applied sciences

Article

Crack Location and Degree Detection Method Based on YOLOX Model

Linlin Wang [1], **Junjie Li** [1,2,*] **and Fei Kang** [1,*]

[1] Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China
[2] College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China
[*] Correspondence: lijunjie@dlut.edu.cn (J.L.); kangfei@dlut.edu.cn (F.K.); Tel.: +86-0411-84708516 (F.K.)

Abstract: Damage detection and evaluation are concerns in structural health monitoring. Traditional damage detection techniques are inefficient because of the need for damage detection before evaluation. To address these problems, a novel crack location and degree detector based on YOLOX is proposed, which directly realizes damage detection and evaluation. Moreover, the detector presents a superior detection effect and speed to other advanced deep learning models. Additionally, rather than at the pixel level, the detection results are determined in actual scales according to resolution. The results demonstrate that the proposed model can detect and evaluate damage accurately and automatically.

Keywords: deep learning; crack location; object detection; crack assessment; YOLOX

check for updates

Citation: Wang, L.; Li, J.; Kang, F. Crack Location and Degree Detection Method Based on YOLOX Model. *Appl. Sci.* **2022**, *12*, 12572. https://doi.org/10.3390/app122412572

Academic Editor: Mahmoud Bayat

Received: 15 November 2022
Accepted: 6 December 2022
Published: 8 December 2022

1. Introduction

Automatic structural health monitoring (SHM) and maintenance have always been a challenging research field because of the aging of concrete structures (such as bridges, buildings, roads, and dams). The causes of concrete structural damage are various and complex, for example, earthquake, freeze injury, salt corrosion, and dry shrinkage [1]. As one of the earliest signs of structural damage, cracks on a concrete surface can be said to be the core detection part of the SHM [2,3]. Therefore, the timely detection and assessment of concrete surface cracks can better grasp the state of a concrete structure to effectively prevent the occurrence of large-scale concrete structure accidents [4].

Currently, the traditional methods for crack inspection are still visual assessment techniques relying on inspectors. They are inefficient and expensive [2,5–7]. Moreover, the assessment results are highly subjective [8,9]. Thus, owing to the inevitable drawbacks of state-of-the-art methods, efficient and economical crack detection technology is urgently needed.

In recent years, using image processing in crack detection is an appropriate way to address the above problems, and this technique is highly applied in the field of civil engineering [1,2,6,10,11]. Both Jahanshahi et al. [10] and Koch et al. [12] have reviewed the image processing methods for crack detection and evaluation of concrete structures. Generally, image processing technologies are practical and valuable. Although this research has made significant progress, there are still some shortcomings. They are too dependent on image characteristics and easily recognize noise (light, shadow, smudge, etc.) as cracks, which are not conducive to retrieving crack properties. Furthermore, most of the existing methods focus on identifying cracks in one single image. If no crack is recognized or the noise is incorrectly recognized as a crack, the detection results cannot be corrected in time. To tackle these issues, an automatic detection and evaluation system needs to be developed.

Deep learning technologies have been widely used in crack detection of various structures, such as roads [13,14], buildings [15], bridges [16,17], tunnels [18,19], and dams [20]. For instance, Zhang et al. [21] proposed an approach to classify each image patch in the obtained road crack images by training a supervised deep convolutional neural network

(DCNN). However, this method ignores the complexity of image regions [22] and always overestimates crack width. Cha et al. [23] adopted a sliding window to divide the image into blocks and used a convolutional neural network (CNN) to classify whether the block contains cracks or not. However, the computational cost of this approach is high because the CNN classifier needs to be applied multiple times to each window in every image. Chen and Jahanshahi [24] combined a CNN and Naïve Bayes data fusion scheme to identify cracks in nuclear power plants through analysis of individual video frames. Nevertheless, this method only locates the crack position without quantifying the properties of cracks. In addition, when the amount of images is small, the system will not work well. Gopalakrishnan et al. [25] applied a transfer learning (TL) technique to perform crack detection on the pavement. The VGG16 network was first pre-trained on an existing dataset and then automatically detected cracks from the FHWA/LTPP dataset by using transfer learning. This is a successful attempt to apply TL in crack detection. Zhang et al. [26] presented an automatic pavement crack detection system based on CNN. This system, called CrackNet, predicted the category of all pixels to complete crack detection. However, the network architecture of CrackNet strictly depends on the input image size [27]. Kim et al. [28] utilized Region CNN (R-CNN) based on a sliding window as a target detector to detect cracks in one concrete bridge. However, R-CNN is susceptible to the region and is inefficient compared with the current deep learning models. Dorafshan et al. [29] compared the DCNN and some traditional edge detectors for image-based concrete structural crack detection. It concluded that the AlexNet DCNN is better at crack detection. Yet the superior model may be further improved when adopting some new networks. Jang et al. [30] offered a CNN-based crack detection technique by combining vision and infrared thermography data which can minimize the false detection rate. Guo et al. [31] introduced a deep-width network (DWN) architecture to classify cracks without handcraft feature extraction. Although this method avoids hyperparameter tuning to improve efficiency and accuracy, the algorithm is complicated, which limits its development and extensive application. Li et al. [32] detected concealed cracks from images using deep learning models. This research compared and discussed some object detection models. However, it does not apply a high-performance GPU, which affects the detection efficiency. In summary, the above research proves that deep learning is excellent in crack identification. However, detecting structural cracks is only the beginning of SHM, as researchers and inspectors often pay more attention to the state of detected cracks. It is not enough to just locate the crack but not identify the damage degrees of the crack. Thus, it is essential to design an application that can directly obtain crack condition information and crack position.

With the fast growth of deep learning methods, many excellent deep learning technologies have been proposed. Concrete structural crack detection includes target detection, which can identify and locate cracks. Object detection techniques based on deep learning could be classified into two types: two-stage approaches and one-stage approaches. Two-stage approaches, for example, R-CNN [33], Fast R-CNN [34], and Faster R-CNN [35], first use the image information to decide the potential position of the target, and then CNNs is used to classify and retrieve the characteristics from these positions. Although this method provides high accuracy, the detection speed is slow. In contrast, one-stage methods use end-to-end CNN to estimate the bounding box position and type of objects in multiple positions. The detection speed can be greatly accelerated because of the simplified detection process. The main representatives of one-stage approaches are You Only Look Once (YOLO) series algorithms [36–41], Single Shot Multibox Detector (SSD) [42], and Deconvolutional Single Shot Detector (DSSD) [43].

The YOLO series as a popular one-stage detection method is applied extensively. YOLOv1 algorithm was originally proposed by Joseph Redmon et al. [36] in 2016. It forms an end-to-end CNN by integrating each independent step in the process of object recognition. Object detection is regarded as a regression process, and the object location and category of the detection image can be obtained by processing the input image. The detection speed of YOLOv1 is fast, but the detection accuracy is low. Subsequently, a series of YOLO algorithms are proposed (YOLOv2 [37], YOLOv3 [38], and YOLOv4 [39]) with the YOLO methods developing and improving rapidly. They gradually balance the detection efficiency and the precision detection. In addition, the detection results of small targets grow better. Recently, two novel YOLO models were updated, that is YOLOv5 [40] and YOLOX [41]. These two versions combine some progressive methods, which have more advantages and more efficiency in real-time image processing. At present, there are few studies on damage recognition using the YOLOv5 or the YOLOX. Consequently, the application effect of newer YOLO series models in concrete damage remains to be further investigated.

