Smart Structures and Systems, *Vol. 27, No. 6 (2021) 1031-1040*
DOI: https://doi.org/10.12989/sss.2021.27.6.1031

A deep learning-based vision enhancement method for UAV assisted visual inspection of concrete cracks

Yanzhi Qi [1], Cheng Yuan [1], Qingzhao Kong [1], Bing Xiong [2] and Peizhen Li*[1]

[1] *Department of Disaster Mitigation for Structures, Tongji University, Shanghai, China*
[2] *State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China*

(*Received November 13, 2020, Revised February 4, 2021, Accepted March 17, 2021*)

Abstract. Implementing unmanned aerial vehicles (UAVs) on concrete surface-crack inspection leads to a promising visual crack detection approach. One of the challenges for automated field visual cracking inspection is image degradation caused by the rain or fog and motion blur during data acquisition. The present study combines two deep neural networks to address the image degradation problem. By using the Variance of Laplacian algorithm for quantifying image clarity, the proposed deep neural networks can remarkably enhance the sharpness of the degraded images. After vision enhancement process, Mask Region Convolutional Neutral Network (Mask R-CNN) was developed to perform automated crack identification and segmentation. Results show a 8~13% enhancement in prediction accuracy compared to the degraded images, indicating that the proposed deep learning-based vision enhancement method can effectivey identify and segment concrete surface cracks from photos captured by UAVs.

Keywords: RRA-GAN; SR GAN; crack detection; Mask R-CNN; deep learning; SHM

1. Introduction

Numerous residential buildings and civil infrastructures were erected several decades ago and consequently have exceeded the design life so far. However, the safety of infrastructures is crucial for humans (Fan *et al*. 2019). Structural health monitoring (SHM) has been rapidly developed recently, and it includes structural damage detection and strategy characterization (Alla and Asadi 2020). The structural damage could be caused by concrete cracks, asphalt cracks, fatigue cracks, concrete peeling and delamination (Spencer *et al*. 2019, Liu and Zhang 2019). Cracking is a common type of damage and usually occurs on the surface of a building, which affects the structure's health and induces other problems (Gavilan *et al*. 2011, Luo *et al*. 2016). The current engineering consensus is that crack detection is crucial in the structural soundness and reliability maintenance of infrastructures (Koch *et al*. 2015, Fang *et al*. 2020, Ji *et al*. 2020).

In some cases, cracking detection still requires manual participation. However, the conventional detection method is labor-intensive, time-consuming, costly, and inefficient (Nhat-Duc *et al*. 2018, Yousaf *et al*. 2018). It also depends on engineering experience, which brings about potential perils and high maintenance costs (Yang *et al*. 2020). With the development of machine learning technology, widespread attention to image-based crack detection has been attracted in both academia and industry (Agarwal and Singh 2015, Hanzaei *et al*. 2017, Huang *et al*. 2020). Early image-processing techniques (IPTs) (Abdel-Qader *et al*. 2003, Nishikawa *et al*. 2012) used in crack detection were thresholding, edge detection, etc. To implement classification and regression, various statistical features of crack images with machine learning algorithms were used in these methods (Yang *et al*. 2020). However, they are sensitive to lighting changes and other noise in real-world conditions. To overcome the above-mentioned problems, Ji *et al*. (2020) proposed an integrated method to automatically inspect and quantify cracks in asphalt pavement for operation and maintenance at the pixel level. Wu *et al*. (2016) proposed an automated crack-recognition method to complete multi-level image-processing targets in road-rating applications. Liao and Lee (2016) used three different detection techniques to recognize objects from digital images. In addition to this research, numerous investigations have also applied machine learning techniques, such as Support Vector Machine (SVM), Histogram of Oriented Gradient (HOG) or Local Binary Patterns (LBP) (Kapela *et al*. 2015, Quintana *et al*. 2016, Varadharajan *et al*. 2014) to enhance the robustness and efficiency of IPT-based methods. However, these methods require tens of thousands of iteration steps (Ji *et al*. 2020) and still need the pre and post-processing techniques, which are time- consuming (Cha *et al*. 2017a).

Deep learning techniques are widely used for crack detection owing to its numerous advantages compared with machine learning (ML) methods. The deep learning technique can perform better in accuracy and robustness than the ML algorithm in many cases (Fadlullah *et al*. 2017, Ohsugi *et al*. 2017). Cha *et al*. (2017b) introduced Convolutional Neural Networks (CNNs) to evaluate

*Corresponding author, Professor,
E-mail: lipeizh@tongji.edu.cn

ISSN: 1738-1584 (Print), 1738-1991 (Online)

concrete cracks without calculating the defect characteristics. The designed CNN detailed about 98% detection accuracy trained on 256 × 256-pixel resolutions images. Since the transfer study has been verified to improve the accuracy and efficiency of a detection classifier, AlexNet (Krizhevsky *et al*. 2012) was proposed in a pixel-wise way to detect cracks based on CNNs (Li *et al*. 2018). VGG-net (Simonyan and Zisserman 2014) and ResNet (He *et al*. 2015) were also proposed to classify crack as pre-trained models of CNN (Bang *et al*. 2018, Silva and Lucena 2018). Although the CNNs are competitive in many object detection fields, it is hard to achieve a suitable sliding window size because the testing images may have all sorts of scales. To enhance the efficiency of learning and recognizing features from training data automatically, deep Convolutional Neural Networks (DCNN) (Dorafshan *et al*. 2018), and Region-based Convolutional Neural Networks (RCNN) (Girshick *et al*. 2014) have been proposed in field application. However, it still requires optimized model initialization values and a large number of training examples, which are difficult to achieve in actual projects.

