applied sciences

MDPI

Article

A Novel Hybrid Approach for Concrete Crack Segmentation Based on Deformable Oriented-YOLOv4 and Image Processing Techniques

Zengsheng He [1], Cheng Su [1,2,3] and Yichuan Deng [1,2,3,*]

[1] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China; 202010101411@mail.scut.edu.cn (Z.H.); cvchsu@scut.edu.cn (C.S.)
[2] State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou 510641, China
[3] Guangdong Artificial Intelligence and Digital Economy Laboratory, Guangzhou 510641, China
* Correspondence: ctycdeng@scut.edu.cn

Abstract: Regular crack inspection plays a significant role in the maintenance of concrete structures. However, most deep-learning-based methods suffer from the heavy workload of pixel-level labeling and the poor performance of crack segmentation with the presence of background interferences. To address these problems, the Deformable Oriented YOLOv4 (DO-YOLOv4) is first developed for crack detection based on the traditional YOLOv4, in which crack features can be effectively extracted by deformable convolutional layers, and the crack regions can be tightly enclosed by a series of oriented bounding boxes. Then, the proposed DO-YOLOv4 is further utilized in combination with the image processing techniques (IPTs), leading to a novel hybrid approach, termed DO-YOLOv4-IPTs, for crack segmentation. The experimental results show that, owing to the high precision of DO-YOLOv4 for crack detection under background noise, the present hybrid approach DO-YOLOv4-IPTs outperforms the widely used Convolutional Neural Network (CNN)-based crack segmentation methods with less labeling work and superior segmentation accuracy.

Keywords: computer vision; convolutional neural network; YOLOv4; image processing techniques; crack detection; crack segmentation

Citation: He, Z.; Su, C.; Deng, Y. A Novel Hybrid Approach for Concrete Crack Segmentation Based on Deformable Oriented-YOLOv4 and Image Processing Techniques. *Appl. Sci.* **2024**, *14*, 1892. https://doi.org/10.3390/app14051892

Academic Editor: João M. F. Rodrigues

Received: 10 January 2024
Revised: 20 February 2024
Accepted: 22 February 2024
Published: 25 February 2024

1. Introduction

Cracks are common defects appearing on the surface of concrete structures, which can visually reflect structural degradation. Therefore, regular crack inspections are essential to the assessment of structural conditions. However, the traditional method of crack inspection is time-consuming and labor-intensive. As an alternative, computer vision methods, including the Image Processing Techniques (IPTs)-based method and Convolutional Neural Network (CNN)-based method, have proved to be efficient and reliable and have been widely applied to engineering practice in the past few decades [1,2].

Recently, the CNN-based segmentation method has attracted much attention for its capability of pixel-level crack segmentation, whose results can be directly used to obtain the geometric parameters of cracks, such as length, width, and direction. Consequently, the potential damages and the relevant causes can be inferred from the morphological characteristics of cracks. Representative research in this field can be found in [3–13]. However, as a deep learning method, CNN-based segmentation suffers from the preparation of pixel-level labels, which is known to be rather costly [14].

In view of this, a hybrid approach for crack segmentation has been proposed [15–17], combining the CNN-based object detection method for locating crack regions and the IPTs-based segmentation method for detecting crack pixels within those regions. The hybrid approach can greatly reduce the workload of labeling, owing to the fact that the labeling

time of box annotations is much less than that of pixel annotations [18]. Moreover, the IPTs-based segmentation method will, in general, exhibit higher computational speed and lower resource consumption compared with the CNN-based segmentation method. Therefore, the hybrid approach can merge the advantages of both CNN-based object detection and IPTs-based segmentation methods, leading to less labeling cost and higher efficiency for crack segmentation. However, the performance of the hybrid approach will significantly deteriorate with the presence of false or missed crack region detection. Therefore, object detection is expected to locate the crack regions precisely so as to improve the performance of the hybrid approach effectively.

Actually, the existing object detection methods, e.g., Faster Region-CNN (Faster R-CNN) [19], Single Shot Multibox Detector (SSD) [20], and You Only Look Once (YOLO) [21,22] are not that effective in detecting cracks. To tackle this problem, Ma et al. [23] introduced a feature extraction network, namely Resnet, with up to 101 convolutional layers, to enhance the performance of the original Faster R-CNN for crack detection. Similarly, Pang et al. [24] developed an improved Faster R-CNN by incorporating the Inception Resnet v2, which consists of more convolutional layers. In contrast to introducing complex feature extraction networks, Xu et al. [25] utilized several lightweight Squeeze and Excitation (SE) blocks to achieve better performance in crack detection with Faster R-CNN. In addition, Yang et al. [26] introduced receptive field blocks with a few convolutional layers to enhance the feature extraction capability of the network, leading to an improved SSD with higher accuracy for crack detection. Zhang et al. [27] enhanced the traditional YOLOv4 through the utilization of convolutional attention blocks, which can identify crack regions from multi-object images. However, the aforementioned studies involve complex networks for crack detection, owing to the utilization of deep feature extraction networks or the incorporation of feature extraction blocks. Moreover, the horizontal bounding boxes adopted in the above object detection methods fail to tightly locate cracks with various distribution directions, making network training challenging with the presence of background interferences.

In this study, several deformable convolutional layers [28] are incorporated into the traditional YOLOv4 to enhance the ability of feature extraction for weakly patterned cracks while keeping the improved network lightweight without loss of accuracy. Furthermore, to reduce the influence of background interferences during network training, a series of oriented bounding boxes is introduced into YOLOv4 to tightly enclose the inclined crack segments, resulting in higher accuracy for crack detection. Based on the above treatments, the improved YOLOv4 adopted in the present study is termed Deformable-Oriented-YOLOv4 (DO-YOLOv4), which will be further used in conjunction with IPTs for crack segmentation, leading to an improved hybrid approach named as DO-YOLOv4-IPTs.

The remainder of this paper starts with a review of CNN-based object detection and IPTs-based segmentation methods. Then, the methodology of the improved hybrid approach is presented. Following that, several comparative experiments are presented, and the experimental results are discussed. Finally, the contributions of this paper are summarized.

2. Related Work

2.1. CNN-Based Object Detection

Typical CNN-based object detection methods are widely used for recognizing and locating objects with bounding boxes, which can be classified into two main categories: the two-stage approach and the one-stage approach. For the two-stage object detection approach, potential regions that may contain objects are first proposed, and then the features in these regions will be used for box classification and regression. The R-CNN [29] is the early version of the two-stage approach, which needs to repeatedly extract features for different regions, resulting in a high computational cost. To improve the detection efficiency, the Fast R-CNN [30] and Faster R-CNN were proposed successively. In particular, the Faster R-CNN is a representative two-stage method for its high detection accuracy. Cha et al. [31] used Faster R-CNN to detect five types of bridge defects, including concrete cracks, steel corrosions, delamination, etc., and the experimental results showed the high

accuracy of Faster R-CNN in defect detection. Deng et al. [32] showed that Faster R-CNN could effectively locate cracks under complex backgrounds, even with the presence of handwriting scripts. Li et al. [33] enhanced Faster R-CNN through the feature fusion technique, which was capable of detecting cracks on the tunnel surface with high precision. Despite the ideal accuracy of Faster R-CNN, it still has the problem of low detection speed due to the region proposal step.

