



Vision-based concrete crack detection using deep learning-based models

Elham Nabizadeh¹ · Anant Parghi²

Received: 11 March 2023 / Accepted: 31 March 2023 / Published online: 11 April 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

It is critical to detect cracks in concrete promptly and effectively to limit further deterioration and to perform timely repairs. Several convolutional neural networks have been proposed in recent years for identifying objects varying in their accuracy and speed. In this study, the YOLOv7, YOLOv5s, YOLOv5m, and YOLOv5x object identification models were trained for crack detection in concrete surfaces. The networks were trained using 1600 images of concrete cracks and analyzed. The different YOLOv5 versions and YOLOv7 are compared using assessment measures including F1 score, recall, and mAP. The research study found that all of the models predicted encouraging results in terms of crack detection in concrete images. According to the results, YOLOv5m and YOLOv5x achieved F1 scores of 0.87 and 0.86, respectively. Differently, YOLOv5s and YOLOv7 acquired an F1-score of 0.85 and 0.84, respectively. As a result, this research verifies the recently introduced deep learning technology, which can replace conventional crack detection and identification techniques with more reliable and efficient alternatives.

Keywords Concrete · Crack detection · Deep learning · Computer vision

Introduction

Infrastructure is essential to any community, such as roads, bridges, and buildings. Structural faults and environmental changes can negatively affect a structure's dependability over time, emphasizing the need for monitoring infrastructure. Protecting the integrity of buildings requires identifying any responsibility as early as possible. The versatility of concrete makes it one of the most frequently used materials in civil engineering. In concrete structures and pavements, cracking is a common problem, and this structural deterioration can threaten safety and reduce service life. There are, however, several causes that can lead to concrete fractures.

Consequently, it is critical to determine the location of fractures in concrete structures to evaluate their structural safety. The most common and earliest method of assessing a system's health is visual inspection, which is expensive and labor intensive. Furthermore, manual review relies on expert judgment and requires highly qualified professionals. Due to these restrictions, industry and research have developed methods for automatic crack detection. Artificial intelligence advancements have made it possible to autonomously detect concrete cracks and structural damage. As a subfield of machine learning, the use of deep learning to handle enormous volumes of data has been shown on a variety of platforms. Three categories of research projects that have utilized deep learning to find fractures include picture classification, object identification, and semantic segmentation. With the help of hidden information in the data, a deep learning model with layers layered on top of one another may be utilized to predict unique patterns. R-CNN, YOLO, and SSD are a few of the architectures that are utilized to accomplish object detection in computer vision. In YOLO designs, the object detection process is complete after one forward propagation of the picture.

Du et al. (8) utilized the YOLO network to swiftly classify and identify pavement degradation using a novel method

✉ Elham Nabizadeh
ElhamNabizadeh20@gmail.com

Anant Parghi
amp@amd.svnit.ac.in

¹ Department of Civil Engineering, Hakim Sabzevari University, Sabzevar, Iran

² Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat 395007, India

for pavement distress identification. Teng et al. (31) carried out the YOLOv2 crack detection technique employing 11 CNN models as the feature extractors. Song Park et al. (27) used YOLO for real-time fracture identification and evaluated crack diameters based on the placements of the laser beams on the structural surface. Yu (36) used a threshold segmentation technique based on the Otsu maximum inter-class variance to link the grayscale change points in crack pictures, merge the noise points with the background, and remove background noise from the photos. This work uses deep learning based on Yolov5s to find crack information in crack pictures. Cui et al. (6) have created a YOLO-v3 algorithm model that is more effective. Yolo-v3 has been enhanced, and it has been discovered that it can more accurately detect concrete erosion damage. Using the edge computing paradigm, Kumar et al. (20) propose a real-time multi-drone damage detection system that employs YOLO-v3 to detect damage to high-rise civil structures. Kim and Cho (19) proposed a vision-based crack detection algorithm to classify and identify cracks using transfer learning accurately. They trained the AlexNet model on the image dataset, which consisted of two subsequent image classes. Based on natural frequencies and mode shapes, Kaveh and Maniat (15) investigated the mechanism for identifying structural damage. MCSS and PSO techniques were used to overcome the optimization problem associated with damage identification. A U-Net network containing an encoder and decoder framework was presented for automatic crack identification by Andrushia et al. (1). They compared the proposed approach to existing state-of-the-art methods and found that it was accurate with an Intersection over Union of 78.12%. Ye et al. (35) gathered a significant number of photos of concrete cracks, and they proposed STCNet I, a deep learning-based architecture to detect concrete cracks in slabs. Han et al. (9)