Unmanned aerial vehicles (UAVs) with techniques based on computer vision represent prospects for crack inspection (Liu *et al*. 2020). In order to enable drones to achieve real-time crack detection (Reddy *et al*. 2019), some frameworks have been proposed accordingly. Wang *et al*. (2020) proposed a deeply supervised object detector (DSOD) to complete the identification of fatigue crack initiation sites (FCISs) from scratch. Park *et al*. (2020) used the YOLOv3-tiny algorithm to detect and qualify structural cracks, and the test results found that the accuracy of real-time damage detection was 94%. Although the crack targets can be well detected by the bounding box-based methods, they cannot provide informative crack varying angle, path, and detection since cracks appear as thin and long strips and lines (Dung and Anh,2019). Moreover, the depth and width of cracks are useful to diagnose the service life of structures (Fujio *et al*. 2020) and infrastructures maintenance (Mohamed *et al*. 2019), and this specific information can be obtained from the segmentation. Therefore, conducting segmentation is vital for crack detection. To quantify the actual crack size, Liu *et al*. (2019) used the U-Net network structure to propose a deep learning model for crack detection and semantic segmentation. Cha *et al*. (2017a) developed a model based on Faster R-CNN to detect five types of surface damages, including concrete cracks. Detailed in the study, the average precision (AP) ratings were respectively 90.6%, 83.4%, 82.1%, 98.1%, and 84.7% for five different damages. Fang *et al*. (2020) proposed a DCNN to recognize the crack orientation, and Faster R-CNN was also used to detect crack patches to enable crack segmentation. Islam and Kim (2019) used a Fully Convolutional Encoder-Decoder Network to do autonomous crack detection and obtained approximately 92% accuracy for both the recall and F1 score. He *et al*. (2018) proposed Mask R-CNN, which is an instance segmentation network. It localizes objects using a bounding box and classifies each pixel into a set of categories with differentiating object instances. Compared to the networks as mentioned earlier, Mask R-CNN performs better in precision, training speed, and segmentation. Many fields like breeding (Xu *et al*. 2020, Yu *et al*. 2020), agriculture (Ganesh *et al*. 2019, Jia *et al*. 2020), and astronomy (Ali-Dib *et al*. 2020) have put this method into use. Different from semantic segmentation algorithm, Mask R-CNN yield an individual mask for each region of interest (ROI) so that individual instance of concrete cracks can be processed separately. Due to these advantages, the algorithm was used in the experiment of this study.

Existing vision-based methods are highly susceptible to changes in visibility-related environmental impacts, such as rain or fog (Spencer *et al*. 2019). The drone lens attached to raindrops always cause the captured image be blurred, which limits this technology in field application. Meanwhile, the image degradation also decreases the accuracy of crack detection during the deep neural network training process. In this paper, A GAN and Mask R-CNN combined technology was developed to improve field captured low-quality images and perform crack segmentation. The proposed method includes three steps as follows. Firstly, the RRA-GAN was employed to remove raindrops from crack images. Secondly, the SR-GAN network was employed to sharp the blurred crack images. Finally, the processed images were trained under Mask R-CNN for crack identification and segmentation. The general framework of the proposed method is demonstrated in Fig. 1.

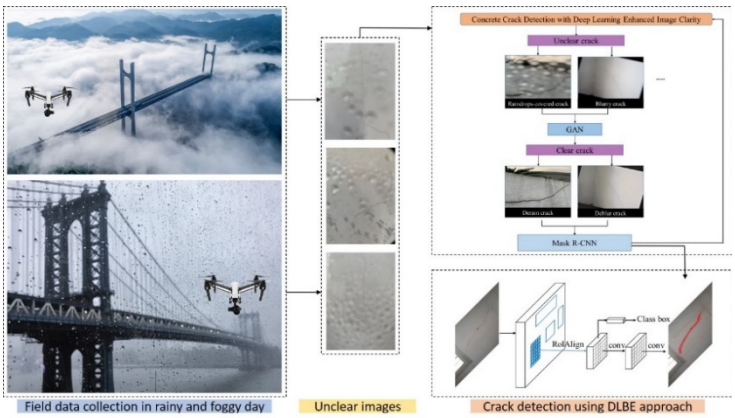

Fig. 1 Framework of the proposed method

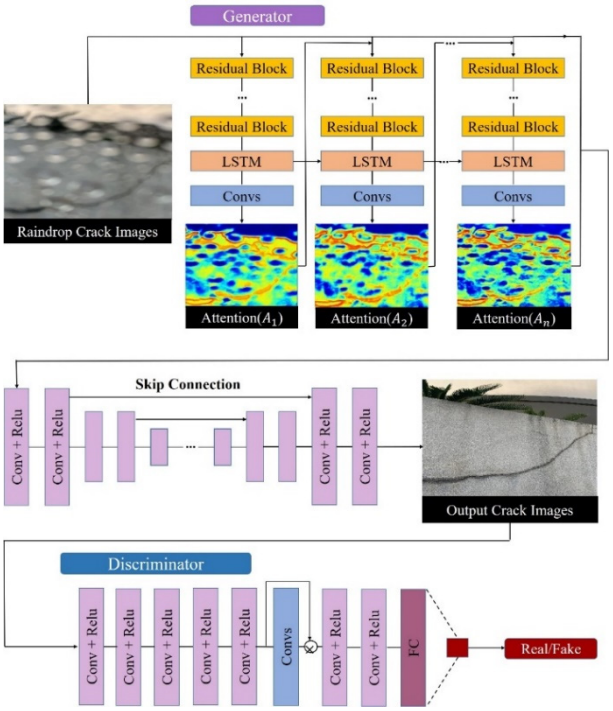

Fig. 2 The architecture of the attentive GAN network

2. Theoretical background of image enhancement

2.1 RRA-GAN network

2.1.1 Attentive generative adversarial network for raindrop removal

Raindrops adhered to UAVs' lens can severely impede the visibility of concrete cracks and significantly reduce the image's clarity. The applied attentive Generative Adversarial Network (GAN) model is adopted to deal with image degradation caused by raindrops (Qian *et al*. 2018). Fig. 2 displays the overall frame of the GAN network. Two main parts are included in the network, namely the generative deep network and discriminative deep network. The generative deep network aims to generate crack image as realistic as possible and remove raindrops. The discriminant deep network is used to verify whether the crack image generated by the generating network is real.

