Journal of Southeast University (English Edition) Vol. 39, No. 4, pp. 393 − 399 Dec. 2023 ISSN 1003—7985

Concrete crack identification in complex environments based on SSD and pruning neural network

Wang Yanhua[1,2] He Junze[1,2] Zhang Mingzhou[1,2] Dai Bowen[1,2] Xu Haoran[3]

([1]Key Laboratory of Concrete and Prestressed Concrete Structures of Ministry of Education, Southeast University, Nanjing 211189, China)

([2]School of Civil Engineering, Southeast University, Nanjing 211189, China)

([3]Chien-Shiung Wu College, Southeast University, Nanjing 211189, China)

Abstract: To solve the problem of poor crack identification algorithm performance in complex environments, an improved method based on a single-shot multibox detector (SSD) algorithm was proposed. This method realized high-precision crack identification for crack images with added noise by adjusting the combination of the number of different resolution prior bounding boxes in the original SSD algorithm. A sufficient number of crack images were captured and preprocessed in actual scenes and laboratories, and noise was added to the crack dataset using a pretzel algorithm to simulate the crack images in complex environments. The improved method was tested along with the original SSD algorithm to identify the crack dataset, and their test results were compared. The results show that the crack identification accuracy of the original SSD algorithm and improved method decreases with increasing noise levels. When a 20% grade of pretzel noise is added at high density, the accuracy in recognizing cracks is 31.7% and 93.0% for the original SSD algorithm and the improved method, respectively. Therefore, the improved method has excellent antinoise ability and can be used for crack identification in complex environments.

Key words: crack identification; pruned neural network; image noise addition; antinoise performance; disease detection

DOI: 10.3969/j. issn. 1003 − 7985. 2023. 04. 008

The most prevalent problem on bridge and road surfaces is cracks, which represent the greatest threat to the safety of bridges and roads among all types of issues associated with pavements and bridges. Cracks on a concrete surface can lead to secondary disasters in its structure. To eliminate possible safety hazards (e. g., cracks), timely inspections must be conducted for any damages to bridges during services[1].

Received 2023-02-10, **Revised** 2023-08-01.

Biography: Wang Yanhua (1977—), female, doctor, researcher-level senior engineer, wyh00737@seu.edu.cn.

Foundation items: The National Major Scientific Research Instrument Development Project (No. 11827801), the National Science and Technology Project (No. 2020YFC1511904).

Citation: Wang Yanhua, He Junze, Zhang Mingzhou, et al. Concrete crack identification in complex environments based on SSD and pruning neural network[J]. Journal of Southeast University (English Edition), 2023, 39(4): 393 − 399. DOI: 10.3969/j. issn. 1003 − 7985. 2023. 04. 008.

Manual surveying was once the most common method for discovering fractures during bridge and highway maintenance, but it was time-consuming and subjective, resulting in low accuracy. Recently, the field of civil engineering inspection has incorporated some computer vision and deep learning techniques, such as crack identification, displacement measurement, and fatigue detection[2−8]. Because deep learning-based detection algorithms can successfully interpret images and correctly identify targets, they have been widely used for crack detection with advancements in artificial intelligence. In 2004, Ouellette et al.[9] originally used convolutional neural networks (CNNs) to solve crack identification problems. The convolutional neural filter obtained by combining a genetic algorithm with the neural network was trained to solve the problem encountered by the neural network algorithm, which is its vulnerability to the interference of minimum value information.

The identification accuracy of crack images can be affected by complex environments, such as weather, light, and noise. In rainy or foggy conditions, low image contrast and image brightness are likely to occur, resulting in reduced crack identification. Many scholars have investigated crack identification with de-raining[10−11] and de-fogging[12−14] models, which were basically improved models based on traditional image processing algorithms. As a result, crack recognition accuracy in rainy or foggy environments was improved, but it remains insufficient. Additionally, interference from the shooting environment generally leads to a certain degree of noise in the captured images. Further, these noise points can cause sharp changes in the image brightness, which will then affect the crack identification accuracy. In 2021, Yang et al.[15] introduced a high antinoise extraction approach based on a dynamic threshold for concrete fractures having complex noise, such as honeycomb pockmarks on the surface.