As high-speed detection is essential for the development of real-time automatic inspection systems, the one-stage approach was proposed by removing the time-consuming region proposal step. The SSD and YOLO are two well-known one-stage methods that have been applied to crack detection in bridges [34] and pavements [26,35,36] in recent years. Generally, the one-stage approach outperforms the two-stage approach in terms of detection speed but compromises in detection accuracy. To tackle this problem, the YOLOv3 [37] was proposed by using several effective techniques, such as the short-cut connection [38], feature pyramid [39], and multi-scale detection [20]. Zhang et al. [40] showed that YOLOv3 achieved comparable accuracy to Faster R-CNN without sacrificing computational efficiency. To further enhance both detection accuracy and speed, the YOLOv4 was proposed with the advanced network architecture consisting of Cross Stage Partial Darknet53 (CSPDarknet53) for feature extraction, Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN) for feature enhancement and YOLO head for bounding box output [22]. On the basis of YOLOv4, Yu et al. [41] developed a novel YOLOv4-FPM for real-time and accurate detection of bridge cracks, which combines the focal loss, pruning algorithm, and multi-scale dataset. Additionally, Zhou et al. [42] enhanced the traditional YOLOv4 by incorporating EfficientNet and depthwise separable convolution, leading to YOLOv4-ED with superior performance in terms of both detection accuracy and efficiency.

Although the CNN-based methods have achieved great success in the area of object detection, there are still two critical issues that need to be addressed when applied to crack detection, i.e., how to establish a lightweight object detection model and how to reduce the influence of background interferences on network training for detecting line-shaped cracks with various propagation directions. These two problems will be solved by introducing deformable convolutional layers and multiple oriented bounding boxes, which will be elaborated in Section 3.

2.2. IPTs-Based Segmentation

The IPTs-based segmentation methods, including edge detection, seed-growing, and thresholding methods, have been extensively studied for crack segmentation [43]. For the edge detection method, the Sobel operator and Canny operator have been widely used for crack edge identification based on gradient characteristics. Ayenu-Prah and Attoh-Okine [44] combined the Bi-dimensional Empirical Mode Decomposition (BEMD) with the Sobel operator to improve the performance of crack edge detection. Abdel-Qader et al. [45] compared the performance of the Fast Fourier Transform (FFT), Fast Haar Transform (FHT), Sobel operator, and Canny operator, and the results showed that FHT is more accurate and reliable for detecting crack edges. The major problem of the edge detection method lies in its difficulty in detecting complete crack edges, leading to the low performance of crack segmentation.

The seed-growing method [46,47] has drawn much attention for its capability of segmenting cracks at high accuracy, in which, to obtain the complete crack, potential crack seeds are first assigned, and the seeds are linked together via the path-searching algorithm. Gavilan et al. [48] employed the multiple directional non-minimum suppression technique to obtain the crack seeds and then connected these seeds by utilizing the minimum distance cost map. To select crack seeds effectively, Zou et al. [49] further constructed a crack probability map by employing the tensor voting technique. Nevertheless, the seed-growing method needs much calculation for seed selection and path searching, resulting in high computational costs when applied to engineering practice.

The thresholding method, in which local or global brightness thresholds are used to identify crack pixels, is known to be fast and simple. Tang et al. [50] employed the histogram thresholding method to obtain the approximate locations of cracks and then refine these locations utilizing the snake model. Following that, Li and Liu [51] proposed a Neighboring Difference Histogram Model (NDHM) considering the fact that crack pixels are darker than their surroundings. Oliveira and Correia [52] further proposed a dynamic thresholding method in consideration of the standard deviation of all pixel intensities and the entropy of each image block. In addition, the computer vision function library OpenCV also provided an adaptive thresholding method for segmentation, in which the adaptive threshold is calculated for each input pixel.

Even though there have been a variety of IPTs-based segmentation methods, the background interferences contained in crack regions will greatly hinder the performance of crack segmentation. To solve this problem, a hybrid approach is proposed for crack segmentation by combining the novel DO-YOLOv4 with the effective adaptive thresholding method, which will be presented in detail in Section 3.

3. Methodology

3.1. Overview of DO-YOLOv4-IPTs

The overview of the proposed hybrid approach DO-YOLOv4-IPTs for accurate crack segmentation is illustrated in Figure 1, in which the crack region is first detected by module DO-YOLOv4, and the crack region is then cropped and processed by module IPTs for crack segmentation. To obtain a lightweight and accurate object detection model, in DO-YOLOv4 shown in Figure 1, the original CSPDarknet53 is improved by introducing deformable convolutional layers, and the single horizontal bounding box in the traditional YOLO head is replaced by multiple oriented bounding boxes tightly enclosing the cracks. Then, the adaptive thresholding method and connected component analysis are adopted to segment cracks in IPTs effectively.

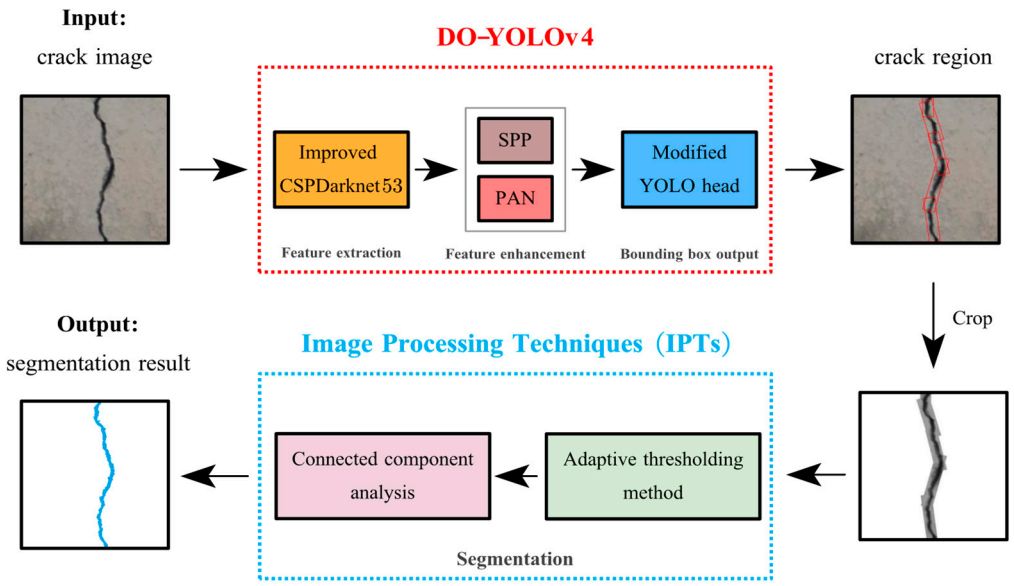


Figure 1. The overview of DO-YOLOv4-IPTs.

3.2. DO-YOLOv4 for Crack Detection

3.2.1. Deformable Convolutional Layers for Feature Extraction

In CNN, the convolutional filter with a size of $N \times N$ is used to extract features through sampling and weighted summation in the process of scanning the input image. However, the sampling locations are limited to a $N \times N$ grid and may be far from the target object with the scanning of the convolutional filter, leading to low performance of feature extraction. This problem can be solved by the deformable convolution [28], in which 2D offsets are

added to the fixed sampling locations, generating deformable sampling locations with the capability of adjusting to various objects with different locations, shapes, and scales. Given an example of a 3×3 convolutional filter, the process of generating deformable sampling locations is illustrated in Figure 2, in which the offset field with 3×3 2D offsets obtained by additional convolutional filters helps distribute the sampling locations in the vicinity of the crack. The deformable convolution has been successfully utilized to identify retinal vessels of various shapes [53], locate vehicles in high-resolution remote sensing images [54], and detect dead fish in aquaculture with a lightweight network [55]. In the present study, it will be further employed to enhance the performance of traditional YOLOv4 for crack detection.