The generative adversarial loss function can be expressed as Eq. (1)

$$\min_G \max_D E_{R\sim P_{clean}}[\log(D(R))] + E_{I\sim P_{raindrop}}[log(1-D(G(I)))] \quad (1)$$

where G represents the generative deep network, and D represents the discriminative network, I refers to the sample extracted from the crack image dataset denoised by raindrops, which is the generative network input, and R refers to the sample extracted from the clean crack image dataset.

2.1.2 Image collection, training, and testing process

Similar to collecting the training dataset (Qian *et al*. 2018), a set of image pairs is needed. In the image pair, one is free from raindrops, and another is degraded by raindrops, which completely contains the same background scene. Both the images should be put into the RRA-GAN to train. DJI Mavic2 Zoom was used to collect the crack images pairs, as shown in Figs. 3(a)-(b). To simulate the raindrops attached to the lens on rainy days and obtain the crack image covered by the raindrops, we used some water droplets on the glass plate on the drone lens. Besides 110 pairs of crack images captured by our UAV, an open source of 864 pairs of images taken by Qian *et al*. (2018) has also been part of the training dataset to improve the robustness of the experiment. After establishing a total of 974 pairs of the corresponding image data, the GAN network was used to conduct the training, and the training time using GPU mode was approximately 20 hours with 120,000 steps. Tensorboard was used to record and visualize the relationship between loss and step.

The testing images were distinguished from the training dataset to strongly prove that the well-trained RRA-GAN network can eliminate raindrops in the crack images. These images are prepared in the same way as building the training dataset. It is found that the network pays more and more attention to the raindrop area and its related structures with the increasing of the time step from the attention map generated by the attentive-recurrent network.

Fig. 4 reveals the results of the training effect: raindrop crack images (input) and derain crack images (output). To validate and quantify the comparisons further, the Variance of Laplacian algorithm (Bansal *et al*. 2016) was employed to quantify the image clarity. It is an effective representation of a blurred image detection scheme based on the edge type

(a) The Mavic2 Zoom drone (b) The drone's working scene

Fig. 3 Equipment used in collecting crack images pairs

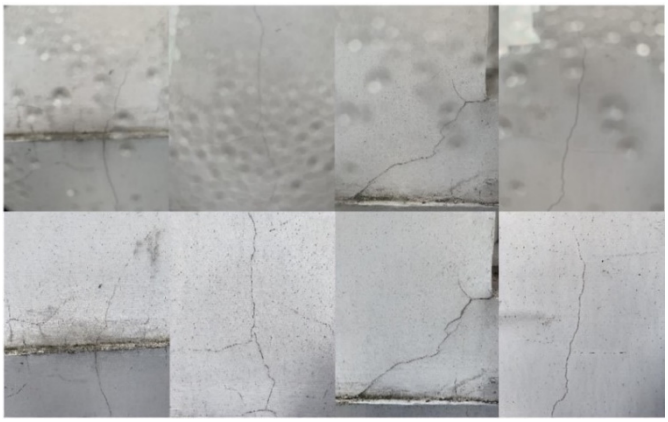

Fig. 4 Some examples of comparing raindrop-covered crack images (top) and derain crack images (bottom)

Fig. 5 The variance of Laplacian for blur detection

and sharpness analysis using the Laplacian operator, determining whether the image is blurred or not through Variance of Laplacian. The low variance of Laplacian means blurred vision, while the higher variance of the Laplacian means clearer vision. Two sets of comparison images are detailed in Fig. 5. It is found that the Variance of Laplacian is significantly improved after Derain, indicating the good effect of the proposed method.

2.2 SR GAN network

2.2.1 Super-resolution generative adversarial network for crack image deblurring

After removing raindrops on the images with concrete cracks, some pictures may still be blurred during shooting. Liu *et al*. (2020) proposed a GAN-based model to solve the motion blur caused by the platform' excessive vibration. Additionally, the clarity of images taken in some practical SHM projects may be affected by equipment, distance, light, and other environment-related issues, which would reduce the image resolution. Furthermore, smartphones may also be used to replace professional photography acquisition equipment for data collection, which would cause a reduction in image pixels. Lower image pixels are not conducive to object segmentation and recognition. To address the issue of blurring the photo mentioned above, a Super-Resolution Generative Adversarial Network (SR GAN) proposed by Ledig *et al*. (2017) was employed to deblur the images. It can recover the finer texture details when the pixel is too low, suitable for cracking pictures. As shown in Eq. (2), Goodfellow *et al*. (2014) defined a discriminator network D_{θ_D} which optimizes alternately along with G_{θ_G} to address the confrontational min-max problem

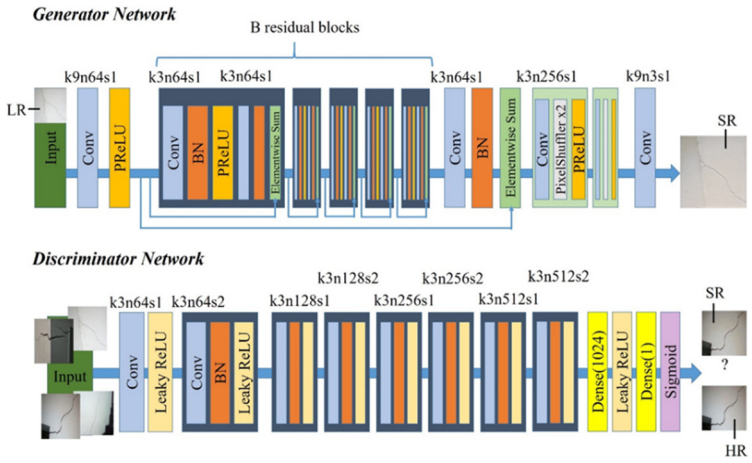

Fig. 6 The architecture of SR GAN

(a) Original image (b) Generated image (0 step) (c) Generated image (15000 steps) (d) Generated image (56000 steps)