used CNNs and digital image processing to detect cracks in images. They trained the AlexNet-based CNN network efficiently and achieves 98.26% accuracy on test data. Kaveh and Zolghadr (18) integrated a developed multi-agent meta-heuristic method named CPA with a guided modal strain energy-based structural damage detection system. Chow et al. (5) present a one-stage automated detection approach for concrete surface defects. EfficientNetB0, the backbone network, and the detector are two parts of their model. The average accuracy for cracks and exposed bars is 76.4% and 89.9%, respectively, according to the results. Bang et al. (3) suggested that structural degeneration can be identified and quantified using deep learning by employing structured lights and a depth sensor. For detecting three types of surface defects, they used Faster R-CNN with Inception Resnet v2. According to the results, the proposed method can identify structural damage with high accuracy. Kang et al. (12) used concrete surface images with complicated backgrounds to evaluate the overall performance of the advanced hybrid concrete crack segmentation technique. Kaveh and Mahdavi (16) assessed the application of two meta-heuristic algorithms for identifying steel truss damage. Results show that for determining the specific location of damaged structural components, the ECBO algorithm outperforms the CBO approach. Joshi et al. (10) presented a deep learning-based architecture for identifying and segmenting concrete-based material defects. For training and testing, they used an image dataset of 3000 surface cracks. Each image's cracks were manually labeled using a bounding box and segmented mask. Yang et al. (33) used AlexNet, VGGNet13, and ResNet18 to detect and recognize structural cracks. They then trained the YOLOv3 model on crack images to identify crack targets. For concrete crack detection, Wan et al. (32) developed a novel strategy combining deep learning with a single-shot