Although concrete crack development is a slow process, the general method for crack identification leads to a low identification rate in high-altitude areas, areas with constant rain and fog, etc. The current algorithm for detecting cracks is flawed because it does not consider crack identification in complex environments similar to those described above and cannot meet requirements for applying it in particular scenarios in actual engineering. In this research, an improved SSD algorithm[16] based on the prior bounding box is proposed for identifying noise-added crack images, enabling the algorithm to better adapt to complex noisy environments.

1 SSD Neural Networks

The SSD algorithm is based on a CNN structure, the current machine learning algorithm with faster detection speed and higher accuracy rate. Firstly, sampling frames with different aspect ratios and sizes are taken at different locations on the image for uniform sampling. Then, the features are extracted by CNN and classified, and the category probability is directly displayed on the image after the test image is detected. Finally, the detection result can be obtained after a single detection. Compared with that of other traditional neural network algorithms, the identification effect is more accurate.

The network structure of SSD is optimized according to the VGG16 network model. SSD uses the method of increasing the number of convolutional layers for object detection of images, enabling the algorithm to utilize more feature maps, and this ability is one of the reasons for the high identification accuracy of the SSD algorithm.

2 Dataset Acquisition

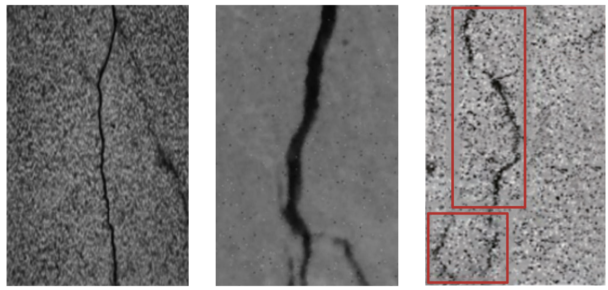
2.1 Crack dataset preprocessing

The crack data used in the experiments were obtained from the field bridge crack images and the damage test images in the laboratory of Southeast University.

Before training the CNN, the crack dataset must be marked. This process is similar to the process of allowing machine learning: The target that must be trained and identified is transmitted to the neural network so that it knows which area the crack image is in. This area is the target that needs to be learned and identified. The LabelImg tool is used to label the cracked images. First, the target position is marked in the image. Then, the file is transformed into XML format and transferred to the framework of a deep learning algorithm for training.

2.2 Noise addition to crack images

To increase the difficulty of crack identification, a complex environment was simulated by adding salt-and-pepper noise based on the original crack atlas, which is a black-and-white noise produced via image sensors, transmission channels, and decoding processes[17]. To better simulate the cracked images in each complex environment and make the algorithm universal and noise-resistant, five noise addition levels were used to add noise to the cracked images, and their signal-to-noise ratios were 5%, 10%, 15%, and 20%. Fig. 1 shows the effect of different degrees of noise addition.



(a) (b) (c)

Fig. 1 Noise-added crack diagram. (a) Pothole crack diagram; (b) Noise-plus-20% crack diagram; (c) High-density noisy crack diagram

3 Construction of SSD-Based Algorithm Model

3.1 Training process

Before the final training, a positive and negative sample matching method was required to better train the dataset; that is, the preselected boxes and the final prediction boxes were matched according to the IoU principle[18]. If the match is successful, it is a positive sample; otherwise, it is a negative sample. The samples with larger loss values (i. e., the set of samples where negative samples are as easily seen as positive samples) were used for training, thus ensuring that the ratio of positive samples to negative samples is approximately 1∶3[19].

The stochastic gradient descent algorithm was used to train the crack data model for its higher recognition rate for the validation set[20]. After continuously training and adjusting the parameters, each parameter is set as shown in Table 1.

3.2 Pruning improvement algorithm based on prior bounding boxes

Deep learning network models have numerous redundant

Table 1 Parameter setting values

Parameter	Description	Value
B_s	Batch size (the number of samples used in one iteration)	32
G	Value of variables of the Gamma function	0.1
L_r	Learning rate[21] (step size for control parameter adjustment)	0.01
L_s	Step size for learning rate update	$[2.8\times10^5, 3.6\times10^5]$
M_I	Maximum number of iterations	5×10^4
M	Momentum parameters (differences in the direction of control parameter updates)	0.9
W_f	Learning rate initialization factor	1/3
W_I	Number of iterations of the warm-up process	500

parameters at all stages oftraining and learning identification, and many of their activation values tend to be zero. Hence, removing these redundant parameters can considerably improve the model expression ability, which is called network pruning.