In this study, an algorithm for detecting and evaluating different degrees of cracks in images is proposed. The defects in the image are labeled according to the width of cracks, which compose a new image dataset with different degrees of crack. Using the latest YOLOX deep learning models, the crack can not only be extracted and localized accurately but also be evaluated automatically. Moreover, the integration of crack detection and crack assessment is an innovative trial. Additionally, some advanced deep learning models, such as Faster R-CNN, YOLOv5, and DSSD, are comprehensively compared to the proposed method. Furthermore, the pixel-level crack degree obtained by the proposed method is converted to actual scale through resolution. Results show that the proposed approach based on YOLOX performs outstandingly in detection accuracy and detection speed.

The remainder of this paper is organized as follows. Section 2 starts with a brief overview of the proposed model. Section 3 provides dataset construction, which introduces the composition of the image acquisition system and dataset acquisition and labeling. Experimental results and analysis are presented in Section 4, which consists of model evaluation metrics, training setting, training results and analysis, test results, and discussion and crack degree transformation test and results. Section 5 summarizes the conclusion and suggestions for future work.

2. Methodology

To quickly assess the crack position and degree, this paper proposes a crack location and degree detector based on YOLOX. This algorithm can detect cracks with different degrees automatically and quickly. Figure 1 shows the flow chart of the method, which is divided into four main parts: (1) dataset construction, which collects data using an image acquisition system and labels these crack images for training; (2) detector training, which is the process of finding the optimal hyperparameters of the model; (3) crack location and degree detection, which outputs model predictions; and (4) actual crack damage assessment, which converts crack degrees in pixel level to actual scale. The detection principle and the transformation of damage levels are explained in the following subsections.

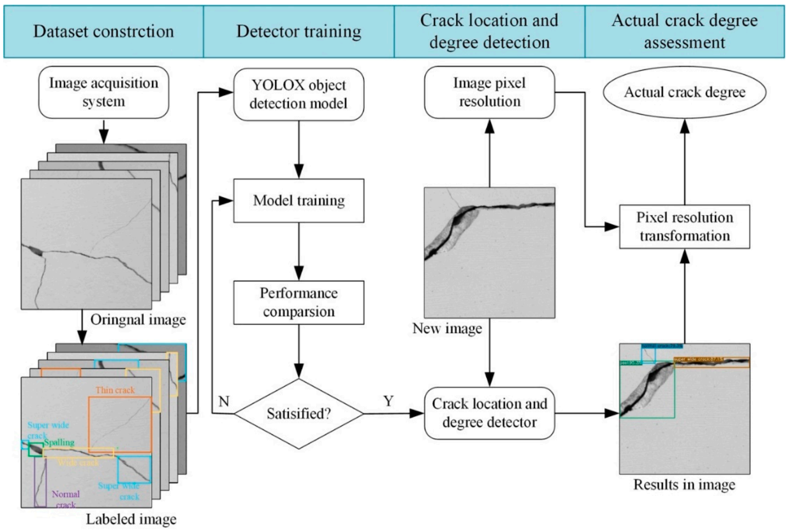


Figure 1. Overview of crack location and degree detection algorithm.

2.1. The Detection Principle of the Proposed Detector

The crack location and degree detector proposed in this study is based on the latest YOLOX model, which cleverly combines some excellent target detection methods. YOLOX is a classical one-stage detection network. Unlike two-stage approaches (such as Faster R-CNN), YOLO algorithm turns the detection problem into the regression problem, where the bounding boxes and confidence of objects can be obtained immediately. By changing the width and depth of the backbone network, four versions of the YOLOX model are obtained, which are YOLOXs, YOLOXm, YOLOXl, and YOLOXx. Figure 2 describes the architecture of YOLOX. It is composed of four general parts, including the input part, backbone network, neck network, and prediction part, corresponding to the red dotted rectangular box in Figure 2. Among them, the CBL is the basic module composed of a convolution layer (Conv), a batch normalization layer (BN), and a Leaky ReLU activation function. Resunit is a residual structure to be utilized to deepen neural network, which contains 2 CBL modules and an Add layer; the Add layer realizes tensors stacked directly, whereas the Concat layer stitches tensors with expanding dimensions of tensors; the spatial pyramid pooling (SPP) [44] is mainly for multi-scale fusion; other network modules are introduced later.

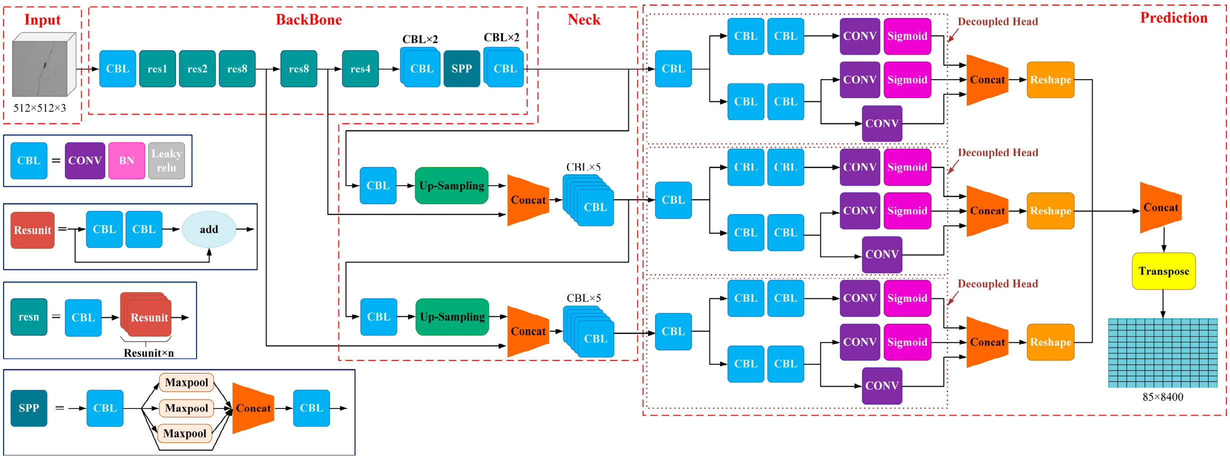


Figure 2. The architecture of YOLOX.

2.1.1. Input Part

The input part is the input image, including data augmentation and adaptive anchor box calculation, which can improve the training speed and accuracy.

YOLOX applies Mosaic and Mixup data augmentation methods. The Mosaic data augmentation method [39] mixes four training images into the image in Figure 3, which are stitched according to random scaling, random clipping, and random arrangement. The Mixup data augmentation method fuses two images by setting weight fusion coefficients, as shown in Figure 4. Both of the two data augmentation methods not only enrich the dataset, but also reduce the need for large memory.

Adaptive anchor box calculation: In the network training, the model outputs the corresponding prediction box based on the initial anchor box, calculates the difference between it and the ground truth box, and performs the reverse update operation to iterate the parameters of the whole network. Thus, setting the initial anchor box is also a key step. The optimal anchor box in different training sets is calculated adaptively during each training.

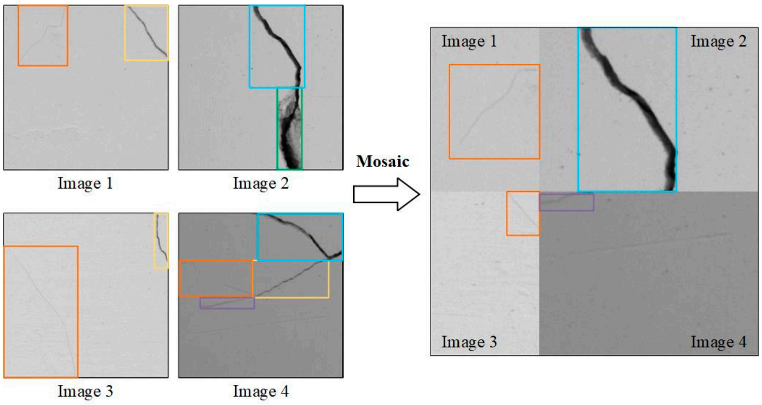


Figure 3. Mosaic data augmentation.