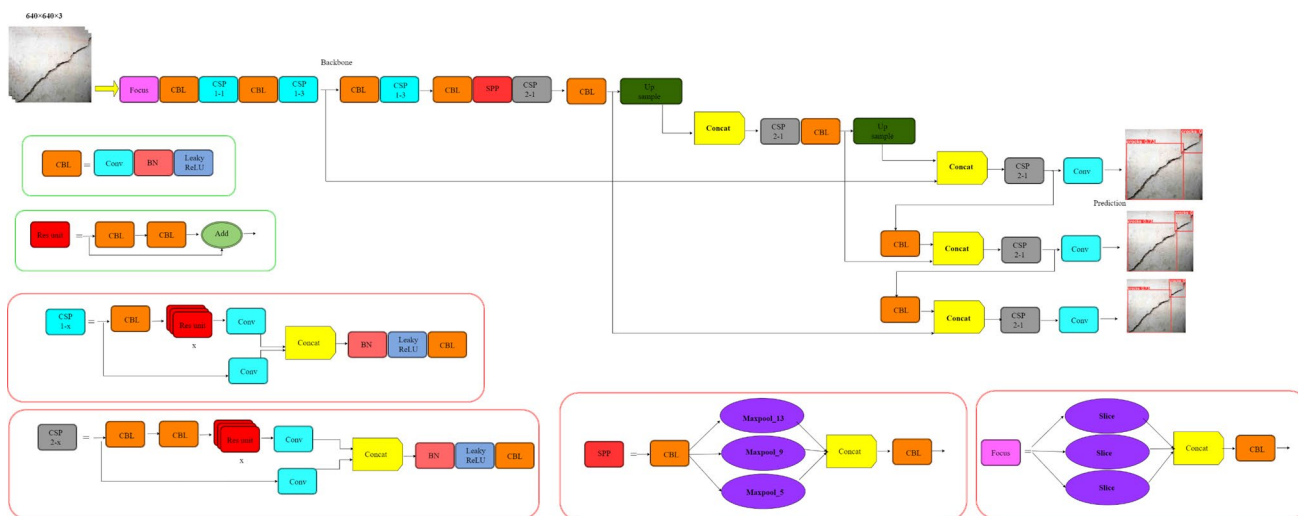


Fig. 1 YOLOv5-based detection network design

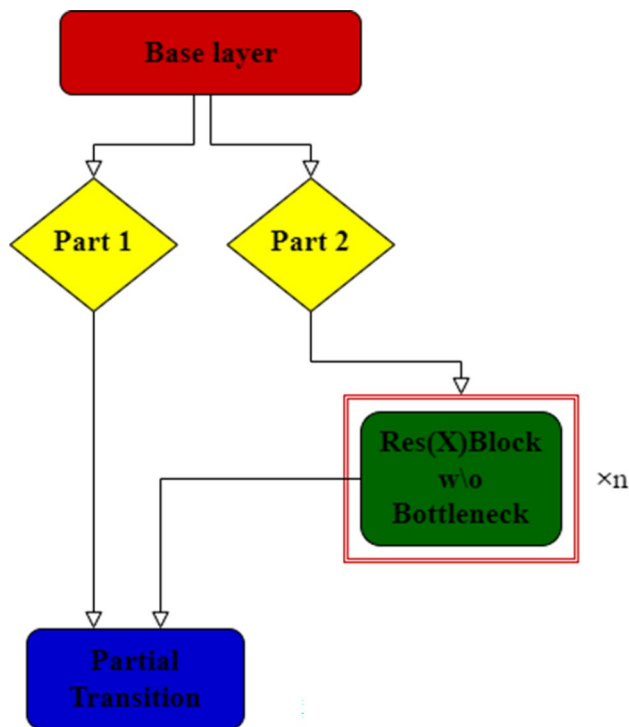


Fig. 2 CSPNet network struct

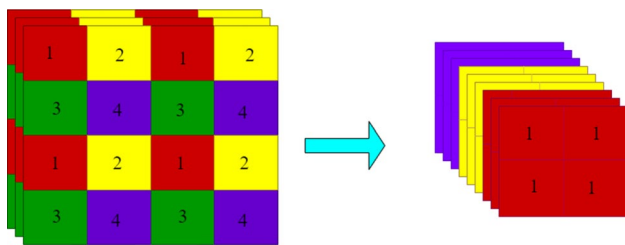


Fig. 3 Focus network struct

multi box detector. As a result of the suggested method, the crack identification process is significantly improved. Li et al., (21, 22) recommended ATCrack for automatic crack detection. In ATCrack, a symmetrical structure consisting of an encoder and a decoder is used to predict end-to-end cracks. Kaveh (13) evaluated the using of meta-heuristic algorithms to address specific key optimization issues in civil engineering. His research looked into how recently developed meta-heuristic algorithms can be applied in various actual situations. Liu et al. (23) presented a U-Net-based method for detecting concrete cracks. Based on comparisons between U-Net and DCNN, U-Net is found to be more elegant, more robust, and more effective than DCNN. Based on the result, the AP of the Faster-R-CNN process was 95%. Liu et al. (24) presented a generative adversarial network for automatic crack inspection. Compared to the deblurring

model, the proposed model achieves remarkable improvements in concrete crack images. Kaveh and Zolghadr (17) examined structural damage detection utilizing changes in natural frequencies, presented as an inverse optimization problem. Ren et al. (29) proposed CrackSegNetto carry out dense pixel-wise crack segmentation. Compared to traditional image processing and existing deep learning-based crack segmentation algorithms, the proposed model exhibits much greater accuracy. Kaveh and Dadras (14) applied an enhanced variation of the optimization method known as the TEO algorithm to solve a damage detection problem. The locations and extents of damage are accurately identified several scenarios using noise-and noise-free modal input. Li et al., (21, 22) employed various image recognition networks for verification, and the YOLO-v4 model is used as the main body of the lightweight convolutional neural network. A combination of the YOLOv5s-HSC algorithm and three-dimensional photogrammetric reconstruction was proposed by Zhao et al., (37) for the exact identification of damages in concrete dams. Mishra et al. (26) presented a two-stage automated concrete crack detection system based on YOLOv5. They used a total of 40,000 photos to train the deep learning model that was created.

YOLOv5 network architecture

The YOLO method is a well-known single-stage detection technique. The YOLO series method is a deep learning-based approach that detects targets more quickly and accurately than SSD and R-CNN. Many image object detection applications use the YOLO method. Initially, it was developed as a single-stage detection method by Redmon. It converts the object detection challenge into a regression problem by dividing the image into a grid and forecasting the object.

Redmon et al. (28) updated the YOLO algorithm and implemented a new algorithm called YOLOv2, which utilizes Darknet-19 as a feature extractor network and improves recall using anchor frames. Recently, the YOLO series was upgraded, and two new versions (YOLO version 5 and YOLO version 7) are available. Both versions have included state-of-the-art algorithms, making them more efficient and adaptable to object detection. They proposed a concrete crack detector based on the YOLO v7 and YOLOv5 networks. The results from the two models were compared among themselves. The second section presents an overview of the architecture of YOLOv5 and YOLOv7 (Lu et al., 25). The next Section discusses the network performance and database generation techniques. Next, a comparison and analysis of the actual detection results follow, along with the evaluation metrics and training outcomes. As a single-stage target detection algorithm,