Fig. 7 Generated images and original images in different steps

$$\min_{\theta_G}\max_{\theta_D} E_{I^{HR}\sim P_{train}(I^{HR})}\left[log\left(D_{\theta_D}\left(I^{HR}\right)\right)\right] + E_{I^{LR}\sim P_G(I^{LR})}\left[log\left(1-D_{\theta_D}\left(G_{\theta_G}\left(I^{LR}\right)\right)\right)\right] \quad (2)$$

The architecture of SR GAN also consists of the generator and discriminator network. The corresponding kernel size (k), the number of feature maps (n), and stride (s) indicated for each convolutional layer is shown in Fig. 6. The generator network G's very deep core is also illustrated in Fig. 6, namely B residual blocks. Followed by batch-normalization layers, two convolutional layers with 3×3 kernels and 64 feature maps were used (Ioffe and Szegedy 2015). The activation function choosed Parametric ReLU (He *et al*. 2015).

2.2.2 Image collection, training, and testing process

There are 234 crack images captured during the process of structural inspection for training. The pictures were taken under different light environments (strong indoor light or weak light) and distances (between 0.5 to 1.5 meters).

Approximately 10 hours were spent in the training time using GPU mode, and the corresponding steps are 60,000. Figs. 7(a)-(d) shows the comparison of generated images and original images in different steps during training. With the increase of steps, the generated crack images are closer and closer to the initial crack images, and then a trained model was obtained after training. A set of low-pixel blurred crack images was fed into the trained model, and then the model generated the deblurred images. It is found that the crack outline is clearer after using SR GAN, and the sharpness of the degraded image is also enhanced.

3. Concrete crack detection using deep neural network

3.1 Methodology

The main algorithm used to identify concrete cracks and segment the corresponding outlines is Mask R-CNN (Girshick *et al*. 2014, He *et al*. 2017). It combines the classical computer vision target detection and improved semantic segmentation. The network classifies individual objects and using a bounding box to localize them. At the same time, each pixel is classified into a set of categories with different object instances. For example, the aim is to detect concrete cracks in each image, and the classes were set as crack and background. Mask R-CNN is currently recognized as a promising method of instance segmentation. A branch for predicting segmentation masks extended by Fast R-CNN was generated on each Region of Interest (RoI), as shown in Fig. 8.

Mask R-CNN adopts a two-stage procedure. The first stage is Region Proposal Network (RPN), and the candidate object bounding boxes are proposed in this stage. The second stage is parallel with the prediction class and the box offset (Ren *et al*. 2016), which outputs a binary mask for each RoI. The loss function is defined as Eq. (3)

$$L = L_{cls} + L_{box} + L_{mask} \quad (3)$$

where L_{cls} refers to the classification loss, and L_{box} represents the bounding-box loss. L_{mask} is defined as the average binary cross-entropy loss. The definition L_{mask} allows the network to generate masks for each class without competition (He *et al*. 2017).

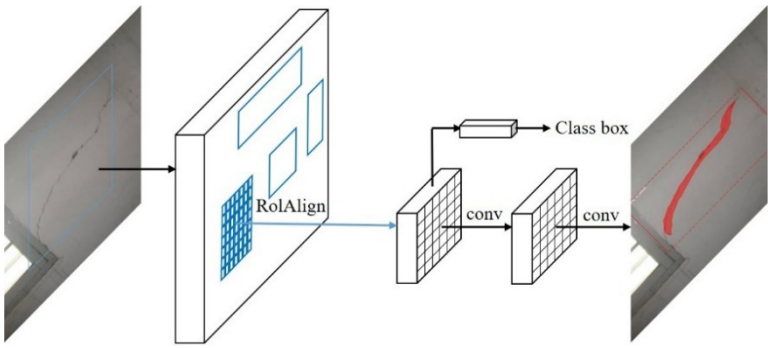

Fig. 8 The Mask R-CNN framework for instance segmentation

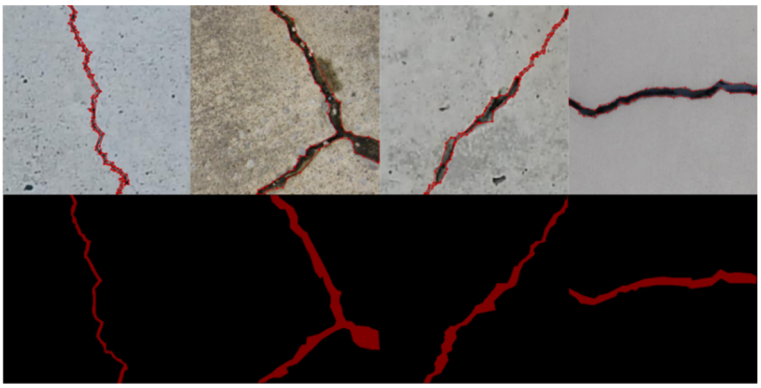

Fig. 9 Example of annotated crack images using LabelMe

3.2 Set-up of image datasets

One of the most effective approaches to enhance the performance of Mask R-CNN is to collect more data in the training set. It has been demonstrated that too little training data would result in poor approximation, and insufficient training data would lead to overfitting and generalization. For these reasons, data collection for concrete crack is divided into two parts: 40,000 concrete crack images of 227 × 227 pixels collected at various campus buildings of Middle East Technical University (Özgenel 2017) and pictures of exterior walls or high places taken by drones. Image augmentation was also used in crack images taken by ourselves to avoid overfitting. Data augmentation is an efficient way to extend datasets by introducing unobserved data or potential variables from existing datasets through different forms of physical transformations, which can be expressed as Eq. (4)