The improved algorithm in this paper is predominantly based on the analysis of the SSD model structure, resulting in a pruning improvement algorithm based on prior bounding boxes. The algorithm focuses on the unique structure of the SSD algorithm[22], and after counting the number of its prior bounding boxes in the training, the candidate prior boxes that can avoid the effect of external noise and the pretzel algorithm are selected.

The structure of the SSD algorithm model is analyzed as shown in Fig. 2, and the number of required localization prediction boxes is calculated as follows: $38\times38\times4+19\times19\times6+10\times10\times6+5\times5\times6+3\times3\times4+1\times1\times4=8\ 732$. Fig. 2 shows that SSD predicts the localization frames after extracting the features of different discriminative layers of the backbone network. Then, all the results of 8 732 prediction boxes are outputted and filtered using nonmaximum suppression (NMS) to obtain the optimal target box. The appropriate bounding boxes are selected statistically and computationally, the resolution bounding boxes with low statistical numbers are eliminated, and the number of prior bounding boxes is reasonably allocated using the analysis of variance method. This enables the model to focus on the more effective resolution during training[23].

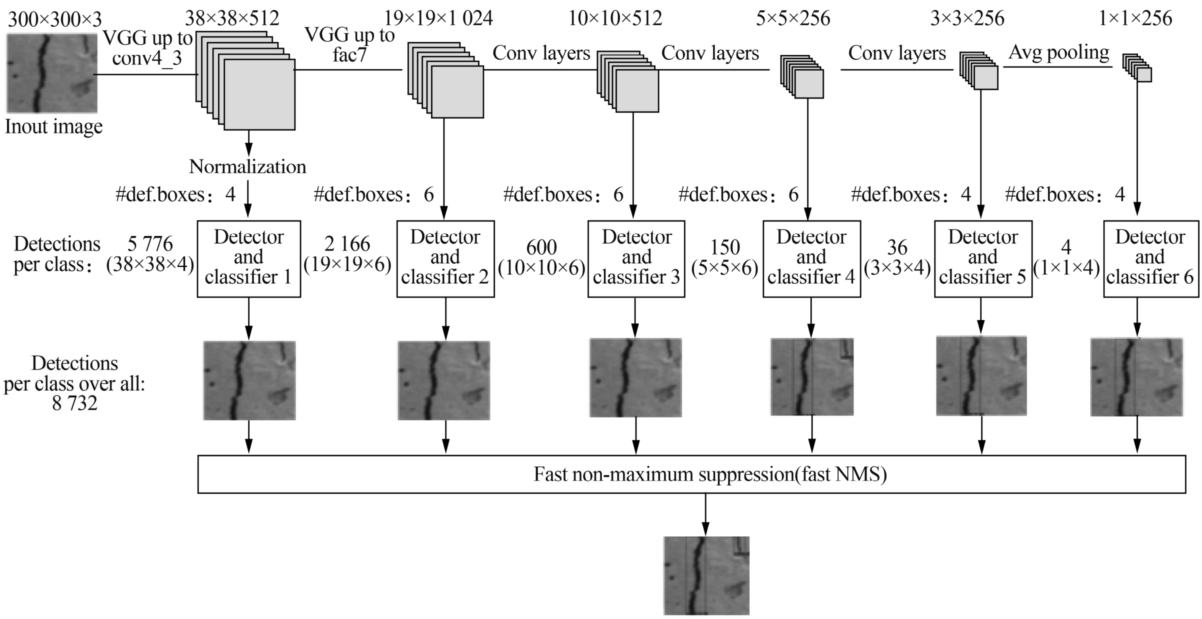


Fig. 2 SSD crack identification process

The original SSD algorithm was statistically tested after counting and analyzing the multiscale and multidirectional bounding boxes. A statistical histogram of the number of correctly predicted target location sources is shown in Fig. 3 for classifying boundingboxes. In the vertical coordinates, parentheses indicate the prior box resolution, and outside the parentheses, the aspect ratio of the prior box is provided.

