Therefore, based on the multi-layer convolutional sparse coding (ML-CSC) model, this paper combines multi-layer iterative soft threshold algorithm (ML-ISTA) and convolutional neural network (CNN) into recurrent neural networks (RNN) to identify crack images across the simulated experiments. And in different noise environments, different training cycles and different epochs, the ML-ISTA is compared with traditional CNN and the layered basis pursuit (LBP), which is another popular algorithm for ML-CSC. Experimental results showed that under the different training conditions with the same parameter setting, the stability and accuracy of the ML-ISTA is better than CNN and LBP. The ML-ISTA can achieve crack identification accuracy of 99.36% efficiently, which demonstrates the effectiveness of convolutional sparse coding in crack detection.

1. Introduction

With the continuous development and improvement of China's infrastructure construction, engineering construction has achieved remarkable results in the specific implementation of roads and bridges. However, in the actual operation of project construction, due to the impact of aging building materials, natural disasters and human-made destruction, these engineering structures show different degrees of damage, and the cracks on the surface of the concrete structure are the most prominent aspects. These cracks bring substantial economic losses and accident risks. Therefore, it is of considerable significance to classify the pavement cracks accurately and efficiently.

Crack detection task can be evaluated as a classification problem of crack presence in essence. Two types of methodological approaches are observed in the course of autonomous visual crack detection. The first type of study is based on the sequential operation of feature extraction and classification utilizing machine learning classifiers [1-3]. The second type of methodological approach is observed



in studies utilizing deep learning methods (e.g., convolutional neural networks) in which the feature extraction stage is conducted within the black-box algorithm. Zhang et al. use a CNN with six convolution layers to conduct binary crack detection task on roads, in which CNN method shows excellent performance [4]. Similarly, Wang et al. utilize CNN with five convolutional layers for classifying the asphalt pavement cracks based on the 3D data input with depth with 1mm resolution [5]. In this study, the RNN model generated by the ML-ISTA algorithm of the convolutional sparse coding strategy [6-8] is extended to the road crack images classification task, where its performance will be fully validated in the next section.

2. ML-CSC and ML-ISTA

Given a set of convolutional dictionaries $\{\mathbf{D}_i\}_{i=1}^L$ of appropriate dimensions, a signal $\mathbf{x}(\gamma_i) \in \mathbb{R}^N$ admits a representation in terms of multi-layer convolutional sparse coding (ML-CSC) model [6-8], i.e. $\mathbf{x}(\gamma_i) \in \mathcal{M}_\lambda$, if

$$\begin{aligned} \mathbf{x} &= \mathbf{D}_1 \gamma_1, \|\gamma_1\|_{0,\infty}^s \leq \lambda_1, \\ \gamma_1 &= \mathbf{D}_2 \gamma_2, \|\gamma_2\|_{0,\infty}^s \leq \lambda_2, \\ &\vdots \\ \gamma_{L-1} &= \mathbf{D}_L \gamma_L, \|\gamma_L\|_{0,\infty}^s \leq \lambda_L. \end{aligned} \quad (1)$$

In this paper, the multi-layer iterative soft threshold algorithm (ML-ISTA) was applied to solve the ML-CSC problem [9], where the structure of recurrent neural network generated by combining ML-ISTA and CNN is shown in Figure 1. More details about ML-ISTA can be found in reference [9].