$$\phi: S \rightarrow T \tag{4}$$

where S refers to the original dataset, T defines for the dataset augmented by a transformation method $\emptyset$. The crack images taken by ourselves have been augmented by shifts, noise, rotation, and flip. Finally, a total of 613 crack images were obtained for detection and segmentation. The training, validation, and testing datasets respectively contain 400, 92, and 121 concrete crack images. The desirable features to identify should be told to the computer before asking it to learn features from the training dataset. Thus, the concrete cracks on the images should be annotated in the training datasets in advance. Feature labeling was manually carried out using LabelMe (Fig. 9), a graphical image annotation tool. After the above steps, the preliminary work of training was finished.

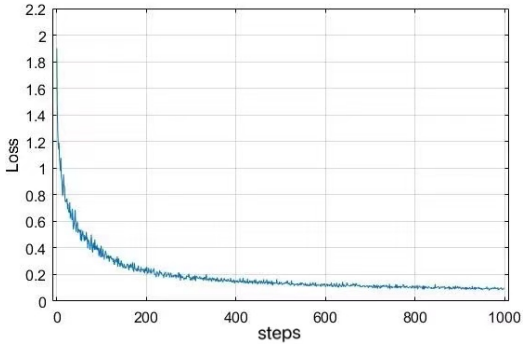

Fig. 10 The loss-step graph generated by the training process

3.3 Training process and examples of test images

The experiment was conducted by applying deep learning for Python 3.7, cuDNN 7.4.1, CUDA 10.0 on a computer with an Intel(R) Core(TM) i9-9900K@3.60GHz CPU and 11 GB NVIDIA GeForce RTX 2080 Ti. At the beginning of the training, the parameters were initialized with random values. Normalization was applied to each pixel value, so the result can converge faster. To facilitate parameter updates between iterations, the momentum was set to 0.9. The RPN anchor scales were respectively (32, 64, 128, 256, 512) and the anchor ratios were (0.5, 1, 2). At the same time, the threshold of RPN was set to 0.75. The present study set the mini batch size for network training to 4, which will contain multiple RoIs sampled from each

Fig. 11 Examples of detection and segmentation results for test images

training image in each iteration, so as to evaluate the gradient of the loss function and update the weights. The number of steps for crack detection using Mask R-CNN is 1000. Additionally, the training option was employed to ensure that the weight of the network decays to 0.0001 and the learning rate was 0.001. Compared to other training models (Ji *et al*. 2020), the loss value decreases rapidly and achieves good detection and segmentation effect, as detailed in Fig. 10. Fig. 11 displays some examples of segmentation on the test dataset, it seems crack path and varying angle can be predicted well by the proposed method in this study.

4. Results and comparisons

4.1 Confidence analysis

Since most of the objects in the dataset were elongated and irregular in shape, an overlap threshold was specified for segmentation evaluation, which was computed as mean Intersection over Union (mIoU) to classify positive and negative testing results. If the IoU was greater than or equal to the predefined threshold, the sample was classified as positive. Otherwise, it was negative. The reason for setting to overlap threshold evaluation was because the predicted segmentation mask was unlikely to exactly match the ground truth mask. In this study, the predefined threshold was set as 0.75.

Table 1 The evaluation parameters

	Ground truth	Predicted result
	Objects (positive)	Objects (negative)
Objects	True positive (TP)	False negative (FN)
Nonobjects	False positive (FP)	True negative (TN)

Fig. 12 Duplicate detection using CSRT tracker

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{6}$$

The model performance was evaluated by Precision (P) and Recall (R). In Table 1 and Eqs. (5)-(6), there are four types in samples: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). P was defined as the percentage of true positive (TP) instances in all positive detection results, while R denoted the proportion of TP to the sum of TP and false negative (FN). The standard metric for bounding box was Average Precision (area under the Precision-Recall curve), and for segmentation evaluation, Average Precision and mIoU were applied.

In practical applications, it usually needs to obtain the information of each damage through the recognition of local damage image. However, the process of crack detection doesn't show whether the damage has been detected before, which may cause follow-up issues. In this study, CSRT tracker was used as the duplicate detector (Farhodov *et al*. 2019, Zhou *et al*. 2020). After receiving the coordinates of the bounding box, CSRT subsequently tracked cracks and then marked each of them with an ID (Fig. 12). As long as the recognized bbox area and the marked frame were detected to overlap with each other, it was judged to be the same damage and was considered as a false positive (FP).

4.2 Performance verification and comparisons

The performance of the proposed method was verified for crack detection and segmentation evaluation tasks. The segmentation results of 121 test images showed that the mIoU for crack reached 83.6% and 81.2% for using Mask R-CNN with RRA-GAN and Mask R-CNN with SR GAN, much higher than 70.4% for using Mask R-CNN only. Table 2 details the comparisons of Average Precision of the mask and bbox using Mask R-CNN only and using the proposed method (Mask R-CNN with RRA-GAN or Mask R-CNN with SR GAN). It can be seen that the crack detection and segmentation performed by Mask R-CNN with RRA-GAN or SR GAN shows higher accuracy. The test AP of the mask is respectively from 74.2% up to 87.6% and 83.4%, that of the bbox is respectively from 70.6% up to 84.4% and 82.2%. As the samples displayed in Fig. 12, the segmentation is more accurate and comprehensive. Without the help of RRA-GAN, heavy rain would cause lower