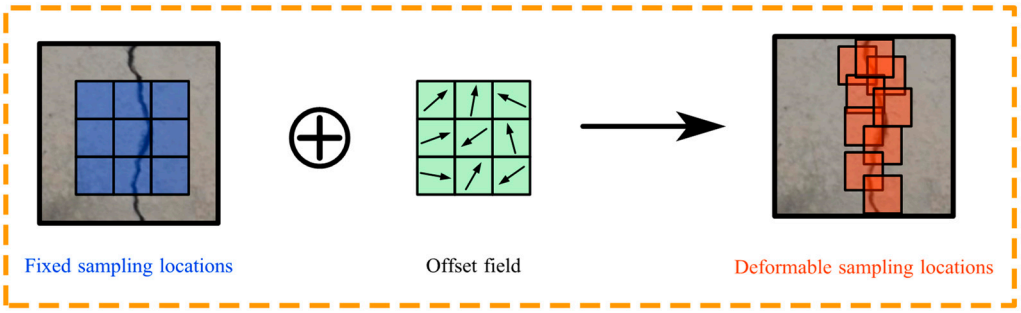


Figure 2. The process of generating deformable sampling locations.

The original CSPDarknet53 in YOLOv4 consists of 6 convolutional layers and 5 CSPblocks with different numbers of Resblocks, as illustrated in Figure 3. The network depth is increased mainly by stacking more Resblocks in the last three CSPblocks so as to ensure the capability of feature extraction. In this study, to obtain the optimal features of cracks with various directions, the five convolutional layers in the original CSPDarknet53 are replaced by five deformable convolutional counterparts, as shown in Figure 4. Owing to the use of deformable sampling locations in the deformable convolutional layers, the numbers of Resblocks employed in CSPblocks c, d, and e are now reduced from 8, 8, and 4 to 2, 2, and 1, respectively, leading to a lightweighted feature extraction network without loss of accuracy.

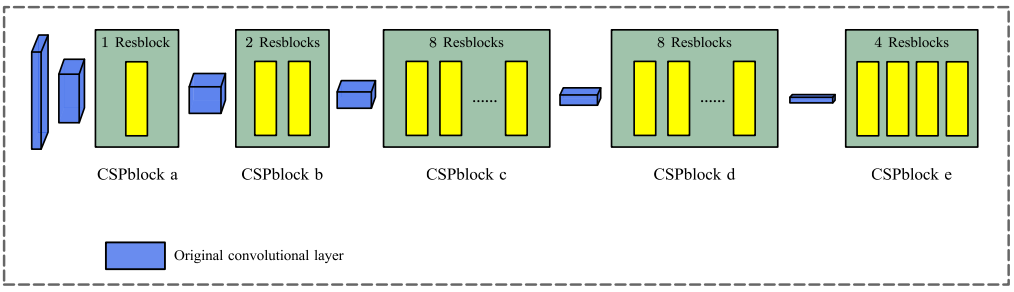


Figure 3. The original CSPDarknet53.

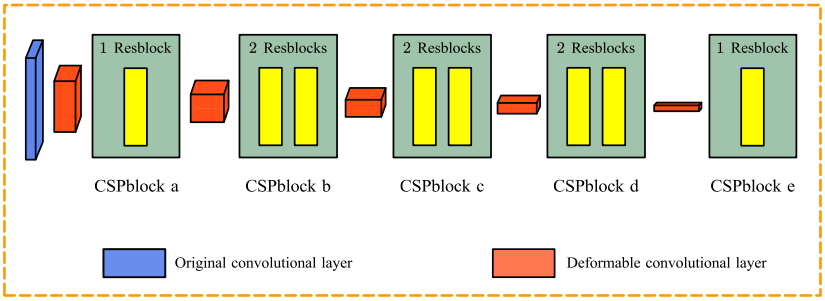


Figure 4. The improved CSPDarknet53 with deformable convolutional layers.

3.2.2. Multiple Oriented Bounding Boxes for Training

In YOLOv4, the traditional single horizontal bounding box is used for object detection. However, for the detection of an inclined crack, the horizontal bounding box, as shown in Figure 5a, will certainly induce much more background interferences, which will have a large influence on the network training and thus deteriorate the performance of crack detection. In view of this, in the proposed DO-YOLOv4, a series of oriented bounding boxes is employed to tightly enclose an inclined crack, as shown in Figure 5b, which will significantly reduce the influence of background interferences on network training and, therefore, greatly enhance the accuracy of crack detection. For each oriented bounding box, it can be located using the parameters (c_x, c_y, w, h, θ) as shown in Figure 6, in which c_x and c_y denote the coordinates of the center point of the bounding box, w and h denote its width and height, and θ denotes the oriented angle from the x-axis.

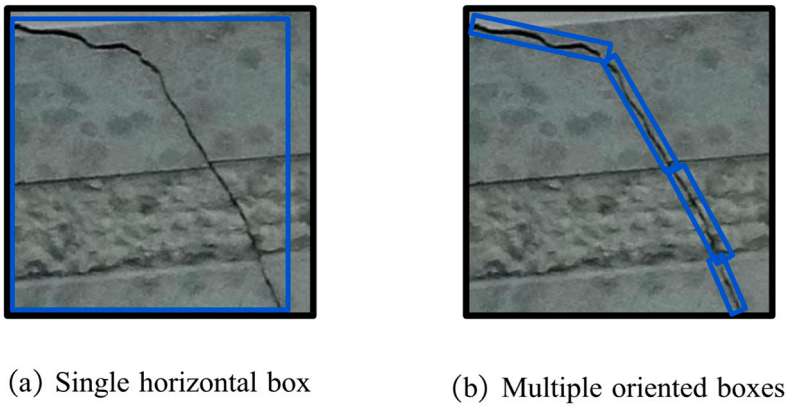


Figure 5. Bounding boxes for crack detection.

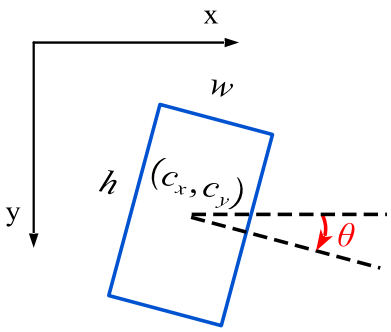


Figure 6. The definition of an oriented bounding box.

3.3. IPTs for Crack Segmentation

In the present hybrid approach, DO-YOLOv4-IPTs, once the crack region is detected by DO-YOLOv4, the IPTs based on the adaptive thresholding method and connected component analysis will be further used for crack segmentation, as shown in Figure 7. To effectively identify crack pixels for each input pixel, the adaptive thresholding method will assign an adaptive threshold as $A_{\text{th}} = I_{\text{m}} - C$, in which I_{m} is the mean pixel intensity of a square region around the input pixel and C is a given constant. When the intensity of the input pixel is lower than the threshold, it is classified as the crack pixel (black), and the contrary is classified as the background pixel (white). For example, for the input pixel A with an intensity of 70, as shown in Figure 7, the adaptive threshold is taken as $A_{\text{th}} = I_{\text{m}} - C = 97 - 5 = 92$, in which $I_{\text{m}} = 97$ is the mean pixel intensity of a region of 3×3 pixels around pixel A and $C = 5$ is the given constant. As the intensity of pixel A is smaller than the adaptive threshold, it is identified as a crack pixel. However, for the input pixel B in Figure 7, it can be identified as a background pixel using the same method.