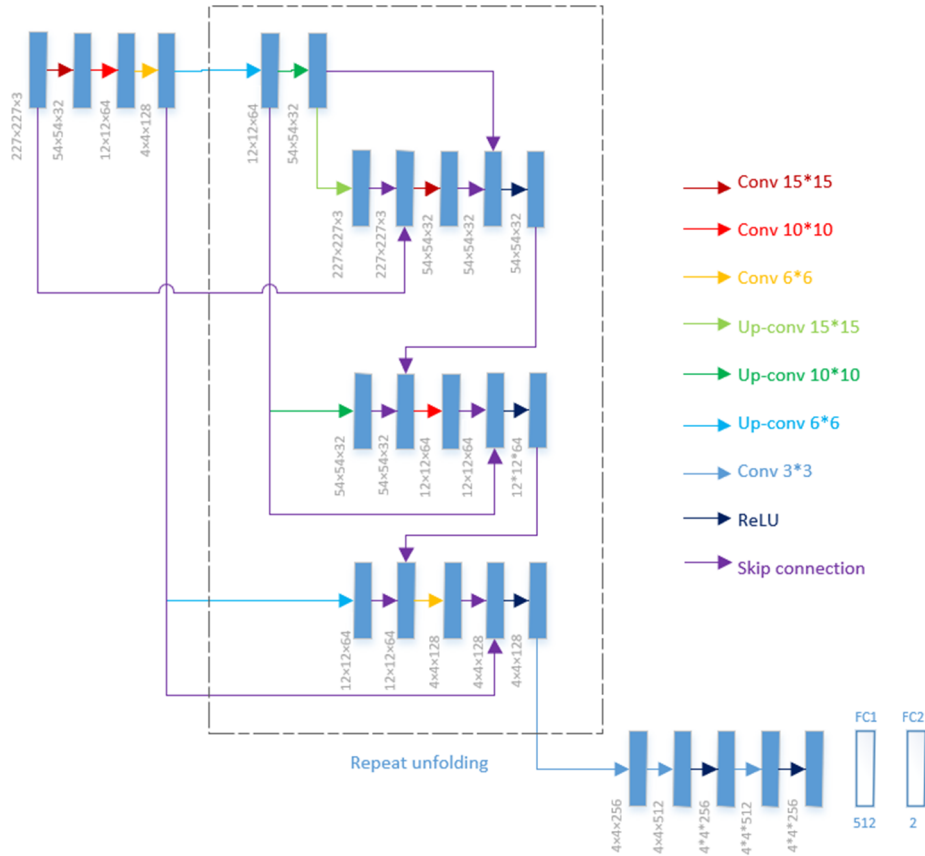


Figure 1. The architecture of recurrent neural network based on ML-CSC with CNN (example for images with 227x227 pixels). Each blue box corresponds to a multi-channel feature map. The x-y-z size is provided at the lower-left edge of the box; white boxes represent fully connected layers; the arrows denote the different operations.

3. Experiments

3.1. Dataset

The original dataset of crack classification comes from [10], which is divided into the positive crack image set and negative crack image set for the image classification task. Each set has 20k images of 227 x 227 pixels. The dataset is generated from 458 high-resolution images (4032x3024 pixels) with the method proposed by [4]. Based on the original dataset, we constructed three datasets with different kinds of noise to validate the performance of RNN based on the ML-CSC model. The experimental flow chart is shown in Figure 2.

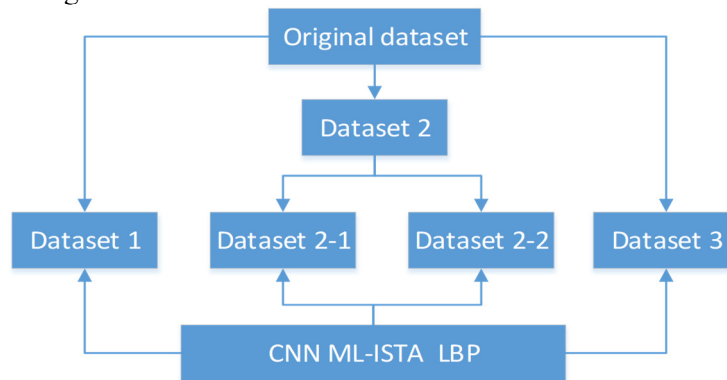


Figure 2. The experimental flow charts.

Dataset 1: we randomly constructed a 12k training set, 4k validation set and 4k testing set from the original dataset. In Dataset 1, the positive and negative crack images are shown in Figure. 3.



Figure 3. The positive crack image (left) and negative crack image (right) of Dataset 1.

Dataset 2-1: we randomly constructed a 12k training set, 4k validation set and 4k testing set from the original dataset, then we added Gaussian noise with a mean of 0 and a variance of 0.05 to the training set of Dataset 2. In the training set of Dataset 2-1, the positive and negative crack images are shown in Figure. 4.