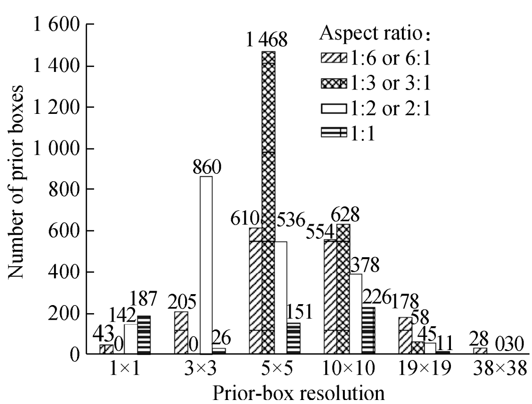


Fig. 3 Statistics on the number of prior bounding boxes

As shown in Fig. 3, the prior bounding box with the largest number of resolutions and shapes is (5×5); furthermore, its shape is rectangular with an aspect ratio of (1∶3 or 3∶1). Therefore, the variable analysis method is used to adjust the number of prior candidate boxes of each resolution based on a sum of 30 prior boxes. The accuracy of crack identification with noise addition of 20% is compared in various cases to determine the final selection of the number of prior boxes.

The analysis in Table 2 shows that the identification accuracy of the proposed algorithm considerably increases after decreasing the number of prior bounding boxes that account for a smaller percentage but occupy a large amount of computing memory. Moreover, after increasing the number of prior bounding boxes that account for the highest percentage, the identification accuracy shows a slight decline due to an increase in the total number of prior bounding boxes, which may require increased computation by the algorithm operation.

Table 2 Comparison of identification accuracy for each prior bounding box number case

Candidate box	38 ×38	19 × 19	10 × 10	5 × 5	3 × 3	1 × 1	Precision/%
Original group	4	6	6	6	4	4	56
Group 1	4	4	6	8	4	4	68
Group 2	2	4	8	8	4	4	85
Group 3	2	2	8	8	6	4	93
Group 4	2	2	8	10	4	4	91

The above statistics and analysis reveal that in image crack identification, large-scale perceptual elements play the most dominant role, while overabundant small-scale identification processes reduce the identification accuracy and increase the training identification time due to the interference of the pretzel noise algorithm. Thus, it is necessary to re-optimize the allocation of candidate frames with the multiscale resolution, reduce the number of small-scale prior bounding boxes, and increase the number of medium- and large-scale prior bounding boxes.

The final number of selected prior bounding boxes is Group 3. After modifying the coefficients of the number of prior bounding boxes, the number of boxes decreases from 8 732 to 4 668, which nearly doubles the number of candidate boxes, and the accuracy also increases slightly. This particularly improves the identification accuracy of cracked images after enhancing the noise level of the pepper algorithm. The results show that such improvements facilitate and expedite the training process and improve the identification accuracy of the detection, particularly in resisting the noise attack of this fragmented black-and-white pretzel algorithm.

4 Results and Analysis

4.1 Analysis of the original algorithm results

A complete crack dataset was produced, and after training with the original algorithm model, crack images with 0%, 5%, 10%, 15%, and 20% noise addition were identified. The identification results are presented statistically in Table 3.

Table 3 Comparison of identification accuracy for each prior bounding box number case

Noise addition/%	AP50	AP75	mAP
0	0.964	0.720	0.791
5	0.844	0.644	0.693
10	0.786	0.606	0.649
15	0.672	0.546	0.571
20	0.615	0.515	0.532