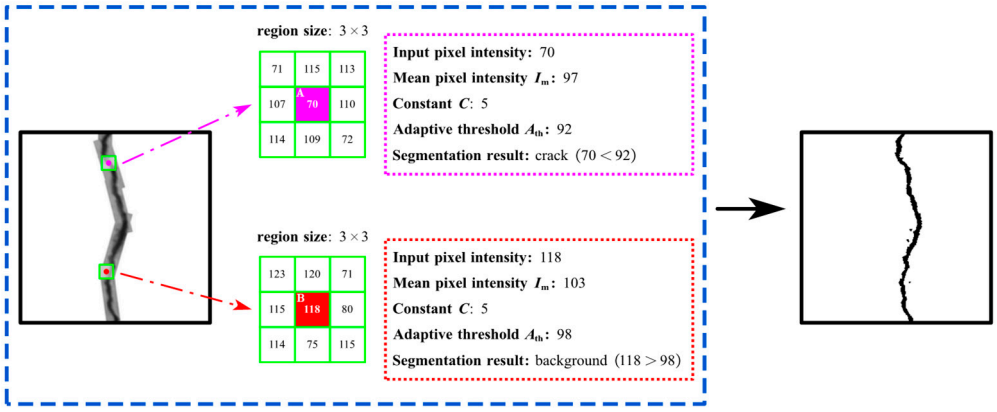


Figure 7. The adaptive thresholding method for crack segmentation.

To further improve the accuracy of crack segmentation, the connected component analysis is conducted to eliminate the falsely classified crack pixels. For illustration, the crack pixels in Figure 8 are first clustered into several connected components, and the respective areas can be calculated based on the number of crack pixels within each component. Then, an area threshold AR_{th} can be determined by analyzing the size of each component. The crack pixels within the components that are smaller than the area threshold AR_{th} can be regarded as falsely classified crack pixels and thus can be removed, leading to a more accurate segmentation result.

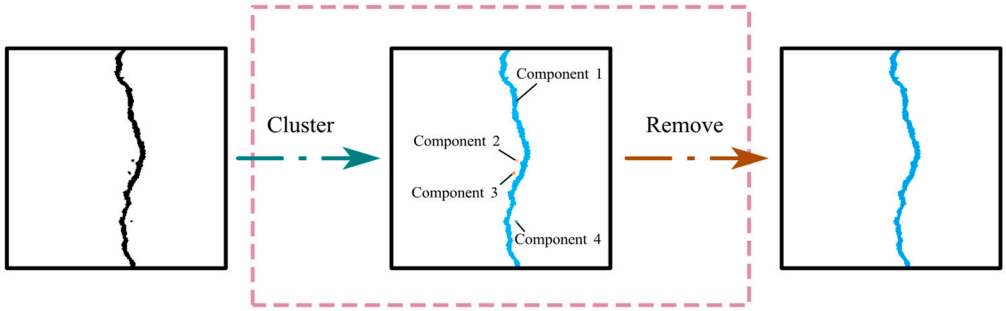


Figure 8. The process of connected component analysis.

4. Experiments

4.1. Dataset Construction and Implementation Details

In the present study, 740 crack images with different background characteristics were selected from a public bridge crack dataset [56], in which 500 and 240 images were chosen for training and testing of the proposed DO-YOLOv4 for crack detection, respectively. Both training and testing crack images were labeled with a series of oriented bounding boxes tightly enclosing the cracks by professional annotation software roLabelImg [57].

The present DO-YOLOv4 was implemented based on the open-source deep learning library PyTorch [58], in which the optimization algorithm Adam [59], widely used in object detection, was selected to train DO-YOLOv4. In this study, the training was performed for 100 epochs with a batch size of 2 for 256 × 256 input images. The initial learning rate of the Adam algorithm was set as 0.001, and the decay rate for the learning rate was taken as 0.9 for each epoch. The Complete Intersection over Union (CIoU) loss function [60] and smooth L1 loss function [19] were adopted to optimize the parameters (c_x, c_y, w, h) and θ of the oriented bounding box, respectively, and the Focal loss function [61] was utilized for the classification of crack or background contained in the oriented bounding boxes.

On the basis of crack detection by DO-YOLOv4, the IPTs were further employed in the present hybrid approach for crack segmentation, which was implemented based on the computer vision library OpenCV [62]. The region size for determining the mean pixel intensity, as illustrated in Figure 7, was taken to be 5 × 5 to 21 × 21 pixels, depending on

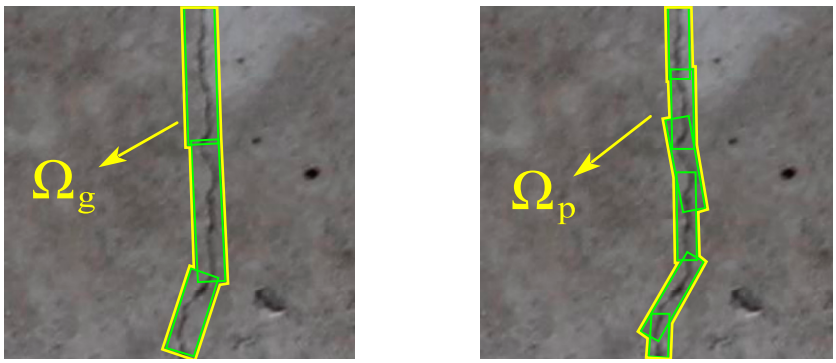
the width of the crack, and the constant C is taken as 5 for determination of the adaptive threshold. In addition, if the size of the connected component was found to be smaller than 30 pixels by the connected component analysis, the pixels contained in this component were classified as false crack pixels and should be removed.

All experiments were implemented on a computer with the Intel® Core™ i5-10400 CPU @ 64-bit 2.90 GHz, 16 GB RAM (Intel, Santa Clara, CA, USA), and NVIDIA GeForce RTX 3080 GPU (NVIDIA, Santa Clara, CA, USA).

4.2. Evaluation of DO-YOLOv4 for Crack Detection

4.2.1. Evaluation Metrics

As the crack region is detected by DO-YOLOv4 using a series of oriented bounding boxes, the widely-used Intersection over Union (IoU) metric is now defined based on the overlap of the whole region Ω_g enclosed by all ground truth boxes and the whole region Ω_p enclosed by all predicted boxes, as shown in Figure 9. If the IoU of the ground truth region Ω_g and the predicted region Ω_p is larger than or equal to 0.5 [63], the detection result is regarded as a true result (true positive), and the contrary is identified as a false one (false positive).



(a) Ground truth region Ω_g (b) Predicted region Ω_p

Figure 9. The ground truth region and predicted region of DO-YOLOv4.