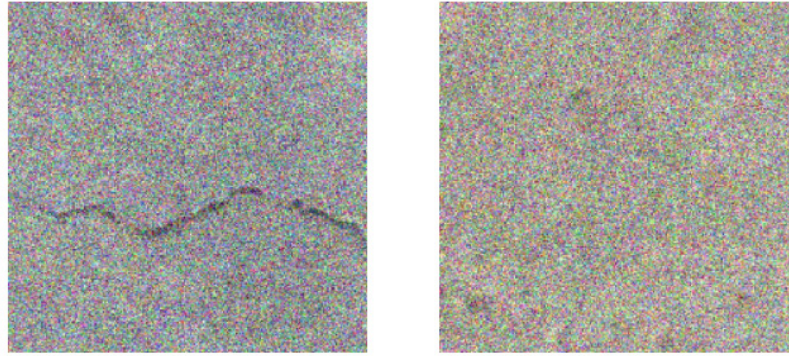


Figure 4. The positive crack image (left) and negative crack image (right) of Dataset 2-1.

Dataset 2-2: we randomly constructed a 12k training set, 4k validation set and 4k testing set from the original dataset, then we added Gaussian noise with a mean of 0 and a variance of 0.075 to the training set of Dataset 2. In the training set of Dataset 2-2, the positive and negative crack images are shown in Figure. 5.

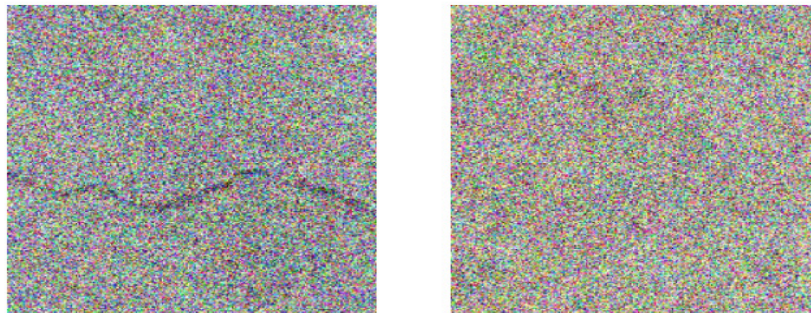


Figure 5. The positive crack image (left) and negative crack image (right) of Dataset 2-2.

Dataset 3: we randomly constructed a 12k training set, 4k validation set and 4k testing set from the original dataset, then add pepper and salt noise with a noise density of 0.03 to the training set. In the training set of Dataset 3, the positive and negative crack images are shown in Figure. 6.

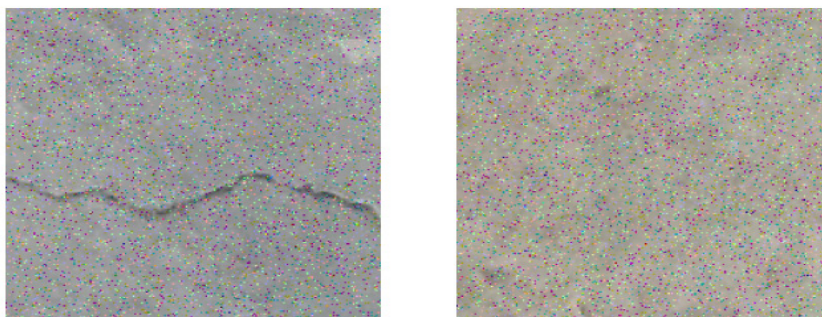


Figure 6. The positive crack image (left) and negative crack image (right) of Dataset 3.

3.2. Experimental Procedure and Results

In this paper, RNN model was constructed through a ML-CSC model with three convolutional layers with 32, 64 and 128 filters, kernel sizes of 15×15 , 10×10 and 6×6 , strides of 4, 4 and 2, and the classifier ζ (γ_6) is a three-layer CNN. A supervised learning setting is adopted to minimize an empirical risk over N training samples of signals y_i with labels h_i . The function $\mathcal{L}$ is a loss function to be minimized during training, where the cross-entropy is used in this paper. Nonnegativity constraints are also enforced on the representations, resulting in the application of ReLUs and biases as shrinkage

operators. Models are trained with stochastic gradient descent (SGD) with momentum, decreasing the learning rate every so many iterations. The details have been shown in Table 1. Moving to a complete comparison, we will demonstrate the ML-ISTA architecture compared with CNN and LBP architecture in different cases and make use of a PyTorch implementation. The configuration of this experiment is 16G memory and one NVIDIA GeForce 1080Ti.

Table 1. Detailed description of the recurrent neural network based on ML-CSC with 3 convolutional layers, unfolding = 6.