Table 2 Performance of using different algorithm(s)

Average precision	Algorithm(s)	Training	Validation	Test
Segmentation mask	Mask R-CNN	77.3%	75.7%	74.2%
	RRA-GAN+Mask R-CNN	89.2%	87.4%	87.6%
	SR GAN+Mask R-CNN	84.7%	83.9%	83.4%
Bounding box	Mask R-CNN	75.8%	73.7%	70.6%
	RRA-GAN+Mask R-CNN	86.3%	84.6%	84.4%
	SR GAN+Mask R-CNN	84.4%	83.6%	82.2%

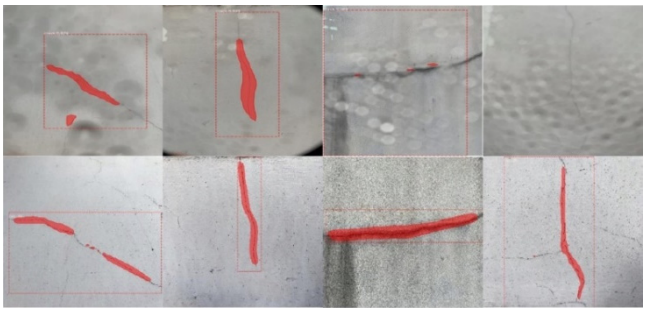

Fig. 12 Comparisons of crack detection between using Mask R-CNN only(top) and using Mask R-CNN+RRA-GAN(bottom)

(a) Using Mask R-CNN only (b) Using Mask R-CNN+SR-GAN

Fig. 13 Comparisons of crack detection between using Mask R-CNN only and using Mask R-CNN+SR-GAN

detection and segmentation accuracy of the Mask R-CNN, while after adding RRA-GAN, the concrete cracks in images could be well-identified and segmented. Figs. 13(a)-(b) shows a remarkable difference in concrete crack segmentation before and after deblur. For blurry images that have not yet eliminated the blur effect, many cracks cannot be detected and segmented. However, for blurry crack images that have been deblurred, the Average Precision was improved approximately 8%~13%, indicating that the proposed deblur approach enhances the crack detection precision.

5. Conclusions

In this study, a deep learning-based vision enhancement method was proposed for automated crack identification and segmentation from photos captured by UAVs. The existing IPT-based methods and traditional CNN or DCNN with a sliding window for crack localization are difficult to determine the window's size and require a tremendous amount of robust image samples. Though the deep learning framework offers a solution to overcome this proplem, current used frameworks such as YOLO, SSD, and Faster R-CNN cannot acchieve good segmentation. Alternatively, Mask R-CNN can effectively perform crack detection and segmentation. By the combination of GAN and Mask R-CNN, results show that the Average Precision can be improved by 8~13% compared with traditional methods.

In the future work, more samples under various environmental conditions will be trained to further improve the accuracy and robustness of the proposed method. Additionally, a video-based crack detection framework with enhanced image sharpness will be studied to provide quasi-real-time and autonomous vision-based concrete surface-crack inspection.

Acknowledgments

This research is supported by the National Key Research and Development Program of China (Grant No. 2018YFC0705602), Science and Technology Commission of Shanghai Municipality (STCSM) (Grant No. 19DZ1201200), and China National Science Foundation (Grant No. 51978507).

References

Abdel-Qader, I., Abudayyeh, O. and Kelly, M.E. (2003), "Analysis of edge-detection techniques for crack identification in bridges", *J. Comput. Civil Eng.*, **17**(4), 255-263. https://doi.org/10.1061/(ASCE)0887-3801(2003)17:4(255)

Agarwal, S. and Singh, D. (2015), "An adaptive statistical approach for non-destructive underline crack detection of ceramic tiles using millimeter wave imaging radar for industrial application", *IEEE Sensors J.*, **15**(12), 7036-7044. https://doi.org/10.1109/JSEN.2015.2469157

Ali-Dib, M., Menou, K., Jackson, A.P., Zhu, C. and Hammond, N. (2020), "Automated crater shape retrieval using weakly-supervised deep learning", *Icarus*, **345**, 113749. https://doi.org/10.1016/j.icarus.2020.113749

Alla, S. and Asadi, S.S. (2020), "Integrated methodology of structural health monitoring for civil structures", *Mater. Today: Proceedings*, **27**(2), 1066-1072. https://doi.org/10.1016/j.matpr.2020.01.435

Bang, S., Park, S., Kim, H., Yoon, Y. and Kim, H. (2018), "A deep residual network with transfer learning for pixel-level road crack detection", *Proceedings of the International Symposium on Automation and Robotics in Construction*, Berlin, Germany, July.

Bansal, R., Raj, G. and Choudhury, T. (2016), "Blur image

detection using Laplacian operator and Open-CV", *Proceedings of 5th International Conference on System Modeling & Advancement in Research Trends*, pp. 63-67, Moradabad, India, November.