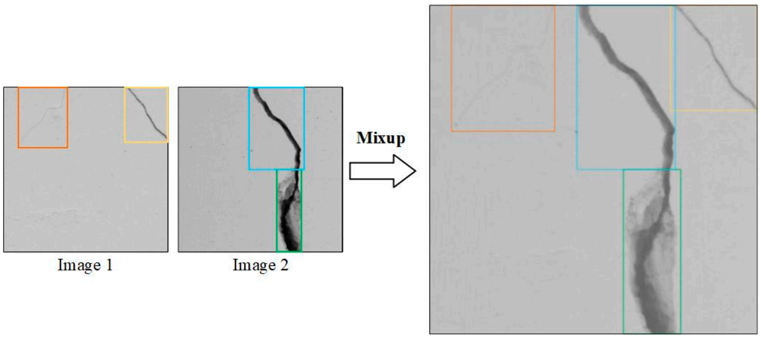


Figure 4. Mixup data augmentation.

2.1.2. Backbone Network

A backbone network is mainly used to retrieve image features. The backbone network of YOLOX selects the backbone network using Darknet53, and the ResX module consists of the CBL module and Resunit.

2.1.3. Neck Network

The neck network is usually between the backbone and head networks. Using this part can detect some complex features and further improve the diversity and robustness of features.

2.1.4. Prediction Part

The prediction part completes the output task of target detection with the help of the decoupled head structure, which greatly improves the convergence speed of the model. Moreover, the decoupled head is very significant for the end-to-end version of YOLO. Figure 5 shows the decoupled head, which contains a 1 × 1 Conv for channel dimensionality reduction, followed by two parallel branches with two 3 × 3 Conv, respectively.

In addition, YOLOX adopts the anchor-free detector again, which avoids the problems of poor generalization ability and high complexity of the anchor mechanism. Furthermore, for the label assignment problem of the algorithm, YOLOX judges the candidate anchor box based on the center point and target box to realize the preliminary screening. Then, YOLOX uses the simple optimal transport assignment (SimOTA) to realize a fine screening. This strategy first calculates pairwise matching degree, that is, the cost or quality of matching for each prediction box and the ground truth box, as shown in Equation (1):

$$\mathrm{c}_{ij} = L_{ij}^{cls} + \lambda\, L_{ij}^{reg} \tag{1}$$

where λ is a balancing coefficient, L_{ij}^{cls} and L_{ij}^{reg} are classification loss and regression loss between ground truth boxes and prediction boxes. Finally, an approximate solution is obtained by the dynamic top-k strategy. The corresponding grids of these positive predictions are designated as positive, whereas the remaining grids are negative. SimOTA used in YOLOX not only accelerates the training speed but also reduces additional solver hyperparameters.

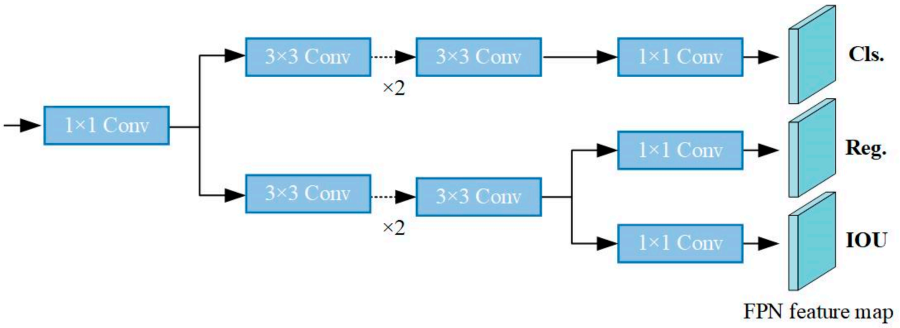


Figure 5. Decoupled head of YOLOX.

2.1.5. The Loss Function of YOLOX

The loss function of YOLOX consists of three parts, which are the loss of predicting the bounding box (L_{iou}), the classification (L_{clc}), and the object position (L_{obj}).

$$Loss = L_{iou} + L_{clc} + L_{obj} \tag{2}$$

In this paper, YOLOX adopts a generalized intersection over union loss (GIOU_Loss) [45] as the bounding box region loss, which effectively solves the problem of noncoincidence of the bounding box and improves the speed and accuracy of prediction box regression.

The calculation equation GIOU_Loss is as follows [45]:

$$\text{GIOU} = \text{IOU} - \frac{|A_c - U|}{|A_c|} \tag{3}$$

where IOU is the intersection of two rectangular areas divided by the union. A_c is the minimum box covering the predicted bounding box and ground-truth bounding box. U is the union of the predicted bounding box and ground truth bounding box. So, the loss function is:

$$L_{iou} = L_{GIOU} = 1 - \text{GIOU} \tag{4}$$

2.2. The Transformation of Crack Degree

The output of the crack location and degree detector is the pixel-level crack degree, and the actual crack degree needs to be obtained by resolution.

According to the imaging principle, the resolution K can be calculated from the field of view (FOV) (or target size) and the image resolution (or the number of pixels of the target). So the actual crack degree, that is, the crack width W_c is:

$$W_\text{c} = W_\text{p}\ K \tag{5}$$

where W_p is the crack pixel width, which is obtained by the crack position and degree detector.

3. Dataset Construction

In this work, an indoor image acquisition system is first built to collect crack images, and then these images are artificially labeled. The pairwise images and labels are used to train the crack location and degree detector.

3.1. Image Acquisition System

To obtain the required dataset and verify the application of the proposed method, an image acquisition system is built as shown in Figure 6. It is composed of a high-speed industrial camera, fill lights, a tripod, a slide rail, a laser rangefinder, and a computer.

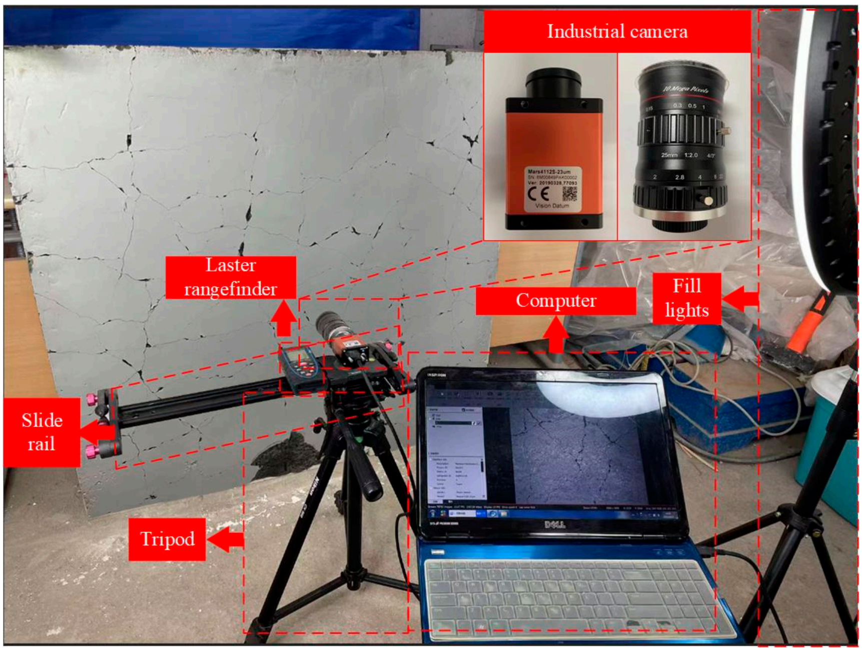


Figure 6. The image acquisition system.