Recurrent neural network	
Input: signal $\mathbf{y}$, dictionaries D_1, D_2, D_3, D_4, D_5 .	
Init: Set $\mathbf{y}_0^k = \mathbf{y}$	
1:	$\mathbf{y}_1^1 = D_1^T \mathbf{y}$
2:	$\mathbf{y}_2^1 = D_2^T \mathbf{y}_1^1$
3:	$\mathbf{y}_3^1 = D_3^T \mathbf{y}_2^1$
4:	for $k=1:6$ do
5:	$\bar{\mathbf{y}}_i = D_{(i+1,3)} \mathbf{y}_3^k$
6:	for $i=1:3$ do
7:	$\mathbf{y}_i^{k+1} = \text{ReLU}(\mathbf{y}_i - D_i^T(D_i \bar{\mathbf{y}}_i - \mathbf{y}_{i-1}^{k+1}))$
8:	$\mathbf{y}_4 = D_4^T \mathbf{y}_3^7$
9:	$\mathbf{y}_5 = \text{ReLU}(D_5^T \mathbf{y}_4)$
10:	$\mathbf{y}_6 = \text{ReLU}(D_5^T \mathbf{y}_5)$
11:	$\min_{\theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(h_i, \zeta_{\theta}(\mathbf{y}_6))$

- CNN: the case of 0 unfoldings corresponds to the typical feed-forward CNN.
- ML-ISTA: the case with 6 unfoldings effectively implements an 18-layers-deep architecture by default.
- Layered Basis Pursuit: the approach proposed in [11], which unrolls the iteration of ISTA for a single-layer BP problem at each layer. In contrast, the proposed ML-ISTA unrolls the iterations of the entire Multi-Layer BP problem.

First of all, the models were trained in the Dataset 1 of 10,000 training samples and 10,000 testing samples. The epoch is set to 30, 60 and 90. In different epochs, we compare the ML-ISTA, CNN, and LBP methods that the testing results are shown in Figure. 7 and Table 2. It can be seen from the resulting chart of Dataset 1 that the LBP algorithm is better than the CNN and ML-ISTA algorithms for crack classification when the epoch is very tiny. As the epoch increases, the accuracy of the three algorithms is increasing. Nevertheless, in the end, the accuracy of the ML-ISTA algorithm for crack classification is significantly higher than that of CNN and LBP algorithms. On Dataset 1, the inferencetimes of CNN, ML-ISTA and LBP algorithms are very similar.

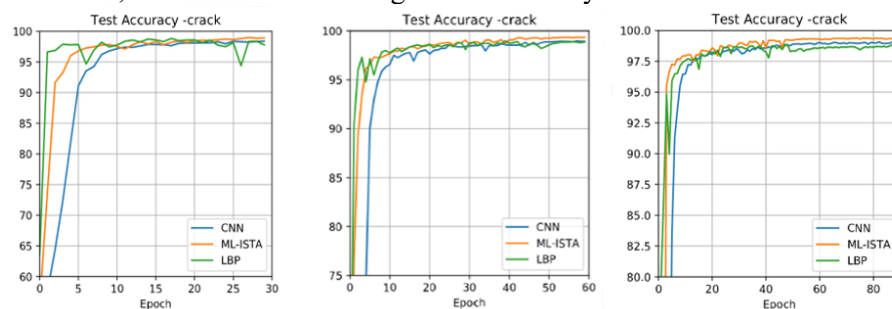


Figure 7. Comparison of different architectures on Dataset 1, all networks have the same number of parameters.

Table 2. Classification results for different architectures for Dataset 1 (epoch=60).

Model	CNN	ML-ISTA	Layered BP
Test accuracy	98.90%	99.36%	98.84%

Secondly, the models were trained in the Dataset 2-1 and Dataset 2-2 of 10,000 training samples and 10,000 testing samples. The epoch is set to 30, 60 and 90. In different epochs, we compare the ML-ISTA, CNN, and LBP methods that the testing results are shown in Figure. 8 and Table 3. It can be concluded that in the Gaussian noise environment with different variances, the LBP algorithm is better than CNN and ML-ISTA algorithms for crack classification when the epoch is very tiny. However, as the epoch increases, the accuracy of the CNN and ML-ISTA algorithm are increasing, and the LBP algorithm has fluctuations sharply. In the end, the LBP algorithm has been unable to converge; the accuracy of the ML-ISTA algorithm for crack classification is significantly higher than that of CNN algorithms.