Cha, Y.J., Choi, W., Suh, G. and Mahmoudkhani, S. (2017a), "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types", *Comput.-Aided Civil Infrastruct. Eng.*, **33**(9), 731-747. https://doi.org/10.1111/mice.12334

Cha, Y.-J., Choi, W. and Büyüköztürk, O. (2017b), "Deep learning-based crack damage detection using convolutional neural networks", *Comput.-Aided Civil Infrastruct. Eng.*, **32**(5), 361-378. https://doi.org/10.1111/mice.12263

Dorafshan, S., Thomas, R.J. and Maguire, M. (2018), "Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete", *Constr. Build. Mater.*, **186**, 1031-1045. https://doi.org/10.1016/j.conbuildmat.2018.08.011

Dung, C.V. and Anh, L.D. (2019), "Autonomous concrete crack detection using deep fully convolutional neural network", *Automat. Constr.*, **99**, 52-58. https://doi.org/10.1016/j.autcon.2018.11.028

Fadlullah, Z. M., Tang, F., Mao, B., Kato, N., Akashi, O., Inoue, T. and Mizutani, K. (2017), "State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems", *IEEE Commun. Surveys Tutor.*, **19**(4), 2432-2455. https://doi.org/10.1109/COMST.2017.2707140

Fan, G., Li, J. and Hao, H. (2019), "Lost data recovery for structural health monitoring based on convolutional neural networks", *Struct. Contorl Health Monitor.*, **26**(10), e2433. https://doi.org/10.1002/stc.2433

Fang, F., Li, L., Gu, Y., Zhu, H. and Lim, J.-H. (2020), "A novel hybrid approach for crack detection", *Pattern Recogn.*, **107**, 107474. https://doi.org/10.1016/j.patcog.2020.107474

Farhodov, X., Kwon, O.H., Kang, K.W., Lee, S. and Kwon, K. (2019), "Faster RCNN detection based OpenCV CSRT tracker using drone data", *Proceedings of International Conference on Information Science and Communications Technologies*, Tashkent, Uzbekistan, February. https://doi.org/10.1109/ICISCT47635.2019.9012043

Fujio, Y., Xu, C.-N., Sakata, Y., Ueno, N. and Terasaki, N. (2020), "Invisible crack visualization and depth analysis by mechanoluminescence film", *J. Alloys Compounds*, **832**, 154900. https://doi.org/10.1016/j.jallcom.2020.154900

Ganesh, P., Volle, K., Volle, K., Burks, T.F., Burks, T.F., Mehta, S.S. and Mehta, S.S. (2019), "Deep orange: Mask R-CNN based orange detection and segmentation", *IFAC-PapersOnLine*, **52**(30), 70-75. https://doi.org/10.1016/j.ifacol.2019.12.499

Gavilan, M., Balcones, D., Marcos, O., Llorca, D. F., Sotelo, M. A., Parra, I., Ocana, M., Aliseda, P., Yarza, P. and Amirola, A. (2011), "Adaptive road crack detection system by pavement classification", *Sensors* (*Basel*), **11**(10), 9628-9657. https://doi.org/10.3390/s111009628

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014), "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, OH, USA, June.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014), "Generative adversarial nets", *Proceedings of the 28th Conference on Neural Information Processing Systems* (*NIPS*), Montreal, Canada, December.

Hanzaei, S.H., Afshar, A. and Barazandeh, F. (2017), "Automatic detection and classification of the ceramic tiles' surface defects", *Pattern Recogn.*, **66**, 174-189. https://doi.org/10.1016/j.patcog.2016.11.021

He, K., Zhang, X., Ren, S. and Sun, J. (2015), "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, December.

He, K., Zhang, X., Ren, S. and Sun, J. (2016), "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June.

He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017), "Mask R-CNN", *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, October.

Huang, X., Liu, Z., Zhang, X., Kang, J. and Guo, Y.J.M. (2020), "Surface damage detection for steel wire ropes using deep learning and computer vision techniques", *Measurement*, **161**, 107843. https://doi.org/10.1016/j.measurement.2020.107843

Ioffe, S. and Szegedy, C. (2015), "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *Proceedings of the 32nd International Conference on Machine Learning* (*ICML*), Lille, France, July.

Islam, M.M. and Kim, J.M. (2019), "Vision-based autonomous crack detection of concrete structures using a fully convolutional encoder–decoder network", *Sensors* (*Basel*), **19**(19), 4251. https://doi.org/10.3390/s19194251

Ji, A., Xue, X., Wang, Y., Luo, X. and Xue, W. (2020), "An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement", *Automat. Constr.*, **114**, 103176. https://doi.org/10.1016/j.autcon.2020.103176

Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J. and Zheng, Y. (2020), "Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot", *Comput. Electron. Agric.*, **172**, 105380. https://doi.org/10.1016/j.compag.2020.105380.

Kapela, R., Sniataa, P., Turkot, A., Rybarczyk, A., Pozarycki, A., Rydzewski, P., Wyczaek, M. and Bloch, A. (2015), "Asphalt surfaced pavement cracks detection based on histograms of oriented gradients", *Proceedings of the 22nd International Conference Mixed Design of Integrated Circuits & Systems* (*MIXDES*), pp. 579-584. https://doi.org/10.1109/MIXDES.2015.7208590

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B. and Fieguth, P. (2015), "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure", *Adv. Eng. Informat.*, **29**(2), 196-210. https://doi.org/10.1016/j.aei.2015.01.008

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012), "Imagenet classification with deep convolutional neural networks", *Adv. Neural Info. Process. Syst.*, **25**, 1097-1105. https://doi.org/10.1145/3065386

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W. (2017), "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, July.

Li, Y., Li, H. and Wang, H. (2018), "Pixel-wise crack detection using deep local pattern predictor for robot application", *Sensors*, **18**(9), 3042. https://doi.org/10.3390/s18093042

Liao, K.-W. and Lee, Y.-T. (2016), "Detection of rust defects on steel bridge coatings via digital image recognition", *Automat. Constr.*, **71**(2), 294-306. https://doi.org/10.1016/j.autcon.2016.08.008

Liu, H. and Zhang, Y. (2019), "Image-driven structural steel damage condition assessment method using deep learning algorithm", *Measurement*, **133**, 168-181. https://doi.org/10.1016/j.measurement.2018.09.081

Liu, Z., Cao, Y., Wang, Y. and Wang, W. (2019), "Computer vision-based concrete crack detection using U-net fully

convolutional networks", *Automat. Constr.*, **104**, 129-139. https://doi.org/10.1016/j.autcon.2019.04.005