The selection of industrial cameras is mainly based on resolution. If the resolution is too low, it will be hard to acquire high-resolution images, which is easy to cause thin cracks to be difficult to identify or misjudge. The image acquisition system in this paper adopts the high-speed industrial camera MARS4112S-23UM, equipped with a LEM2520CD-H2 industrial lens, and the main parameters are shown in Table 1. To avoid ghosting and darkness, this study tries to acquire crack images in the daytime with sufficient sunshine. Dual LED ring fill lights are used to ensure uniform illumination. The working distance is measured by a laser rangefinder. Fixing the acquisition devices on the slide rail can help to keep the images collected at the same working distance.

Table 1. The main parameters of the industrial camera.

Brand	Model	Main Parameters
Vision Datum Mars	MARS4112S-23UM	Sensor: 1.1″ CMOS Pixel size: 3.45 × 3.45 μm Resolution: 4096 × 3000
Vision Datum Mars	LEM5020CD-H2	Focal length: 25 mm M.O.D: 0.15 m

3.2. Establish Image Dataset with Labels

The dataset used in this paper is collected and produced by the authors. In the laboratory, a 1 × 1 × 0.05 m C20 concrete slab with reinforcement mesh was made and many different degree cracks were artificially made on the concrete slab. Figure 7a,b shows the complete concrete slab and the damaged concrete slab, respectively.

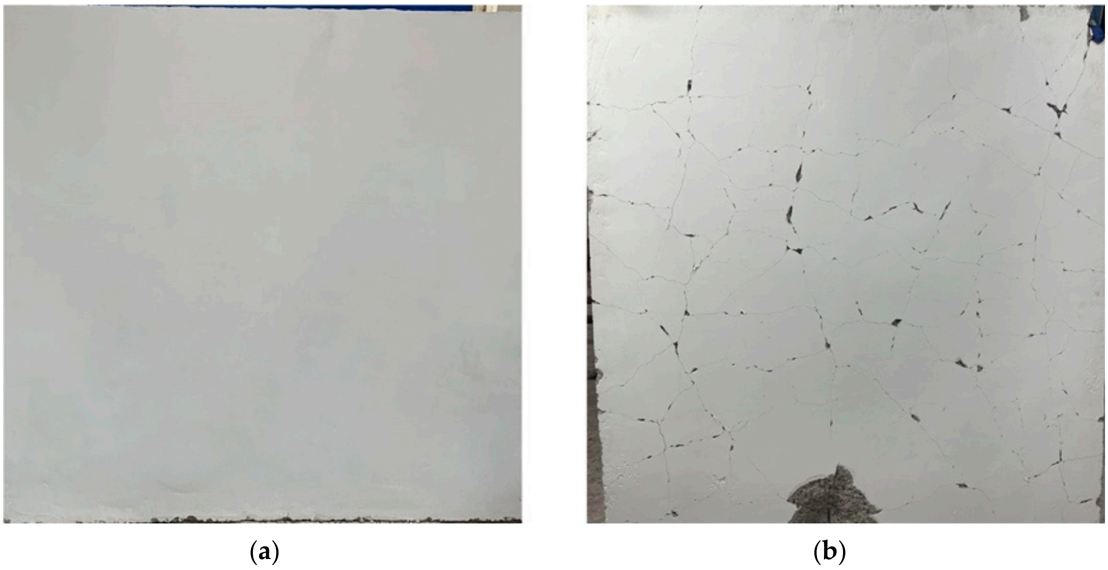


Figure 7. Complete concrete slab and concrete slab with damage. (**a**) The complete concrete slab; (**b**) the damaged concrete slab.

The concrete crack images are collected by the image acquisition system. The camera is perpendicular to the surface of the structure. The distance between the camera and the concrete slab (the working distance) is 0.35 m and the FOV is 144 mm × 197 mm. A total of 150 crack images of resolution 3000 × 4096 pixels were acquired. In order to reduce the computational cost of training, 150 images were cropped into small images with a resolution of 500 × 500 by automatic segmentation. This research aims at detecting the concrete surface crack of different scales, so 2177 images with defect characteristics were finally selected as the final image dataset. In this dataset, these 2177 images were divided into 3 parts in proportion randomly: 1525 images were selected as the training set, 435 images were selected as the validation set, and 217 images were selected as the test set, and the ratio is 7:2:1. The training set is used to train the network model, the validation set is used to test the model during training, and the test set is used for testing after the network model training is completed.

The LabelImg software is used to manually label the dataset images to generate some Pascal VOC format [46] samples. In general, the maximum allowable width of the crack cannot be over 0.3 mm, and the crack width shall not be greater than 0.2 mm in special situations. According to this, the structural surface crack category is marked into 5 types: super wide crack, wide crack, normal crack, thin crack, and spalling. The detailed description of each category is shown in Table 2. Therefore, super wide crack, wide crack and spalling can be the focus of structural health monitoring. The normal crack and thin crack may generate into critical damage, which can also be used as long-term monitoring objects. The crack width is measured by an HC-CK103 crack width gauge (as shown in Figure 8a). The equipment has a self-calibration ability and is easy to operate and reliable. The main parameters are shown in Table 3. Figure 8b presents crack width measurements recorded on a concrete slab.

Table 2. The structural surface crack category and characteristics.

Crack Type	Super Wide Crack	Wide Crack	Normal Crack	Thin Crack	Spalling [1]
Crack characteristics	Crack width > 0.3 mm	Crack width between 0.2 mm and 0.3 mm	Crack width between 0.1 mm and 0.2 mm	Crack width ≤0.1 mm	Block Spalling

[1] The authors label spalling as peel, that is, the type of peel damage is spalling.

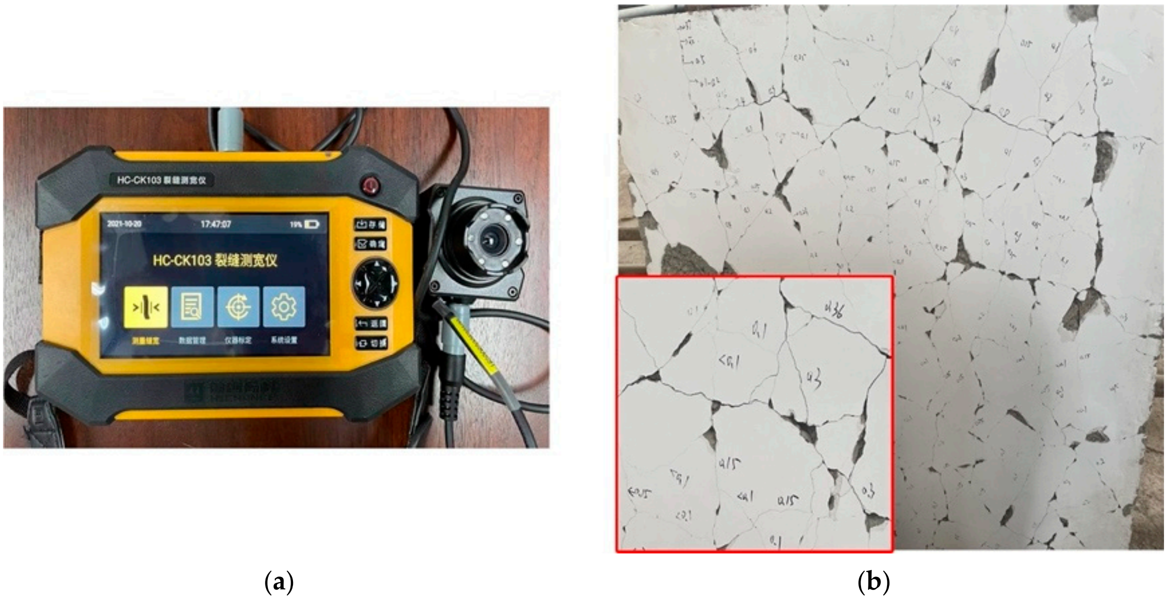


Figure 8. The crack width measuring instrument and the concrete slab with width marks. (**a**) The crack width measuring instrument; (**b**) the concrete slab with width marks.