For the current experiment, a total of 240 crack images with 282 cracks were labeled for testing, leading to 282 ground truth regions $\Omega_{g\,i}$ $(i = 1, 2, \cdots, 282)$. Accordingly, a total of 282 predicted regions $\Omega_{p\,i}$ $(i = 1, 2, \cdots, 282)$ were obtained by DO-YOLOv4. For the predicted region $\Omega_{p\,i}$, take its confidence C_i to be the average confidence of different predicted boxes involved in $\Omega_{p\,i}$. To consider different confidence thresholds, assume the i-th confidence threshold $C_{t\,i}$ to be C_i $(i = 1, 2, \cdots, 282)$. Corresponding to $C_{t\,i}$, the Precision and Recall metrics, P_i and R_i can be obtained as follows [49]:

$$P_i = \frac{TP_i}{TP_i + FP_i} (i = 1, 2, \cdots, 282) \quad (1)$$

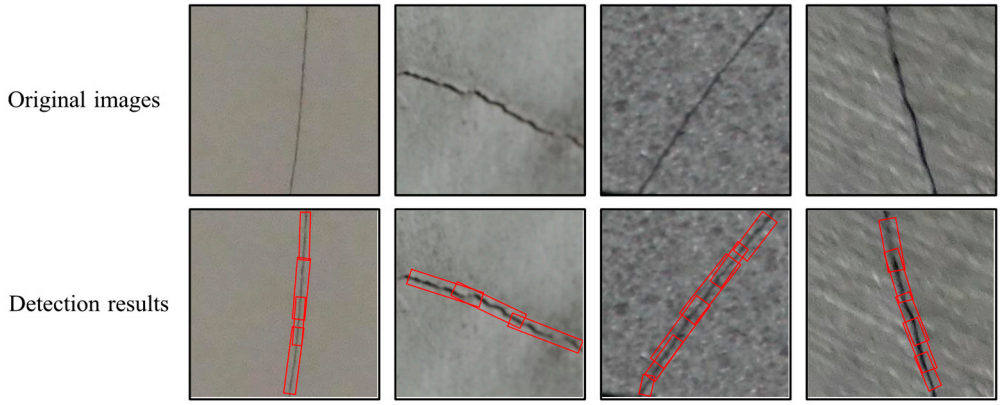
$$R_i = \frac{TP_i}{TP_i + FN_i} (i = 1, 2, \cdots, 282) \quad (2)$$

where TP_i and FP_i respectively represent the number of true and false detection results among those predicted regions with confidences larger than or equal to $C_{t\,i}$, and FN_i denotes the number of non-detected ground truth regions, i.e., $FN_i = 282 - TP_i$.

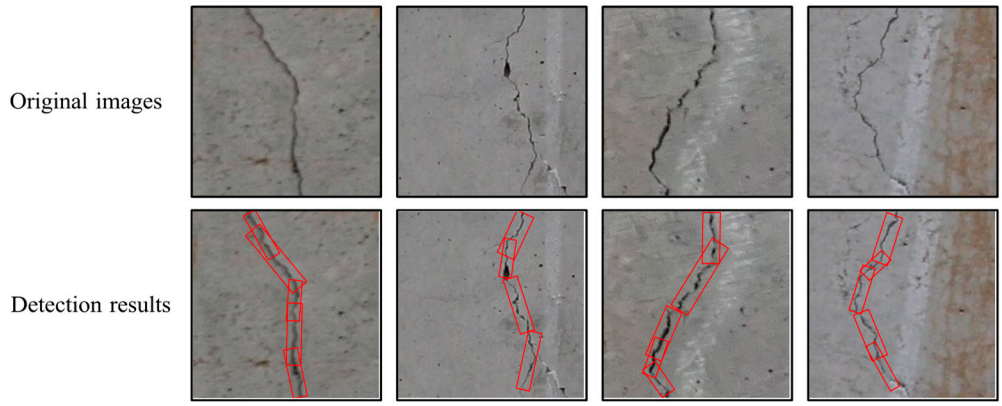
Using (P_i, R_i) $(i = 1, 2, \cdots, 282)$ calculated above, the Precision-Recall (P-R) curve can be plotted, and the average precision (AP) metric can then be obtained as the area under the P-R curve [49], which can be used to measure the capability of crack detection by DO-YOLOv4.

4.2.2. Testing Results and Analysis

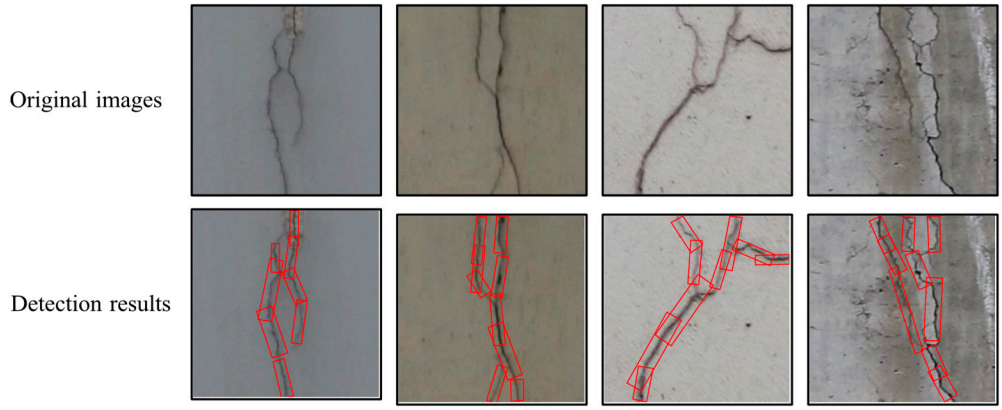
Several typical testing results of the trained DO-YOLOv4 for crack detection are depicted in Figure 10, from which it can be observed that various inclined cracks, including those with individual directions, with different extension directions, and even with multiple branches, can be well-located and tightly enclosed by a series of oriented bounding boxes. This indicates that DO-YOLOv4 is capable of adjusting to a wide variety of cracks with different sizes and directions and multiple branches, even in the presence of background interferences.



(**a**) Inclined cracks with individual directions



(**b**) Inclined cracks with different extension directions



(**c**) Inclined cracks with multiple branches

Figure 10. Detection results of cracks by DO-YOLOv4.

To better evaluate the performance of DO-YOLOv4, its feature extraction network and crack detection accuracy were compared with those of the traditional object detection methods, including Faster R-CNN, SSD, YOLOv3, and YOLOv4. For the feature extraction

network, the improved CSPDarknet53 is adopted in DO-YOLOv4, in which the original convolutional layers connecting different CSPblocks are replaced by the deformable convolutional layers, and a total of 15 resblocks can be reduced, as shown in Figures 3 and 4. Therefore, as shown in Table 1, the size of improved CSPDarknet53 accounts for roughly 1/10, 1/2, 1/3, and 1/2 of the size of VGG16, Resnet50, Darknet53, and CSPDarkent53, respectively, leading to the lightweight feature of DO-YOLOv4. For the detection accuracy, as multiple oriented bounding boxes are employed for crack detection, the area under the P-R curve of DO-YOLOv4 is much larger compared to those of the traditional object detection methods, as depicted in Figure 11. Correspondingly, the AP metric of DO-YOLOv4 calculated using the method stated in Section 4.2.1 reaches up to 80.43%, as also shown in Table 1, which is 21.54%, 39.39%, 35.27% and 27.33% higher than the AP metrics of Faster R-CNN, SSD, YOLOv3 and YOLOv4, respectively, indicating the high accuracy of DO-YOLOv4.

Table 1. Comparison of different object detection methods.