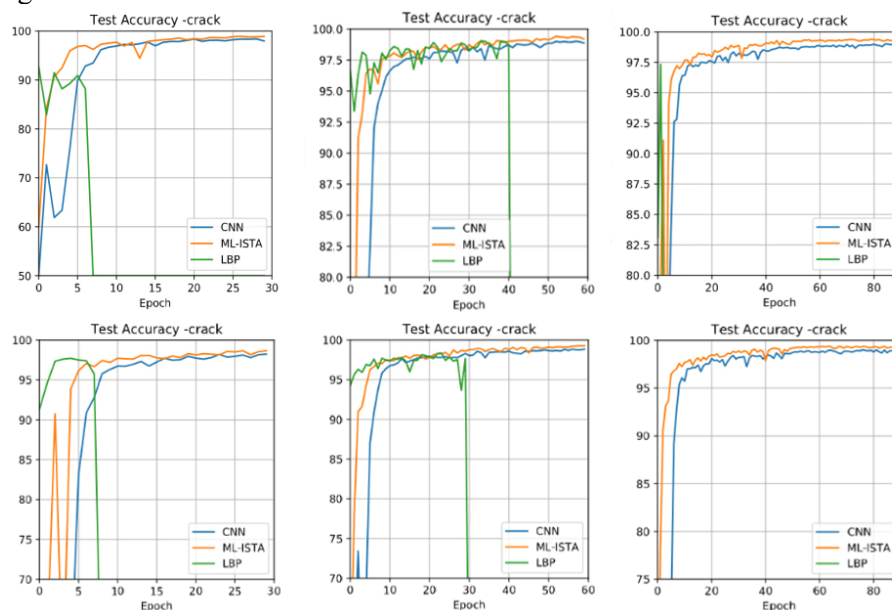


Figure 8. Comparison of different architectures on the Dataset 2-1 (top) and Dataset 2-2 (bottom), all networks have the same number of parameters.

Table 3. Classification results for different architectures for Dataset 2-1 and Dataset 2-2. The LBP algorithm has been unable to converge (epoch=60).

Test accuracy of Model	CNN	ML-ISTA	Layered BP
Dataset 2-1	98.88%	99.20%	-
Dataset 2-2	98.85%	99.17%	-

Finally, the models were trained in the Dataset 3 of 10,000 training samples and 10,000 testing samples. The unfolding of ML-ISTA and LBP is set to 3, 6 and 9 and the epoch is set to 90. In different epochs, we compare the ML-ISTA, CNN, and LBP methods that the testing results are shown in Figure. 9 and Table 4. It can be seen from Figure.9 and Table 4. Under different unfolding, the accuracy of the ML-ISTA algorithm for crack classification is significantly higher than the results of the CNN and LBP algorithms. In Dataset 3 the case with 3 unfoldings of the test accuracy of the ML-ISTA algorithm is the highest.

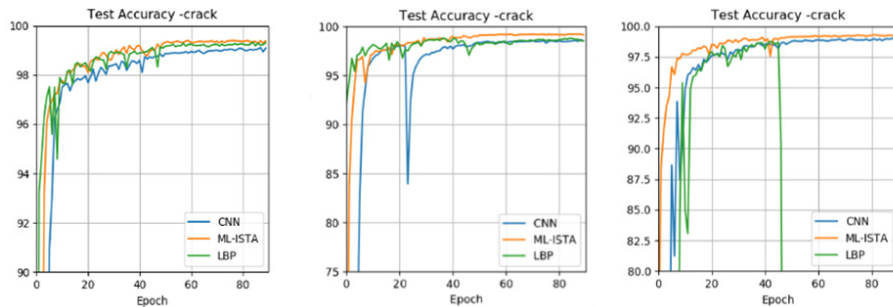


Figure 9. Comparison of different architectures and different unfoldings on Dataset 3, all networks have the same number of parameters, unfolding = 3 (left), unfolding = 6 (middle) and unfolding = 9 (right).

Table 4. Classification results for different architectures for different unfoldings. In the case of unfolding=9, the LBP algorithm has been unable to converge (epoch=90).