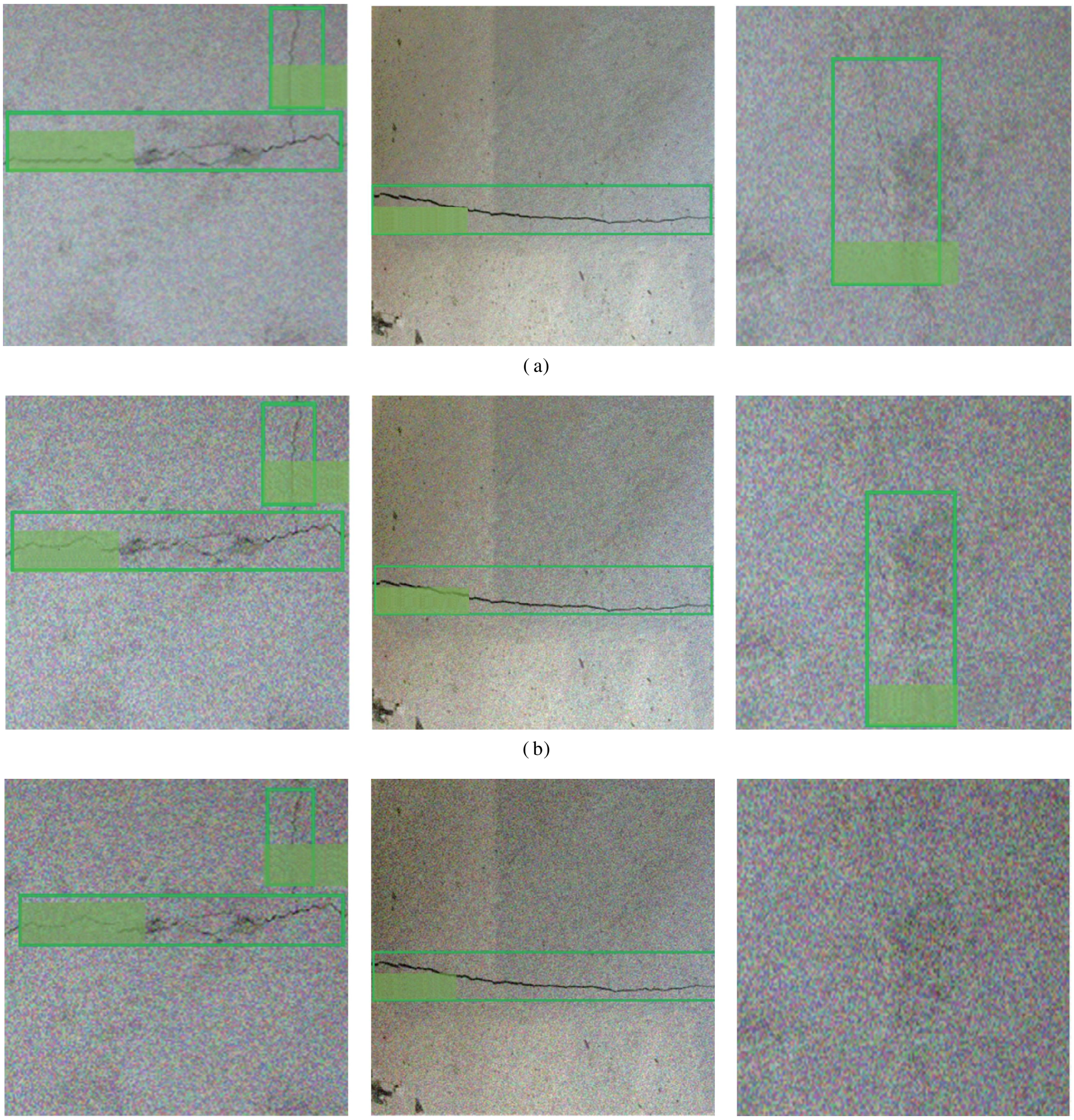
The identification effect of the original SSD algorithm gradually decreased with increasing noise addition levels. The pepper algorithm made identifying crack images more difficult; however, it is crucial for simulating external environmental noise. The identification accuracy of the SSD algorithm reached 0.791 when the crack image was not noise-added, whereas its identification accuracy decreased to 0.532 when the noise addition reached 20%, indicating that the original algorithm has poor noise immunity and the identification effect was influenced by the deepened external environmental noise.

4.2 Analysis of experimental results

In this experimental model, training and validation sets of 5 000 and 3 000 images, respectively, were set. Crack images with normal no-noise addition and noise addition of 5%, 10%, 15%, and 20% were recognized. For verifying the improved algorithm, two types of noise processing were applied to the crack images, one a high-density pretzel algorithm noise addition to the crack images, and the other was a different level of film grain noise addition to the crack images. Then, the algorithm was used to recognize these two images. The film grain results from the poor quality of film sensitization, and this noise is similar to the pepper algorithm noise addition but with richer visual sensory colors. This can provide sufficient validation regarding the identification ability of the improved algorithm. Part of the verification focuses on crack identification, as shown in Fig. 4.