Table 3. The main parameters of the crack width measuring instrument.

Brand	Model	Test Range	Measurement Accuracy	Magnification
HaiChuangGaoKe	HC-CK103	0~8 mm	0.01 mm	40

The number of labels and box position information marked on 2177 images is 10,388 in total, including 2102 super wide crack, 1993 wide crack, 1856 normal crack, 2321 thin crack, and 2116 spalling. Sample images with marked bounding boxes and labels are shown in Figure 9.

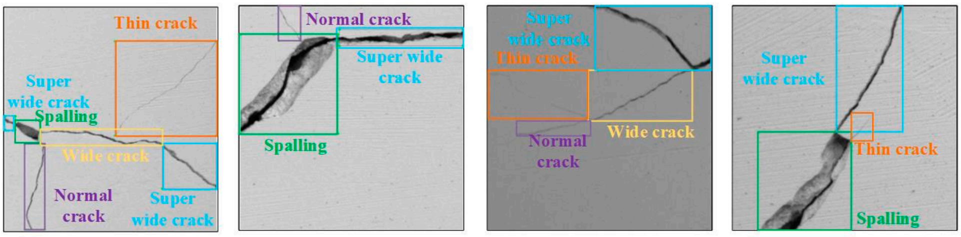


Figure 9. Sample images with marked bounding boxes and labels.

4. Experimental Results and Analysis

This section first introduces the metrics of model evaluation and some settings of the training model and then analyzes the performance of the trained model to find the optimal model result. Next, the test results of the constructed detector are compared with other classic models. Finally, the effectiveness of the proposed method is verified by a crack degree transformation test.

4.1. Model Evaluation Metrics

In this paper, the performance of the deep learning networks is comprehensively evaluated by (1) average precision (AP) and mean average precision (mAP) and (2) detection speed and inference time.

4.1.1. AP and mAP

First, IOU is a measurable standard in the accuracy of detecting corresponding targets in a specific dataset. According to the experimental results, if the crack type is positive (P) and the model detection is P, it is marked as a true positive. In general, when IOU ≥ 0.5, the detection result can be considered as a true positive (TP). Similarly, if the crack type is negative (N) but the model detection is P, it is marked as a false positive (FP), that is, when IOU < 0.5, the detection result is considered as a false positive. If the crack type is P, but the model does not detect the crack type, it is marked as a false negative (FN), and the model does not detect the ground truth box at this time. If the crack type is N but the model detection is correct, it is marked as a true negative (TN).

Then, Precision and Recall are defined, which are two basic evaluating indicators in object detection, as shown in Equations (5) and (6). Precision represents the proportion of correctly recognized targets among all detected targets, while Recall represents the proportion of correctly recognized targets among all detected positive samples:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

where TP is the number of the crack type correctly recognized, TN is the number of the non-crack type correctly recognized, FP is the number of the non-crack type regarded as the crack type, and FN is the number of the crack type regarded as the non-crack type.

Finally, AP, as shown in Equation (8), could be calculated by the area under the Precision and Recall curve, which completely considers the impact of the Precision and Recall. In addition, the average value of each crack type of AP in the data set is mAP, which can reflect the accuracy of crack detection. Additionally, mAP can be used as a criterion for the comparison among difference detection models.

$$\text{AP} = \int_0^1 \text{P}(\text{R})\, \text{dR} \tag{8}$$

4.1.2. Detection Speed and Inference Time

The frame per second (FPS) is used to evaluate the detection speed, that is, the number of images processed per second. When the FPS of the network is over 30, it is considered to have real-time detection capability [47]. Moreover, this study also uses inference time to evaluate processing speed of the model, which is the time consumed to deal with an image.

4.2. Training Setting

The training platform is based on the PyTorch framework and all experiments are performed using one GPU mode workstation equipped with the following configuration: i9-10900X CPU @ 3.70 Hz RAM 64 G, NVIDIA Geforce GTX 2080Ti GPU. The software configuration is as follows: Ubuntu 18.04, CUDA10.0, cuDNN 7.5, Python 3.7, Pytorch 1.1.

Most of the parameters of the YOLOX model adopt the default initial parameters. Some main hyperparameters are set as follows: the weight_decay coefficient is 0.0005; the momentum is 0.937; the batch size is 8; and the epoch is 100. It is worth mentioning that YOLOX series models require that the size of the input image must be a multiple of 32, so the network automatically modifies the input image size from 500 × 500 pixels to 512 × 512 pixels.

4.3. Training Results and Analysis

There are four standard versions of YOLOX, namely YOLOXs, YOLOXm, YOLOXl, and YOLOXx. Each version holds the same network structure, but the depth and width are different. The model depth and width can be set by the pre-weights required for training.

The initial learning rate (lr) is an important hyperparameter of the YOLOX model, which affects the accuracy and convergence speed by controlling the step size of the weight update. If the learning rate is too large, it is likely to exceed the optimal value and make the loss function unable to converge. If the learning rate is too small, the network optimization efficiency is too low, the loss function cannot converge for a long time, and it is easy to make the network fall into local optimization. Therefore, an appropriate learning rate needs to be found through continuous attempts, which does not only ensure model convergence as soon as possible but also makes the model recognition effect the best.

This study discusses the impact of different initial learning rate settings on the performance of four YOLOX models and selects the optimal result file as the crack location and degree detector. The initial learning rates are set to 0.01, 0.001, and 0.0001, respectively, and the validation results of the model after training with the same parameters are shown in Figure 10. YOLOX models obtain the best results on the validation set when the initial learning rate is 0.001, with mAP values of 88.57% (YOLOXs), 89.39% (YOLOXm), 90.05% (YOLOXl), and 91.17% (YOLOXx), respectively. Therefore, the proposed detector is trained with an initial learning rate of 0.001.

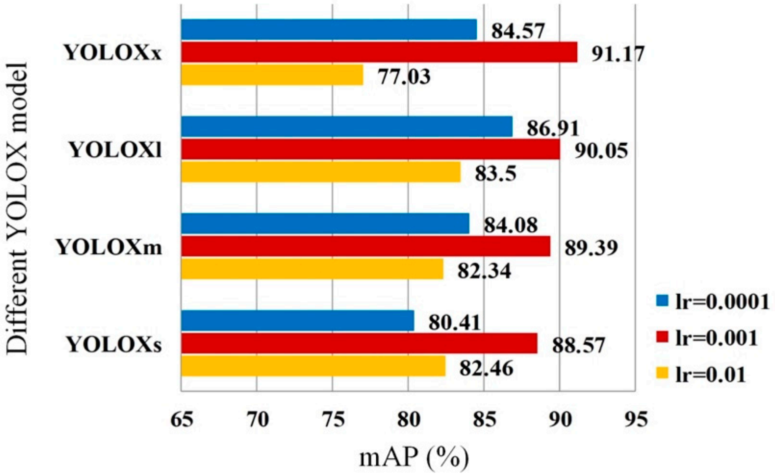


Figure 10. Validation results of the YOLOX model at different learning rates.