Model	Test accuracy
CNN	98.84%
ML-ISTA, unfolding=3	99.35%
ML-ISTA, unfolding=6	99.25%
ML-ISTA, unfolding=9	99.27%
Layered BP, unfolding=3	99.29%
Layered BP, unfolding=6	98.83%
Layered BP, unfolding=9	-

Table 5. Classification results for different architectures for different datasets. In the case of Dataset 2, the LBP algorithm has been unable to converge (epoch = 90, unfolding = 6).

Model	CNN	ML-ISTA	Layered BP
Dataset 1	98.90%	99.36%	98.84%
Dataset 2-1	98.88%	99.20%	-
Dataset 2-2	98.85%	99.17%	-
Dataset 3	98.84%	99.25%	98.83%

4. Conclusion

Through the above experiments, we found that ML-CSC is a promising model for crack detection, but ML-CSC is excessively dependent on the accuracy of the solving algorithm and the number of unfolding. The accuracy of the LBP on the testing set is better than ML-ISTA and CNN when the training epochs are small, but with the increase of the training epochs, the accuracy of ML-ISTA on the testing set is better than LBP and CNN. Moreover, LBP is very unstable, and it will fail to converge when the noise is large and the number of unfoldings is too high. For the dataset of this experiment, when the number of ML-ISTA unfoldings is 3, the testing set has the highest accuracy. No matter in different noise environments, different unfoldings, or different training epochs, the final test accuracy of ML-ISTA is better than CNN and LBP. Also, the speed is comparable to CNN or LBP. In the future, the ML-CSC model combining with current mainstream methods will be used to design a sparse convolutional neural network for object detection and semantic segmentation of the crack images.

Acknowledgments

This research was supported by National Natural Science Foundation of China (No. 61701318), Natural Science Research Project of Jiangsu Higher Education Institutions (No.18KJB416001,

19KJB160001), Project of “Six Talent Peaks” of Jiangsu Province (No. SWYY-017) and the Lianyungang city Haiyan project Grant 2019-QD-004.

References

- [1] Wu, W., Liu, Z., & He, Y. (2015) Classification of defects with ensemble methods in the automated visual inspection of sewer pipes. *Pattern Analysis and Applications.*, 18(2): 263-276.
- [2] Zhu, A., Zhu, Y., Wang, N., & Chen, Y. (2020) A robust self-driven surface crack detection algorithm using local features. *Insight-Non-Destructive Testing and Condition Monitoring.*, 62(5): 269-276.
- [3] De-Fang, W., Wei-Ming, Z., & Ni-Zhuan, W. (2015) Road crack detection under uneven illumination using improved k-means algorithm. *Computer Applications and Software.*, 32(7): 244-247.
- [4] Zhang, L., Yang, F., Zhang, Y. D., & Zhu, Y. J. (2016) Road crack detection using deep convolutional neural network. In: *IEEE international conference on image processing (ICIP)*. Phoenix. pp. 3708-3712.
- [5] Wang, K. C., Zhang, A., Li, J. Q., Fei, Y., Chen, C., & Li, B. (2017) Deep learning for asphalt pavement cracking recognition using convolutional neural network. In: *Airfield highway pavements 2017: Airfield pavement technology and safety*, Philadelphia. pp. 166-177.
- [6] Sulam, J., Pappas, V., Romano, Y., & Elad, M. (2018) Multilayer convolutional sparse modeling: Pursuit and dictionary learning. *IEEE Transactions on Signal Processing.*, 66(15): 4090-4104.
- [7] Pappas, V., Romano, Y., Sulam, J., & Elad, M. (2018) Theoretical foundations of deep learning via sparse representations: A multilayer sparse model and its connection to convolutional neural networks. *IEEE Signal Processing Magazine.*, 35(4): 72-89.
- [8] Aberdam, A., Sulam, J., & Elad, M. (2019) Multi-layer sparse coding: The holistic way. *SIAM Journal on Mathematics of Data Science.*, 1(1): 46-77.
- [9] Sulam, J., Aberdam, A., Beck, A., & Elad, M. (2020) On multi-layer basis pursuit, efficient algorithms and convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence.*, 42(8): 1968-1980.
- [10] Özgenel, Ç. F., & Söğüt, A. G. (2018) Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In: *Proceedings of the International Symposium on Automation and Robotics in Construction*. Berlin. pp. 1-8.
- [11] Pappas, V., Romano, Y., & Elad, M. (2017) Convolutional neural networks analyzed via convolutional sparse coding. *The Journal of Machine Learning Research.*, 18(1): 2887-2938.