According to the identification map of cracks after partial noise addition, the improved algorithm can accurately recognize most of the cracks and has higher identification accuracy for the cracks after noise addition. However, it failed to recognize some of the images with unclear cracks, or it has difficulty recognizing some fine cracks when they are disturbed by high-density noise, resulting in missed detection. The overall performance of the improved algorithm is better than before, and the identification accuracy is shown in Table 4. The improved before and after algorithms were also used to train and recognize the crack images after the high-density pepper algorithm with noise. Finally, the results are compared and analyzed, as shown in Fig. 5.

As seen from Fig. 5, the identification accuracy of the initial SSD algorithm gradually decreases with increasing noise levels. When the pepper noise of 20% grade is added at high density, the identification accuracy of the algorithm only reaches 31.7%. The identification accuracy also decreases substantially when the noise is added slightly. In the training curve before the improvement, the loss rate is gradually reduced only with increasing training iterations. After improving the algorithm, the



(a)

(b)

(c)

Fig. 4 Film particle noise verification results. (a) Twenty-film grain level noise crack identification; (b) Thirty-film grain level noise crack identification; (c) Forty-film grain level noise crack identification

Table 4 Improved algorithm identification accuracy

Noise addition/%	Based SSD	Improved SSD
0	0.926	0.966
5	0.621	0.962
10	0.524	0.958
15	0.426	0.951
20	0.317	0.930

identification accuracy can reach 93.0% even under the 20%-level of the pepper algorithm plus noise. Furthermore, the improved training curve shows that the loss rate has almost converged to 0 after 5 000 training iterations, and the convergence is more rapid. The improvement strengthened the identification ability of the algorithm and considerably enhanced its noise immunity. The accuracy of the original SSD and proposed algorithms in identifying cracks decreases with increasing noise levels.

5 Conclusions

1) For recognizing crack images with various types of noise information, this study proposed an improved SSD crack identification algorithm based on the prior bounding box, achieving better results than the original SSD algorithm.

2) After training and adjusting parameters, the initial SSD object detection algorithm has an identification accuracy of only 92.6% and 31.7% for noise-free and noise-added 20% crack images, respectively. Hence, it cannot meet the requirements for accurate identification of crack images with noisy information.

3) According to the characteristics of the SSD network structure, an improved SSD algorithm based on the prior bounding box is proposed for the first time. The identification of crack images with various noise levels is

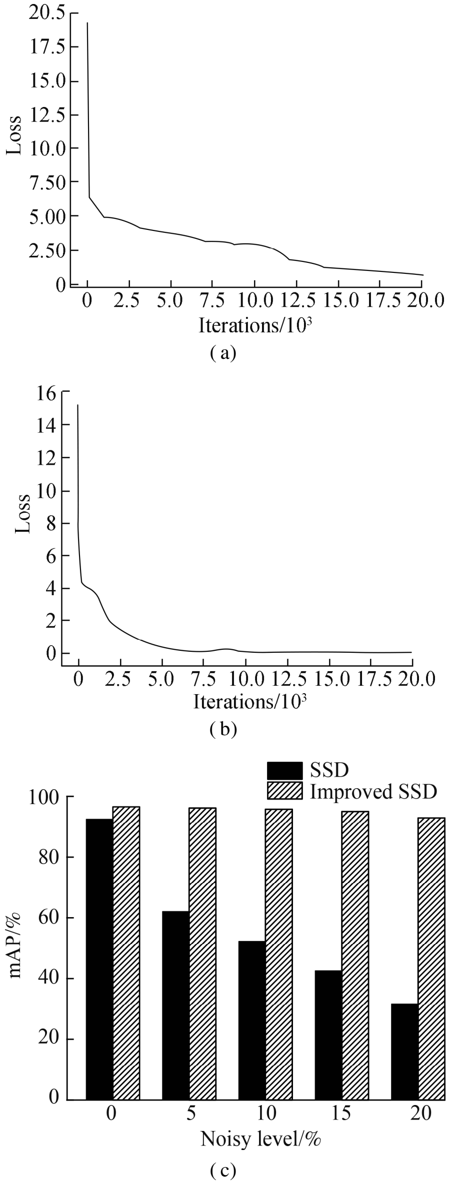


Fig. 5 Comparison before and after training. (a) Pre-improvement training curves; (b) Post-improvement training curves; (c) Comparison of identification accuracy

performed using the improved algorithm, and its identification accuracy reaches 96.6% and 93.0% for no-noise-added and noise-added 20% crack images, respectively. The identification effect is good for crack images with strong interference noise information.