4.4. Test Results and Discussion

To verify the superiority of using the YOLOX model as a crack location and degree detector, this study comparatively analyzed the recognition performance of 10 models with 4 different deep learning networks, i.e., YOLOX (YOLOXs, YOLOXm, YOLOXl, and YOLOXx), Faster R-CNN, DSSD, and YOLOv5 (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x). To ensure fairness, this study is calculated on the same dataset and training platform, and each model has been effectively trained, and their parameters are selected with the best choice.

Table 4 lists all testing results of different deep learning models, including mAP value to measure the detection effect, FPS and inference time to measure the real-time performance and the model training time to measure the complexity. In general, the mAP values of YOLOX series models all exceed 88.5%, especially the YOLOXx. The best mAP value is 91.05% in YOLOXx. Even in YOLOXs the mAP value is 88.71%. These results demonstrate the applicability of YOLOX series models in the detection of cracks with different degrees. In contrast, the mAP value of the Faster R-CNN is only 69.77%, the mAP value of the YOLOv5 is the highest at 88.11%, and the mAP value of the DSSD is slightly higher at 88.92%, but other performance indicators are bad. Therefore, the YOLOX model has significantly higher accuracy in identifying the location and extent of cracks.

Table 4. Performance comparison of different deep learning models.

Model	mAP (%)	FPS	Inference Time (ms)	Training Time (h)
YOLOXs	88.78	284.90	3.51	0.583
YOLOXm	89.98	159.74	6.26	0.900
YOLOXl	90.17	112.36	8.90	1.517
YOLOXx	91.05	68.49	14.60	2.533
Faster R-CNN	69.77	22.69	44.07	1.867
DSSD	88.92	97.47	10.26	4.822
YOLOv5s	85.85	500.00	2.00	0.343
YOLOv5m	86.33	212.77	4.70	0.598
YOLOv5l	87.85	121.95	8.20	0.889
YOLOv5x	85.11	65.79	15.20	1.811

Furthermore, object detection needs to ensure better real-time performance. The FPS and inference time are also shown in Table 4. With the exception of the FPS of Faster R-CNN being only 22.69, the other models all exceed 30, which can complete the real-time detection. The FPS of YOLOv5_s is up to 500, and the inference time is also the fastest. It takes just 2 ms to detect the cracks with extent in an image. Of course, the inference times of YOLOX are also rapid, all with 15 ms, and the inference time of the YOLOXs is only 1.51 ms longer than that of the YOLOv5s.

Additionally, Table 4 shows the training time of each model; the longer the training time, the more complex the model network structure. The training time of the YOLOXs for the training set in this paper is 0.583 h, which is only 0.24 h more than the training time of the YOLOv5s, whereas the training of the DSSD is the most time-consuming, requiring 4.822 h.

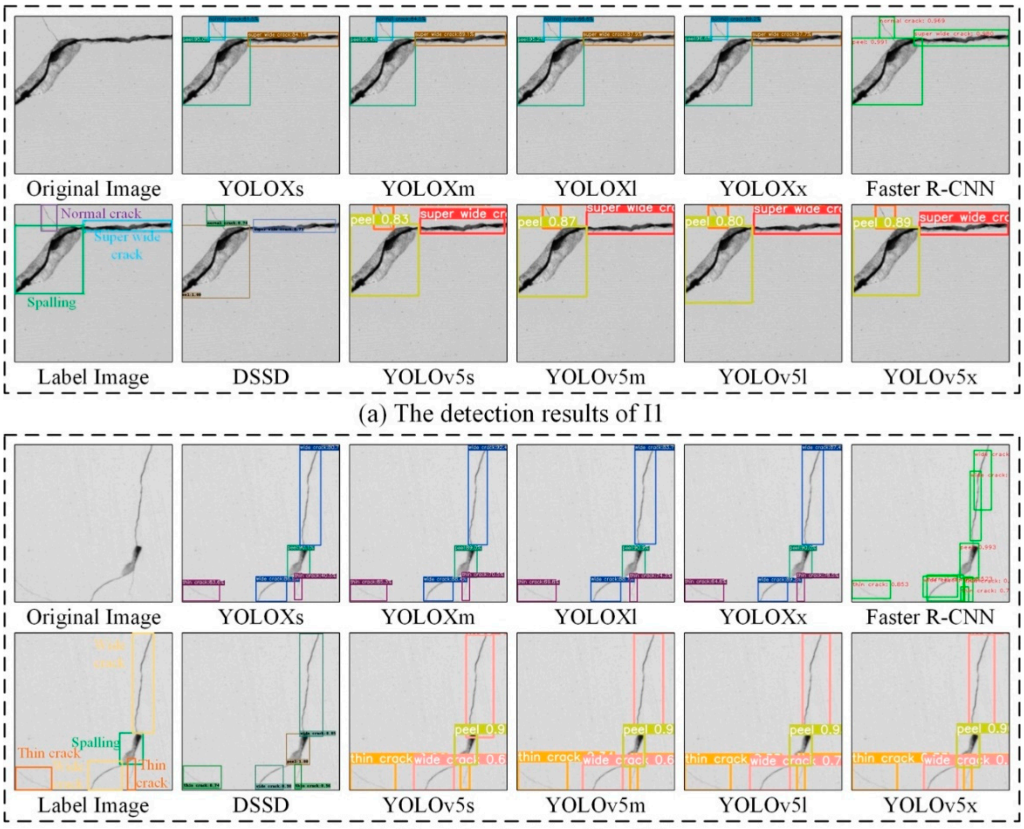
In this study, the crack degrees are divided into five categories according to the different crack widths. Table 5 presents the detection test results of each crack degree, which is expressed by AP value. The YOLOX has superior detection results for wide crack, normal crack, thin crack, and spalling, which are 92.59% (YOLOXl), 91.65% (YOLOXx), 84.75% (YOLOXx), and 99.86% (YOLOXs), respectively. The best AP value of the super wide crack is 95.36% from YOLOv5x. Generally, super wide crack and spalling are easy to detect due to their severe damage and prominent features. In contrast, thin crack, with a width of less than 1 mm, is more difficult to detect because of the small degree. In this study, the crack location and degree detector constructed by the YOLOX model can realize the identification of thin crack with a high accuracy, which is a great breakthrough. The AP value of YOLOX for thin crack is above 74%, which highlights the advantages of the crack location and degree detector for the detection of small features.

Table 5. Test results of different object detection models.

Model	mAP (%)	AP (%)				
		Super Wide Crack	Wide Crack	Normal Crack	Thin Crack	Spalling [1]
YOLOXs	88.78	90.58	91.98	86.98	74.48	**99.86**
YOLOXm	89.98	91.67	91.88	86.04	80.90	99.43
YOLOXl	90.17	90.08	**92.59**	88.44	80.28	99.47
YOLOXx	**91.05**	89.83	90.04	**91.65**	**84.75**	99.01
Faster R-CNN	69.77	78.86	81.95	65.21	33.80	89.01
DSSD	88.92	91.43	89.57	82.21	84.72	96.69
YOLOv5s	85.85	92.16	87.51	77.56	73.35	98.67
YOLOv5m	86.33	92.91	87.12	82.05	71.11	98.46
YOLOv5l	87.85	95.23	91.10	81.35	73.04	98.53
YOLOv5x	88.11	**95.36**	88.57	81.32	76.58	98.72

[1] The authors label spalling as peel, that is, the type of peel damage is spalling.