Method	Feature Extraction Network (Size)	Bounding Box	AP (%)
Faster R-CNN	VGG16 (527.8 MB)	Single, horizontal	58.89
SSD	Resnet50 (89.67 MB)	Single, horizontal	41.04
YOLOv3	Darknet53 (154.82 MB)	Single, horizontal	45.16
YOLOv4	CSPDarknet53 (101.5 MB)	Single, horizontal	53.10
DO-YOLOv4	Improved CSPDarknet53 (53.3 MB)	Multiple, oriented	80.43

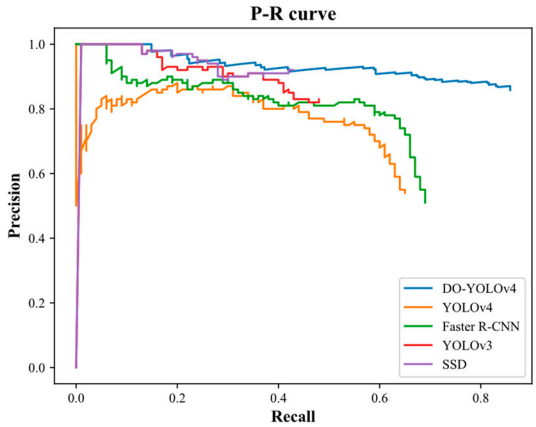


Figure 11. The P-R curve for different object detection methods.

4.3. Evaluation of DO-YOLOv4-IPTs for Crack Segmentation

Once the cracks are detected by DO-YOLOv4 using multiple oriented bounding boxes, the IPTs are further employed in the regions enclosed by those boxes for crack segmentation, by which the crack sizes, e.g., the lengths, widths, and areas, etc., can be quantified on the pixel level.

To evaluate the performance of DO-YOLOv4-IPTs for crack segmentation, several existing CNN-based crack segmentation methods, including FCN [9], Unet [10], CrackSegNet [64] and CrackPix [65], were selected for comparison study, and the results of different methods are depicted in Figure 12, in which the crack areas are denoted in terms of the numbers of pixels in the brackets. It can be seen from image (a) and image (b) of Figure 12 that, for the cases with fewer background interferences, CrackSegNet, CrackPix, and DO-YOLOv4-IPTs perform well with similar segmentation results, while certain discrepancies can be observed in the results of FCN and Unet. For the very tiny crack with background interferences shown in image (c) of Figure 12, all four CNN-based methods fail to identify the crack, while DO-YOLOv4-IPTs is still capable of segmenting the crack correctly. Furthermore, to investigate the influence of crack-like interferences, the crack images shown in image (d) and image (e) of Figure 12 were employed for crack segmentation. It can be observed that the

CNN-based methods are very sensitive to crack-like interferences, and over-segmentation results were more or less obtained by the CNN-based methods. Whereas DO-YOLOv4-IPTs still exhibits ideal performance under crack-like interferences.

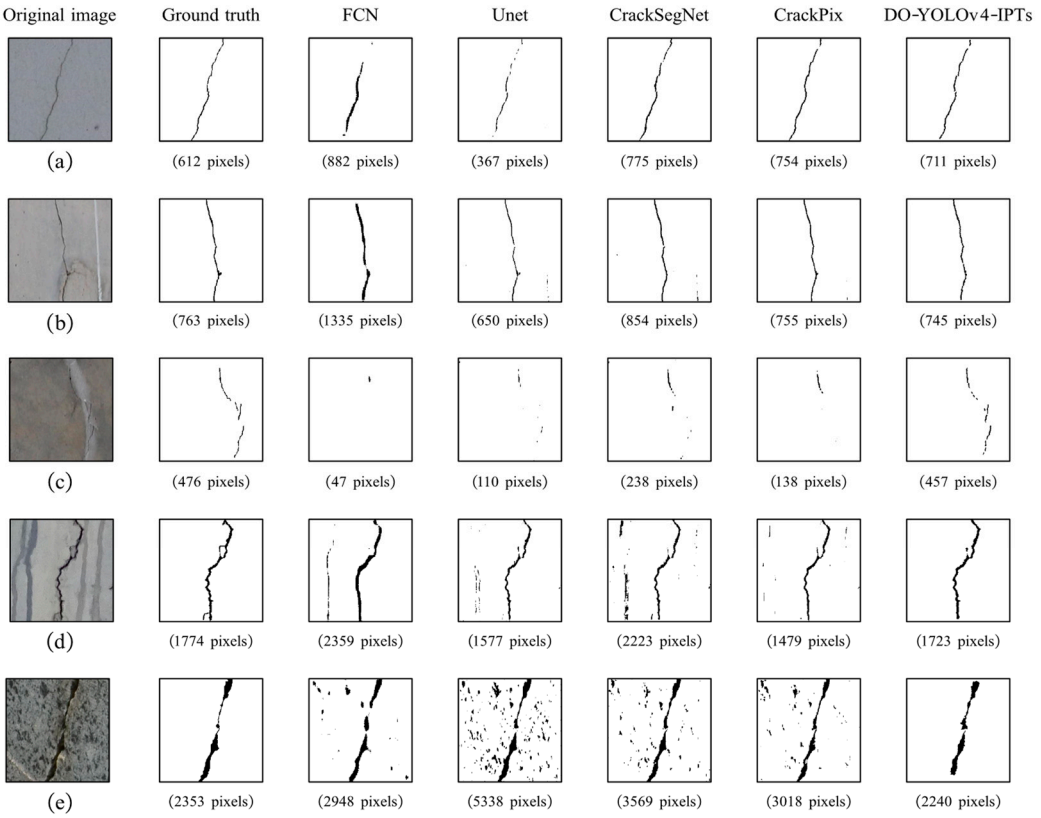


Figure 12. Crack segmentation results of CNN-based methods and DO-YOLOv4-IPTs. (Note: The number of pixels in the bracket is used to denote the crack area).

To quantify the performance of different crack segmentation methods, the pixel-level IoU (PIoU) metric is defined herein as the ratio of intersection to union of the ground-truth crack area and the segmented crack area in terms of the pixel numbers. The PIoU metrics of different methods for crack segmentation of the images shown in Figure 12 are presented in Table 2, from which it can be seen that, in general, DO-YOLOv4-IPTs has higher PIoU metrics than CNN-based methods, in particular for the case of image (c). The mean PIoU metrics of different segmentation methods are also presented in Table 2, from which it can be found that DO-YOLOv4-IPTs has the highest mean PIoU metric, as expected. In summary, DO-YOLOv4-IPTs can maintain a good standard for crack segmentation even with the presence of complex background within the crack images, which is mainly due to the high accuracy of DO-YOLOv4 for crack detection, as discussed in Section 4.2.

Table 2. Comparison of CNN-based methods and DO-YOLOv4-IPTs for crack segmentation.

Method	PIoU (%)					Mean PIoU (%)	Mean Time Cost of Labeling (min)
	Image (a)	Image (b)	Image (c)	Image (d)	Image (e)		
FCN	39.10	47.43	6.95	52.96	55.73	40.44	5
Unet	54.17	69.63	18.38	65.32	40.35	49.57	5
CrackSegNet	64.53	70.39	21.43	49.98	53.54	51.97	5
CrackPix	68.85	77.54	19.22	67.08	52.67	57.07	5
DO-YOLOv4-IPTs	72.49	75.14	61.42	78.24	70.93	71.64	0.5

In addition to the ideal accuracy of DO-YOLOv4-IPTs, the mean time cost of labeling of the present hybrid approach in this experiment accounts for roughly 1/10 of those of the CNN-based methods, as also shown in Table 2. The reason for this lies in the fact that only box annotations are required for crack detection in DO-YOLOv4-IPTs, while for the CNN-based methods, time-consuming pixel annotations are needed for crack segmentation.