4) The next stage of research will apply the algorithm for unmanned intelligent inspection to identify and improve the bridge cracks in complex environments while continuing to improve the algorithm and testing its limits.

References

[1] Ni Y H, Lu H, Ji C, et al. Comparative analysis on bridge corrosion damage detection based on semantic segmentation[J]. *Journal of Southeast University* (*Natural Science Edition*), 2023, **53**(2): 201 − 209. DOI: 10.3969/j.issn.1001-0505.2023.02.003. (in Chinese)

[2] Cheng J M, Gao Y Q, Yan H D, et al. Statistical analysis of nondestructive testing results of cast steel joints in civil engineering structures[J]. *Journal of Southeast University* (*English Edition*), 2022, **38**(1): 1 − 8. DOI: 10.3969/j.issn.1003-7985.2022.01.001.

[3] Jiang J W, Ni F J. Evaluation of fatigue property of asphalt mixtures based on digital image correlation method [J]. *Journal of Southeast University* (*English Edition*), 2017, **33**(2): 216 − 223. DOI: 10.3969/j.issn.1003-7985.2017.02.015.

[4] Li Z J, Wang H, Wang R G, et al. Experimental study on fatigue performance of diaphragm openings of orthotropic steel bridge decks based on 3D-DIC[J]. *Journal of Southeast University* (*Natural Science Edition*), 2019, **49**(6): 1116 − 1123. DOI: 10.3969/j.issn.1001-0505.2019.06.014. (in Chinese)

[5] Ngeljaratan L, Moustafa M A. Structural health monitoring and seismic response assessment of bridge structures using target-tracking digital image correlation[J]. *Engineering Structures*, 2020, **213**: 110551. DOI: 10.1016/j.engstruct.2020.110551.

[6] Zhang L, Yang F, Daniel Z Y, et al. Road crack detection using deep convolutional neural network[C]//*IEEE International Conference on Image Processing* (*ICIP*). Phoenix, AZ, USA, 2016: 3708 − 3712. DOI: 10.1109/ICIP.2016.7533052.

[7] Cha Y-J, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convolutional neural networks: Deep learning-based crack damage detection using CNNs[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2017, **32**(5): 361 − 378. DOI: 10.1111/mice.12263.

[8] Lang H, Wen T, Lu J, et al. 3D pavement crack detection method based on deep learning[J]. *Journal of Southeast University* (*Natural Science Edition*), 2021, **51**(1): 53 − 60. DOI: 10.3969/j.issn.1001-0505.2021.01.008. (in Chinese)

[9] Oullette R, Browne M, Hirasawa K. Genetic algorithm optimization of a convolutional neural network for autonomous crack detection[C]// *Proceedings of the* 2004 *Congress on Evolutionary Computation*. Portland, ME, USA, 2004: 516 − 521. DOI: 10.1109/CEC.2004.1330900.

[10] Xu J, Zhao W, Liu P, et al. Removing rain and snow in a single image using guided filter[C]// *IEEE International Conference on Computer Science and Automation Engineering* (*CSAE*). Zhangjiajie, 2012: 304 − 307. DOI: 10.1109/CSAE.2012.6272780.

[11] Xu J, Zhao W, Liu P, et al. An improved guidance image based method to remove rain and snow in a single image[J]. *Computer and Information Science*, 2012, **5**(3): 49 − 55. DOI: 10.5539/cis.v5n3p49.

[12] Tan R T, IEEE. Visibility in bad weather from a single image[C]// *The IEEE Conference on Computer Vision and Pattern identification*. Anchorage, AK, USA, 2008: 1 − 8. DOI: 10.1109/cvpr.2008.4587643.