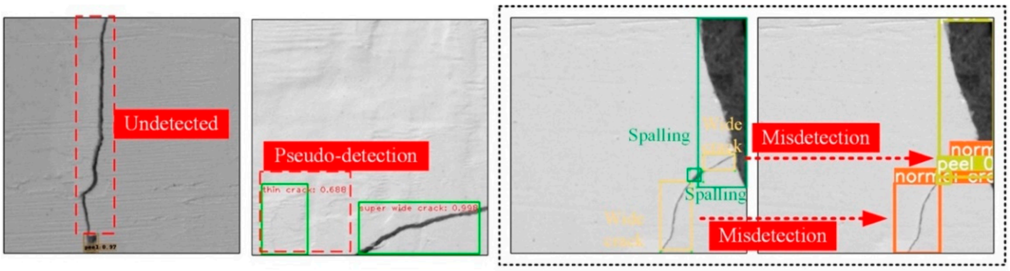
Figure 11 visualizes the partially recognized results of the 10 networks from the same image. There are three crack degrees in the example image 1 (I1), which are a super wide crack, a normal, and a spalling, respectively. Example image 2 (I2) contains two thin cracks, two wide, and a spalling. Although the above models could accurately identify these cracks in the image, some false detections exist during the test process, which directly impact the mAP values. However, the proposed method can effectively avoid too many wrong identifications. Figure 12a–c describes some examples of detected errors, which are roughly divided into three categories. The first type is undetected, as shown in Figure 12a from DSSD, that is, the algorithm does not detect the crack in the image; the second type is pseudo-detection, as shown in Figure 12b from Faster R-CNN, that is, the algorithm recognizes a non-existent crack; the third type is misdetection, as shown in Figure 12c from YOLOv5, that is, the algorithm incorrectly recognizes one crack degree as another crack degree.



(a) The detection results of I1

(b) The detection results of I2

Figure 11. Examples of detection results by different deep learning models.



(a) False detection 1 (b) False detection 2 (c) False detection 3

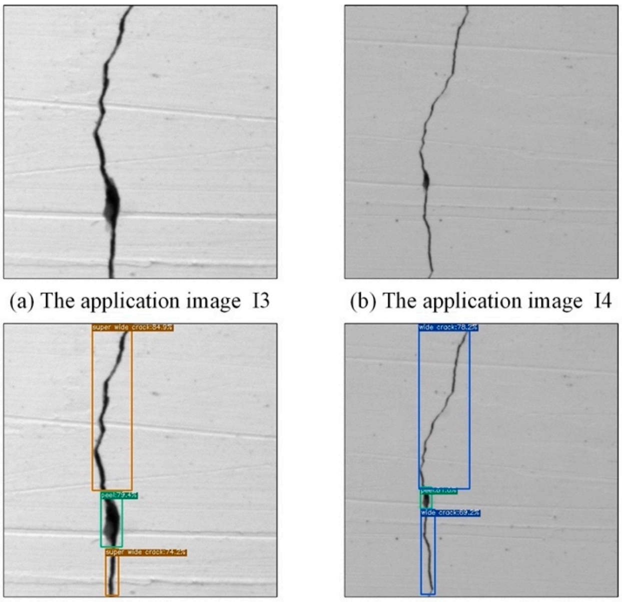
Figure 12. Examples of false detections.

In summary, the YOLOX performs well in the detection and assessment of different types of crack. First, the great mAP values suggest that the model's ability to detect cracks with different degrees is a relatively reliable performance. Secondly, high FPS and short inference time represent that the detection speed is very rapid and can be content with real-time detection. Thirdly, the short training time indicates that the network structure is not complex. Finally, the high AP for thin crack suggests that this model is suitable for small crack detection. Therefore, the trained YOLOX model is selected as the crack location and degree detector.

4.5. Crack Degree Transformation Test and Results

The output of the crack location and degree detector proposed in this study is the pixel level, and the actual crack degree needs to be obtained according to the resolution.

The application images are collected under different working distances by the constructed image acquisition system. The application image I3 is still obtained at a working distance of 0.35 m, the pixel resolution K1 is 0.048 mm/pixel; the application image I4 is obtained when the working distance is 0.70 m, and the FOV is 291 mm × 397 mm, the corresponding pixel resolution K2 is 0.097 mm/pixel. The application images are shown in Figure 13a,b, and the actual width of the crack measured by the crack width measuring instrument is 0.45 mm.



(a) The application image I3

(b) The application image I4

(c) The detection result of I3

(d) The detection result of I4

Figure 13. Detection results of application images.

Figure 13c,d show the detection results of application images using the proposed crack location and degree detector. There are two super wide cracks (the width greater than 0.3 mm) in I3, which is consistent with the actual situation and is correctly identified. Nevertheless, the detection results are two wide cracks for I4, because of the difference in resolution. The pixel resolution of the data set in this paper is 0.048 mm/pixel, and the defined wide crack is a crack with a width between 0.2 mm and 0.3 mm. Therefore, the corresponding crack pixel width in the image is 4–6 pixels. According to the pixel resolution transformation relationship, K2 is 0.097 mm/pixel. The wide crack is between 0.388 mm and 0.582 mm. The actual crack width is 0.45 mm; it is also a correct detection. Therefore, the actual crack degree can be obtained by the proposed detector.

5. Conclusions and Future Work

Structural damage detection and evaluation have always been a research concern in the field of SHM and are also a significant part. The traditional visual structural crack detection method is time-consuming and laborious. Additionally, most of the SHM methods based on image processing are only for a single image, a process which is non-automatic and inefficient. To address these issues, a detector based on YOLOX is utilized to detect different degrees of crack in this study. This method can detect and assess the structural state quickly, accurately, and automatically. An image dataset with different extents of crack is constructed in this paper. The dataset is mainly marked according to the width of cracks, which are divided into five types: super wide cracks, wide cracks, normal cracks, thin cracks, and spalling, respectively. It is a bold attempt to combine crack detection and assessment. This work can directly identify and evaluate different degrees of defect, rather than recognize crack first and then assess the characteristics of the detected crack. Moreover, some advanced deep learning models were systematically compared to illustrate the superiority of using YOLOX as the detector. From the training and testing, the mAP values of YOLOX exceed 88.5% on a whole, and the maximum mAP value is 91.05% in YOLOXx, whereas the mAP values of YOLOv5 and DSSD network are relatively stable at about 86% and the mAP value of the Faster R-CNN model is only 69.77%. Compared with other algorithms, the FPS and reference time of the proposed model are relatively shorter, which can meet the needs of real-time detection. Furthermore, thin crack detection is a challenge in object detection. The proposed method is especially strong at detecting thin cracks. For example, the AP values of thin cracks from YOLOX are at least 74%. Comprehensively, the proposed model demonstrates the most balanced detection performance and detection speed. Finally, the feasibility of the proposed method is proved by the crack degree transformation test.

This study is groundbreaking in that it combines damage detection and assessment. However, one common limitation of almost all deep learning approaches is that they require numerous training data to obtain excellent results. Therefore, a long-term data collection plan is inevitable and will be pursued.