Liu, Y., Yeoh, J.K. and Chua, D.K. (2020), "Deep Learning–Based Enhancement of Motion Blurred UAV Concrete Crack Images", *J. Comput. Civil Eng.*, **34**(5), 04020028. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000907

Luo, D., Yue, Y., Li, P., Ma, J., Zhang, L., Ibrahim, Z. and Ismail, Z. (2016), "Concrete beam crack detection using tapered polymer optical fiber sensors", *Measurement*, **88**, 96-103. https://doi.org/10.1016/j.measurement.2016.03.028

Mohamed, Y.S., Shehata, H.M., Abdellatif, M. and Awad, T.H. (2019), "Steel crack depth estimation based on 2D images using artificial neural networks", *Alexandria Eng. J.*, **58**(4), 1167-1174. https://doi.org/10.1016/j.aej.2019.10.001

Nhat-Duc, H., Nguyen, Q.-L. and Tran, V.-D. (2018), "Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network", *Automat. Constr.*, **94**, 203-213. https://doi.org/10.1016/j.autcon.2018.07.008

Nishikawa, T., Yoshida, J., Sugiyama, T. and Fujino, Y. (2012), "Concrete crack detection by multiple sequential image filtering", *Comput.-Aided Civil Infrastruct. Eng.*, **27**(1), 29-47. https://doi.org/10.1111/j.1467-8667.2011.00716.x

Ohsugi, H., Tabuchi, H., Enno, H. and Ishitobi, N. (2017), "Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment", *Sci. Rep.*, **7**, 9425. https://doi.org/10.1038/s41598-017-09891-x

Özgenel, Ç.F. (2017), "Concrete crack images for classification", Mendeley Data, v1.

Park, S.E., Eem, S.-H. and Jeon, H. (2020), "Concrete crack detection and quantification using deep learning and structured light", *Constr. Build. Mater.*, **252**, 119096. https://doi.org/10.1016/j.conbuildmat.2020.119096

Qian, R., Tan, R.T., Yang, W., Su, J. and Liu, J. (2018), "Attentive Generative Adversarial Network for Raindrop Removal from A Single Image", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June.

Quintana, M., Torres, J. and Menendez, J.M. (2016), "A simplified computer vision system for road surface inspection and maintenance", *IEEE Transact. Intell. Transport. Syst.*, **17**(3), 608-619. https://doi.org/10.1109/TITS.2015.2482222

Reddy, A., Indragandhi, V., Ravi, L. and Subramaniyaswamy, V. (2019), "Detection of cracks and damage in wind turbine blades using artificial intelligence-based image analytics", *Measurement*, **147**, 106823. https://doi.org/10.1016/j.measurement.2019.07.051

Ren, S., He, K., Girshick, R. and Sun, J. (2016), "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Transact. Pattern Anal. Mach. Intell.*, Barcelona, Spain, December.

Silva, W.R.L.D. and Lucena, D.S.D. (2018), "Concrete cracks detection based on deep learning image classification", *Proceedings of The 18th International Conference on Experimental Mechanics*, **2**(8), 489. https://doi.org/10.3390/ICEM18-05387

Simonyan, K. and Zisserman, A. (2014), "Very deep convolutional networks for large-scale image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June.

Spencer Jr., B.F., Hoskere, V. and Narazaki, Y. (2019), "Advances in computer vision-based civil infrastructure inspection and monitoring", *Eng.*, **5**(2), 199-222. https://doi.org/10.1016/j.eng.2018.11.030

Varadharajan, S., Jose, S., Sharma, K., Wander, L. and Mertz, C. (2014), "Vision for road inspection", *Proceedings of IEEE Winter Conf. on Applications of Computer Vision* (*WACV*), Steamboat Springs, CO, USA, March, pp. 115-122. https://doi.org/10.1109/WACV.2014.6836111

Wang, S.Y., Zhang, P.Z., Zhou, S.Y., Wei, D.B., Ding, F. and Li, F.K. (2020), "A computer vision based machine learning approach for fatigue crack initiation sites recognition", *Computat. Mater. Sci.*, **171**, 109259. https://doi.org/10.1016/j.commatsci.2019.109259

Wu, L., Mokhtari, S., Nazef, A., Nam, B. and Yun, H.-B. (2016), "Improvement of crack-detection accuracy using a novel crack defragmentation technique in image-based road assessment", *J. Comput. Civil Eng.*, **30**(1), 04014118. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000451

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A. and Schneider, D. (2020), "Automated cattle counting using Mask R-CNN in quadcopter vision system", *Comput. Electron. Agric.*, **171**, 105300. https://doi.org/10.1016/j.compag.2020.105300

Yang, Q., Shi, W., Chen, J. and Lin, W. (2020), "Deep convolution neural network-based transfer learning method for civil infrastructure crack detection", *Automat. Constr.*, **116**, 103199. https://doi.org/10.1016/j.autcon.2020.103199

Yousaf, M.H., Azhar, K., Murtaza, F. and Hussain, F. (2018), "Visual analysis of asphalt pavement for detection and localization of potholes", *Adv. Eng. Info.*, **38**, 527-537. https://doi.org/10.1016/j.aei.2018.09.002

Yu, C., Fan, X., Hu, Z., Xia, X., Zhao, Y., Li, R. and Bai, Y. (2020), "Segmentation and measurement scheme for fish morphological features based on Mask R-CNN", *Info. Process. Agriculture*, **7**(4), 523-534. https://doi.org/10.1016/j.inpa.2020.01.002

Zhou, Z., Wu, Q.J., Wan, S., Sun, W. and Sun, X. (2020), "Integrating SIFT and CNN feature matching for partial-duplicate image detection", *IEEE Transact. Emerg. Topics Computat. Intell.*, **204**(5), 593-604 https://doi.org/10.1109/TETCI.2019.2909936

BS