[13] Satrasupalli S, Daniel E, Guntur S R. Single image haze removal based on transmission map estimation using encoder-decoder based deep learning architecture[J]. *Optik*, 2021, **248**: 168197. DOI: 10.1016/j.ijleo.2021.168197.

[14] Ling Z, Fan G, Gong J, et al. Learning deep transmis-

sion network for efficient image dehazing[J]. *Multimedia Tools and Applications*, 2019, **78**(1): 213 − 236. DOI: 10.1007/s11042-018-5687-0.

[15] Yang C Q, Li S, Wang B K, et al. High anti-noise extraction and identification method for concrete cracks based on dynamic threshold[J]. *Journal of Southeast University* (*Natural Science Edition*), 2021, **51**(6): 967 − 972. DOI: 10.3969/j.issn.1001-0505.2021.06.007. (in Chinese)

[16] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]// 14*th European Conference on Computer Vision* (*ECCV*). Amsterdam, Netherlands, 2016: 21 − 37. DOI: 10.1007/978-3-319-46448-0_2.

[17] Azzeh J, Zahran B, Alqadi Z. Salt and pepper noise: effects and removal[J]. *JOIV*: *International Journal on Informatics Visualization*, 2018, **2**(4): 252 − 256. DOI: 10.30630/joiv.2.4.151.

[18] Padilla R, Netto S L, Da Silva E. A. B. A survey on performance metrics for object-detection algorithms [C]// 27*th International Conference on Systems, Signals and Image Processing* (*IWSSIP*). Niteroi, Brazil, 2020: 237 − 242. DOI: 10.1109/IWSSIP48289.2020.9145130.

[19] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE Trans Pattern Anal Mach Intell*, 2010, **32**(9): 19. DOI: 10.1109/TPAMI.2009.167.

[20] Wilson A C, Roelofs R, Stern M, et al. The marginal value of adaptive gradient methods in machine learning [C]// 31*st Annual Conference on Neural Information Processing Systems* (*NIPS*). Long Beach, CA, USA, 2017: 4148 − 4158. DOI: 10.48550/arXiv.1705.08292.

[21] Smith L N. Cyclical learning rates for training neural networks[C]// *IEEE Winter Conference on Applications of Computer Vision* (*WACV*). Santa Rosa, CA, USA, 2017: 464 − 472. DOI: 10.1109/wacv.2017.58.

[22] Girshick R. Fast R-CNN[C]// *IEEE International Conference on Computer Vision* (*ICCV*). Santiago, Chile, 2015: 1440 − 1448. DOI: 10.1109/ICCV.2015.169.

[23] Wan C, Xiong X, Wen B, et al. Crack detection for concrete bridges with imaged based deep learning[J]. *Science Progress*, 2022, **105** (4): 1 − 10. DOI: 10.1177/00368504221128487.

基于 SSD 及剪枝神经网络的复杂环境下混凝土裂缝识别

王燕华[1,2]　何俊泽[1,2]　张明洲[1,2]　戴博闻[1,2]　徐浩然[3]

([1] 东南大学混凝土及预应力混凝土结构教育部重点实验室, 南京 211189)
([2] 东南大学土木工程学院, 南京 211189)
([3] 东南大学吴健雄学院, 南京 211189)

摘要:为解决裂缝识别算法在复杂环境下性能不佳的问题,提出了一种基于单激发多框检测器(SSD)算法的改进方法.该方法通过调整原始 SSD 算法中不同分辨率先验框数量的组合,实现对存在噪声的裂缝图像的高精度裂缝识别.在真实场景和实验室中采集足够数量的裂缝图像并进行预处理,利用椒盐算法对裂缝数据集添加噪声模拟复杂环境中的裂缝图像.在识别裂缝数据集时,对改进方法与原始 SSD 算法进行对比分析.结果表明,原始 SSD 算法和改进方法识别裂缝的准确性均随噪声水平的增加而降低.在高密度下添加 20% 等级的椒盐噪声时,原始 SSD 算法识别裂缝的准确率仅为 31.7%,而改进方法的准确率则高达 93.0%.因此,改进方法具有较强的抗噪能力,可用于复杂环境下的裂缝识别.

关键词:裂缝识别;剪枝神经网络;图像加噪;抗噪性能;病害检测

中图分类号:TP391.41