Author Contributions: Conceptualization, J.L. and F.K.; methodology, L.W.; validation, L.W.; formal analysis, J.L. and F.K.; writing—review and editing, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant numbers 51779035, 51769033, and 51979027, 52079022; Fundamental Research Funds for the Central Universities grant number DUT21TD106.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Dawei Ren and Dezhi Zai for their help in the manufacturing of concrete slabs.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jang, K.; An, Y.-K. Multiple crack evaluation on concrete using a line laser thermography scanning system. *Smart Struct. Syst.* **2018**, *22*, 201–207. [CrossRef]
2. Kim, H.; Lee, J.; Ahn, E.; Cho, S.; Shin, M.; Sim, S.-H. Concrete crack identification using a UAV incorporating hybrid image processing. *Sensors* **2017**, *17*, 2052. [CrossRef] [PubMed]
3. Kim, B.; Yuvaraj, N.; Sri, K.R.; Arun, R. Surface crack detection using deep learning with shallow CNN architecture for enhanced computation. *Neural Comput. Appl.* **2021**, *33*, 9289–9305. [CrossRef]
4. Zhang, H.; Li, J.; Kang, F.; Zhang, J. Monitoring and evaluation of the repair quality of concrete cracks using piezoelectric smart aggregates. *Constr. Build. Mater.* **2022**, *317*, 125775. [CrossRef]

5. Liu, Y.; Cho, S.; Spencer, B.F.; Fan, J. Automated assessment of cracks on concrete surfaces using adaptive digital image processing. *Smart Struct. Syst.* **2014**, *14*, 719–741. [CrossRef]
6. Li, G.; Li, X.; Zhou, J.; Liu, D.; Ren, W. Pixel-level bridge crack detection using a deep fusion about recurrent residual convolution and context encoder network. *Measurement* **2021**, *176*, 109171. [CrossRef]
7. Savino, P.; Tondolo, F. Automated classification of civil structure defects based on convolutional neural network. *Front. Struct. Civ. Eng.* **2021**, *15*, 305–317. [CrossRef]
8. Jahanshahi, M.R.; Masri, S.F.; Sukhatme, G.S. Multi-image stitching and scene reconstruction for evaluating defect evolution in structures. *Struct. Health Monit.* **2011**, *10*, 643–657. [CrossRef]
9. Yang, C.; Chen, J.; Li, Z.; Huang, Y. Structural crack detection and recognition based on deep learning. *Appl. Sci.* **2021**, *11*, 2868. [CrossRef]
10. Jahanshahi, M.R.; Kelly, J.S.; Masri, S.F.; Sukhatme, G.S. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Struct. Infrastruct. Eng.* **2009**, *5*, 455–486. [CrossRef]
11. Wang, W.; Hu, W.; Wang, W.; Xu, X.; Wang, M.; Shi, Y.; Qiu, S.; Tutumluer, E. Automated crack severity level detection and classification for ballastless track slab using deep convolutional neural network. *Autom. Constr.* **2021**, *124*, 103484. [CrossRef]
12. Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210. [CrossRef]
13. Li, B.; Wang, K.C.P.; Zhang, A.; Yang, E.; Wang, G. Automatic classification of pavement crack using deep convolutional neural network. *Int. J. Pavement Eng.* **2018**, *21*, 457–463. [CrossRef]
14. Mei, Q.; Gül, M.; Azim, M.R. Densely connected deep neural network considering connectivity of pixels for automatic crack detection. *Autom. Constr.* **2020**, *110*, 103018. [CrossRef]
15. Zheng, M.; Lei, Z.; Zhang, K. Intelligent detection of building cracks based on deep learning. *Image Vis. Comput.* **2020**, *103*, 03987. [CrossRef]
16. Bae, H.; Jang, K.; An, Y.-K. Deep super resolution crack network (SrcNet) for improving computer vision–based automated crack detectability in in situ bridges. *Struct. Health Monit.* **2021**, *20*, 1428–1442. [CrossRef]
17. Saleem, M.R.; Park, J.-W.; Lee, J.-H.; Jung, H.-J.; Sarwar, M.Z. Instant bridge visual inspection using an unmanned aerial vehicle by image capturing and geo-tagging system and deep convolutional neural network. *Struct. Health Monit.* **2020**, *20*, 1760–1777. [CrossRef]
18. Xu, Y.; Li, D.; Xie, Q.; Wu, Q.; Wang, J. Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. *Measurement* **2021**, *178*, 109316. [CrossRef]
19. Huang, H.-W.; Li, Q.-T.; Zhang, D.-M. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunn. Undergr. Space Technol.* **2018**, *77*, 166–176. [CrossRef]
20. Tang, J.; Mao, Y.; Wang, J.; Wang, L. Multi-task enhanced dam crack image detection based on Faster R-CNN. In Proceedings of the IEEE 4th International Conference on Image, Vision and Computing, Xiamen, China, 5–7 July 2019; pp. 336–340. [CrossRef]
21. Zhang, L.; Yang, F.; Zhang, Y.D.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the IEEE International Conference on Image Processin, Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712. [CrossRef]
22. Wang, B.; Zhao, W.; Gao, P.; Zhang, Y.; Wang, Z. Crack damage detection method via multiple visual features and efficient multi-task learning model. *Sensors* **2018**, *18*, 1796. [CrossRef]
23. Cha, Y.-J.; Choi, W.; Büyüköztürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Comput. Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [CrossRef]
24. Chen, F.-C.; Jahanshahi, M.R. NB-CNN: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4392–4400. [CrossRef]
25. Gopalakrishnan, K.; Khaitan, S.K.; Choudhary, A.; Agrawal, A. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* **2017**, *157*, 322–330. [CrossRef]
26. Zhang, A.; Wang, K.C.P.; Li, B.; Yang, E.; Dai, X.; Peng, Y.; Fei, Y.; Liu, Y.; Li, J.Q.; Chen, C. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Comput. Aided Civ. Infrastruct. Eng.* **2017**, *32*, 805–819. [CrossRef]
27. Fan, Z.; Wu, Y.; Lu, J.; Li, W. Automatic pavement crack detection based on structured prediction with the convolutional neural network. *arXiv* **2018**, arXiv:1802.02208.
28. Kim, I.H.; Jeon, H.; Baek, S.C.; Hong, W.H.; Jung, H.J. Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle. *Sensors* **2018**, *18*, 1881. [CrossRef]
29. Dorafshan, S.; Thomas, R.J.; Maguire, M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr. Build. Mater.* **2018**, *186*, 1031–1045. [CrossRef]
30. Jang, K.; Kim, N.; An, Y.K. Deep learning–based autonomous concrete crack evaluation through hybrid image scanning. *Struct. Health Monit.* **2019**, *18*, 1722–1737. [CrossRef]
31. Guo, L.; Li, R.; Jiang, B.; Shen, X. Automatic crack distress classification from concrete surface images using a novel deep-width network architecture. *Neurocomputing* **2020**, *397*, 383–392. [CrossRef]
32. Li, S.; Gu, X.; Xu, X.; Xu, D.; Zhang, T.; Liu, Z.; Dong, Q. Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm. *Constr. Build. Mater.* **2021**, *273*, 121949. [CrossRef]

33. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
34. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]
35. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
36. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NA, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
37. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
38. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
39. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
40. Jocher, G. Yolov5 Github Repository. 2021. Available online: https://github.com/ultralytics/yolov5 (accessed on 1 August 2021).
41. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
42. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
43. Fu, C.Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd: Deconvolutional single shot detector. *arXiv* **2017**, arXiv:1701.06659.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
45. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A metric and a loss for bounding box regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666. [CrossRef]
46. Everingham, M.; Gool, L.V.; Williams, C.K.I.; Winn, J.; Zisserman, A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
47. Ma, H.; Liu, Y.; Ren, Y.; Yu, J. Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3. *Remote Sens.* **2020**, *12*, 44. [CrossRef]