5. Conclusions

To avoid the time-consuming process of pixel-level labeling brought by CNN-based crack segmentation methods and to improve the performance of crack segmentation of concrete structures with the presence of background interferences, a novel hybrid approach is presented in this study by combining DO-YOLOv4 for crack detection and IPTs for crack segmentation. Owing to the use of deformable convolutional layers and multiple oriented bounding boxes, the proposed DO-YOLOv4 has a strong learning ability for crack characteristics with its lightweight network and is capable of locating a wide variety of cracks at high accuracy, even with background noise. Following this, the present hybrid approach DO-YOLOv4-IPTs can effectively reduce the labeling cost by means of box annotations and meanwhile improve the performance of crack segmentation, in particular for the images with very tiny cracks and/or crack-like interferences.

For the experiments conducted in the present study, the AP metric of DO-YOLOv4 goes as high as 80.43% and is 21.54%, 39.39%, 35.27% and 27.33% higher than those of Faster R-CNN, SSD, YOLOv3 and YOLOv4, respectively, indicating the high accuracy of DO-YOLOv4 for crack detection, and with that, the hybrid approach DO-YOLOv4-IPTs turns out to have the highest mean PIoU metric of 71.64% with only 1/10 of the mean time cost of labeling for the traditional methods, showing the promising performance of crack segmentation by DO-YOLOv4-IPTs.

Despite the success achieved above, it has been observed that the parameters adopted in IPTs, namely the mean pixel intensity I_{m}, the constant C, and the area threshold AR_{th}, need to be determined by manual intervention with prior knowledge. The values of such parameters will have a certain influence on the crack segmentation results by IPTs within the bounding boxes. The adaptive optimization of such parameters is crucial to the automatic crack segmentation under different circumstances, which deserves further research for the present approach.

Furthermore, the robust learning ability facilitated by the deformable convolution and the oriented bounding boxes can be further utilized to address object detection challenges in other fields, such as retinal vessel detection, multi-class concrete defect identification, etc.

Author Contributions: Z.H.: Methodology (lead), Formal analysis, Investigation, Writing—original draft. C.S.: Resources, Methodology, Validation, Writing—review and revising, Funding acquisition. Y.D.: Conceptualization, Methodology, Writing—review and editing, Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The research is funded by the National Natural Science Foundation of China (52308314), Guangdong Provincial Key Laboratory of Modern Civil Engineering Technology (2021B1212040003), Guangdong Basic and Applied Basic Research Foundation (2022A1515010174, 2023A1515030169), Guangzhou Science and Technology Program (202201010338), and Guangzhou Science and Technology Program (201804020069).

Data Availability Statement: The code of DO-YOLOv4-IPTs has been uploaded to the website https://github.com/DO-YOLOv4-IPTs/main (accessed on 8 February 2024), allowing interested readers to apply it in their respective research fields.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Hu, W.B.; Wang, W.D.; Ai, C.B.; Wang, J.; Wang, W.J.; Meng, X.F.; Liu, J.; Tao, H.W.; Qiu, S. Machine vision-based surface crack analysis for transportation infrastructure. *Autom. Constr.* **2021**, *132*, 103973. [CrossRef]
2. Ali, R.; Chuah, J.H.; Talip, M.S.A.; Mokhtar, N.; Shoaib, M.A. Structural crack detection using deep convolutional neural networks. *Autom. Constr.* **2022**, *133*, 103989. [CrossRef]
3. Zhang, J.M.; Lu, C.Q.; Wang, J.; Wang, L.; Yue, X.G. Concrete Cracks Detection Based on FCN with Dilated Convolution. *Appl. Sci.* **2019**, *9*, 2686. [CrossRef]
4. Kang, D.H.; Cha, Y.J. Efficient attention-based deep encoder and decoder for automatic crack segmentation. *Struct. Health Monit.* **2022**, *21*, 2190–2205. [CrossRef]
5. Chen, J.; He, Y. A novel U-shaped encoder-decoder network with attention mechanism for detection and evaluation of road cracks at pixel level. *Comput. Aided Civ. Infrastruct. Eng.* **2022**, *37*, 1721–1736. [CrossRef]
6. Tong, Z.; Yuan, D.D.; Gao, J.; Wang, Z.J. Pavement defect detection with fully convolutional network and an uncertainty framework. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 832–849. [CrossRef]
7. Wang, W.J.; Su, C.; Han, G.H.; Zhang, H. A lightweight crack segmentation network based on knowledge distillation. *J. Build. Eng.* **2023**, *76*, 107200. [CrossRef]
8. Zhu, Y.; Tang, H. Automatic Damage Detection and Diagnosis for Hydraulic Structures Using Drones and Artificial Intelligence Techniques. *Remote Sens.* **2023**, *15*, 615. [CrossRef]
9. Yang, X.C.; Li, H.; Yu, Y.T.; Luo, X.C.; Huang, T. Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. *Comput. Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1090–1109. [CrossRef]
10. Liu, Z.Q.; Cao, Y.W.; Wang, Y.; Wang, W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom. Constr.* **2019**, *104*, 129–139. [CrossRef]
11. Lee, S.J.; Hwang, S.H.; Choi, I.Y.; Choi, Y. Estimation of crack width based on shape-sensitive kernels and semantic segmentation. *Struct. Control. Health Monit.* **2020**, *27*, e2504. [CrossRef]
12. Ni, F.T.; He, Z.L.; Jiang, S.; Wang, W.G.; Zhang, J. A Generative adversarial learning strategy for enhanced lightweight crack delineation networks. *Adv. Eng. Inform.* **2022**, *52*, 101575. [CrossRef]
13. Tabernik, D.; Suc, M.; Skocaj, D. Automated detection and segmentation of cracks in concrete surfaces using joined segmentation and classification deep neural network. *Constr. Build. Mater.* **2023**, *408*, 133582. [CrossRef]
14. Miao, P.Y.; Srimahachota, T. Cost-effective system for detection and quantification of concrete surface cracks by combination of convolutional neural network and image processing techniques. *Constr. Build. Mater.* **2021**, *293*, 123549. [CrossRef]
15. Kang, D.H.; Benipal, S.S.; Gopal, D.L.; Cha, Y.J. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. *Autom. Constr.* **2020**, *118*, 103291. [CrossRef]
16. Li, S.Y.; Zhao, X.F. Pixel-level detection and measurement of concrete crack using faster region-based convolutional neural network and morphological feature extraction. *Meas. Sci. Technol.* **2021**, *32*, 065010. [CrossRef]
17. Li, C.; Xu, P.J.; Niu, L.J.L.; Chen, Y.; Sheng, L.S.; Liu, M.C. Tunnel crack detection using coarse-to-fine region localization and edge detection. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1308. [CrossRef]
18. Dai, J.; He, K.; Sun, J. BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. In Proceedings of the International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1635–1643.
19. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, Canada, 7–12 December 2015; pp. 1137–1149.
20. 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 European Conference on Computer Vision*; Springer: Berlin, Germany, 2016; pp. 21–37.
21. 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, LasVegas, NV, USA, 27–30 June 2016; pp. 779–788.
22. Bochkovskiy, A.; Wang, C.; Liao, H.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
23. Ma, D.; Fang, H.Y.; Xue, B.H.; Wang, F.M.; Msekn, M.A.; Chan, C.L. Intelligent detection model based on a fully convolutional neural network for pavement cracks. *Comput. Model. Eng. Sci.* **2020**, *123*, 1267–1291. [CrossRef]
24. Pang, J.; Zhang, H.; Feng, C.C.; Li, L.J. Research on crack segmentation method of hydro-junction project based on target detection network. *KSCE J. Civ. Eng.* **2020**, *24*, 2731–2741. [CrossRef]
25. Xu, G.Y.; Han, X.; Zhang, Y.W.; Wu, C.Y. Dam Crack Image Detection Model on Feature Enhancement and Attention Mechanism. *Water* **2023**, *15*, 64. [CrossRef]
26. Yang, J.; Fu, Q.; Nie, M.X. Road Crack Detection Using Deep Neural Network with Receptive Field Block. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *782*, 042033. [CrossRef]
27. Zhang, J.; Qian, S.R.; Tan, C. Automated bridge surface crack detection and segmentation using computer vision-based deep learning model. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105225. [CrossRef]
28. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.

29. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
30. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
31. Cha, Y.J.; Choi, W.; Suh, G.; Mahmoudkhani, S.; Buyukozturk, O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 731–747. [CrossRef]
32. Deng, J.H.; Lu, Y.; Lee, V.C.S. Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network. *Comput. Aided Civ. Infrastruct. Eng.* **2019**, *35*, 373–388. [CrossRef]
33. Li, D.W.; Xie, Q.; Gong, X.X.; Yu, Z.H.; Xu, J.X.; Sun, Y.X.; Jun, W. Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. *Adv. Eng. Inform.* **2021**, *47*, 101206. [CrossRef]
34. Teng, S.; Liu, Z.C.; Chen, G.F.; Cheng, L. Concrete Crack Detection Based on Well-Known Feature Extractor Model and the YOLO_v2 Network. *Appl. Sci.* **2021**, *11*, 813. [CrossRef]
35. Du, Y.C.; Pan, N.; Xu, Z.H.; Deng, F.W.; Shen, Y.; Kang, H. Pavement distress detection and classification based on YOLO network. *Int. J. Pavement Eng.* **2021**, *22*, 1659–1672. [CrossRef]
36. Cao, M.T.; Tran, Q.V.; Nguyen, N.M.; Chang, K.T. Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. *Adv. Eng. Inform.* **2020**, *46*, 101182. [CrossRef]
37. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
38. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
39. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.M.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
40. Zhang, C.B.; Chang, C.C.; Jamshidi, M. Concrete bridge surface damage detection using a single-stage detector. *Comput. -Aided Civ. Infrastruct. Eng.* **2020**, *35*, 389–409. [CrossRef]
41. Yu, Z.W.; Shen, Y.G.; Shen, C.K. A real-time detection approach for bridge cracks based on YOLOv4-FPM. *Autom. Constr.* **2021**, *122*, 103514. [CrossRef]
42. Zhou, Z.; Zhang, J.J.; Gong, C.J. Automatic detection method of tunnel lining multi-defects via an enhanced You Only Look Once network. *Comput. Aided Civ. Infrastruct. Eng.* **2022**, *37*, 762–780. [CrossRef]
43. 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]
44. Ayenu-Prah, A.; Attoh-Okine, N. Evaluating Pavement Cracks with Bidimensional Empirical Mode Decomposition. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 861701. [CrossRef]
45. Abdel-Qader, I.; Abudayyeh, O.; Kelly, M.E. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. *J. Comput. Civil Eng.* **2003**, *17*, 255–263. [CrossRef]
46. Li, Q.Q.; Zou, Q.; Zhang, D.Q.; Mao, Q.Z. FoSA: F* Seed-growing Approach for crack-line detection from pavement images. *Image Vis. Comput.* **2011**, *29*, 861–872. [CrossRef]
47. Zhou, Y.X.; Wang, F.; Meghanathan, N. Seed-Based Approach for Automated Crack Detection from Pavement Images. *Transp. Res. Recode* **2016**, *2589*, 162–171. [CrossRef]
48. Gavilan, M.; Balcones, D.; Marcos, O.; LIorca, D.F. Adaptive Road Crack Detection System by Pavement Classification. *Sensors* **2011**, *11*, 9628–9657. [CrossRef] [PubMed]
49. Zou, Q.; Cao, Y.; Li, Q.Q.; Mao, Q.Z.; Wang, S. CrackTree: Automatic crack detection from pavement images. *Pattern Recognit. Lett.* **2012**, *33*, 227–238. [CrossRef]
50. Tang, J.S.; Gu, Y.L. Automatic Crack Detection and Segmetnation Using A Hybrid Algorithm for Road Distress Analysis. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 3026–3030.
51. Li, Q.Q.; Liu, X.L. Novel approach to pavement image segmentation based on neighboring difference histogram method. In Proceedings of the 1st International Congress on Image and Signal Processing, Sanya, China, 27–30 May 2008; pp. 792–796.
52. Oliveira, H.; Correia, P. Automatic road crack segmentation using entropy and image dynamic thresholding. In Proceedings of the 17th European IEEE Signal Processing Conference, Glasgow, Scotland, 24–28 August 2009; pp. 622–626.
53. Jin, Q.G.; Meng, Z.P.; Pham, T.D.; Chen, Q.; Wei, L.Y.; Su, R. DUNet: A deformable network for retinal vessel segmentation. *Knowl. Based Syst.* **2019**, *178*, 149–162. [CrossRef]
54. Wang, Y.H.; Ye, S.J.; Bai, Y.; Gao, G.M.; Gu, Y.F. Vehicle detection using deep learning with deformable convolution. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 2329–2332.
55. Zhao, S.L.; Zhang, S.; Lu, J.M.; Wang, H.; Feng, Y.; Shi, C.; Li, D.L.; Zhao, R. A lightweight dead fish detection method based on deformable convolution and YOLOV4. *Comput. Electron. Agric.* **2022**, *198*, 107098. [CrossRef]
56. Ye, X.W.; Jin, T.; Li, Z.X.; Ma, S.Y.; Ding, Y. Structural Crack Detection from Benchmark Data Sets Using Pruned Fully Convolutional Networks. *J. Struct. Eng.* **2021**, *147*, 04721008. [CrossRef]
57. RoLabelImg. 2020. Available online: https://github.com/roLabelImg-master (accessed on 29 June 2020).

58. PyTorch. Available online: https://pytorch.org (accessed on 1 January 2017).
59. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. *arXiv* **2015**, arXiv:1412.6980.
60. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12993–13000. [CrossRef]
61. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef]
62. OpenCV. Available online: https://opencv.org (accessed on 4 July 2011).
63. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
64. Ren, Y.P.; Huang, J.S.; Hong, Z.Y.; Lu, W.; Yin, J.; Zou, L.J.; Shen, X.H. Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Constr. Build. Mater.* **2020**, *234*, 117367. [CrossRef]
65. Alipour, M.; Harris, D.K.; Miller, G.R. Robust Pixel-Level Crack Detection Using Deep Fully Convolutional Neural Networks. *J. Comput. Civil Eng.* **2019**, *33*, 04019040. [CrossRef]