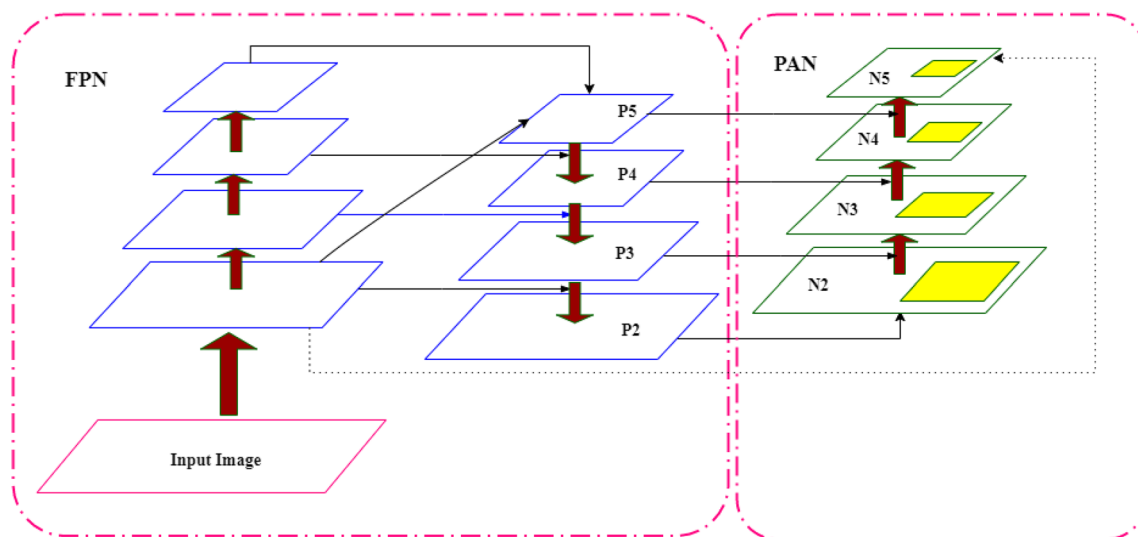


Fig. 4 PANet network structure

the YOLOv5 method has gained popularity because of its straightforward processes, high detection speed, and high accuracy. Figure 1 depicts the YOLOv5 network architecture.

Compared to YOLOv4, YOLOv5 continues to employ the CSP design and adds the CSP design to the backbone and neck to improve the network's ability to fusion features. The backbone of YOLOv5 uses the Focus network to divide the feature maps, increasing input channels by four while decreasing the algorithm's computational time. Each frame in the last YOLO series corresponds to one positive sample, and each existing structure can only be predicted by one previous frame during training. However, with YOLOv5, the number of positive samples is increased, and many preceding frames can anticipate every actual frame during training, which speeds up the model's training efficiency. The YOLOv5 series, as opposed to earlier iterations, comes in four versions: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Among these four versions of structures, the significant difference is achieved by varying the depth multiplier and width multiplier by varying the number of residual components in different cross-phase partial networks (CSPNs) to create networks of varying depths and by using varying numbers of convolution kernels in a variety of focusing structures and CSPNs to produce networks that are of various widths. The CSPDarknet architecture is employed as the backbone feature extraction network for YOLOv5. A residual network can be made more accurate by adding many depths. Jump connections are used by residual fast in YOLOv5 to overcome the gradient disappearance problem that comes with extending the network. As shown in Fig. 2, the backbone feature extraction network uses a CSPN to divide the

residual block stack into two parts. The central part stacks the residual blocks, while the remaining residual edges are attached to the end between them after some processing.

The Focus network structure is employed in the YOLOv5 feature extraction process. Figure 3 illustrates the Focus architecture and as shown in the figure. The slicing procedure is carried out before the input image reaches the backbone and each image is divided into four complementary images with no information lost by taking every other pixel from each image. Compared with the original RGB three-channel design, the input channels have been increased by four, the number of stitched-together images has been increased to 12, and the generated image is combined to produce a down-sampled image. The Focus network image slicing process converted the original 640×640 pixel image into 320×320 -pixel feature maps. Since YOLOv5m has 48 convolution kernels, an additional convolution operation is carried out to produce $320 \times 320 \times 48$ feature maps. As a result of the Focus network structure's slicing of input pictures, computations can be minimized and calculation performance can be boosted.

PANet networks are designed to enhance features by combining shallow and deep features. The FPN in YOLOv4 is based on the spatial pyramid pooling SPP module. In YOLOv5, the backbone feature extraction network is based on the SPP module. A-Max pool layer with various-size pooling kernels is used to pool the retrieved features to the maximum extent possible. The bottom route is directly connected rather than pooled, and the results are channel stitched to extend the network's perception range.

A total of three feature layers must be retrieved in the feature usage section to identify objects in YOLOv5. There are three feature layers in the feature extraction network

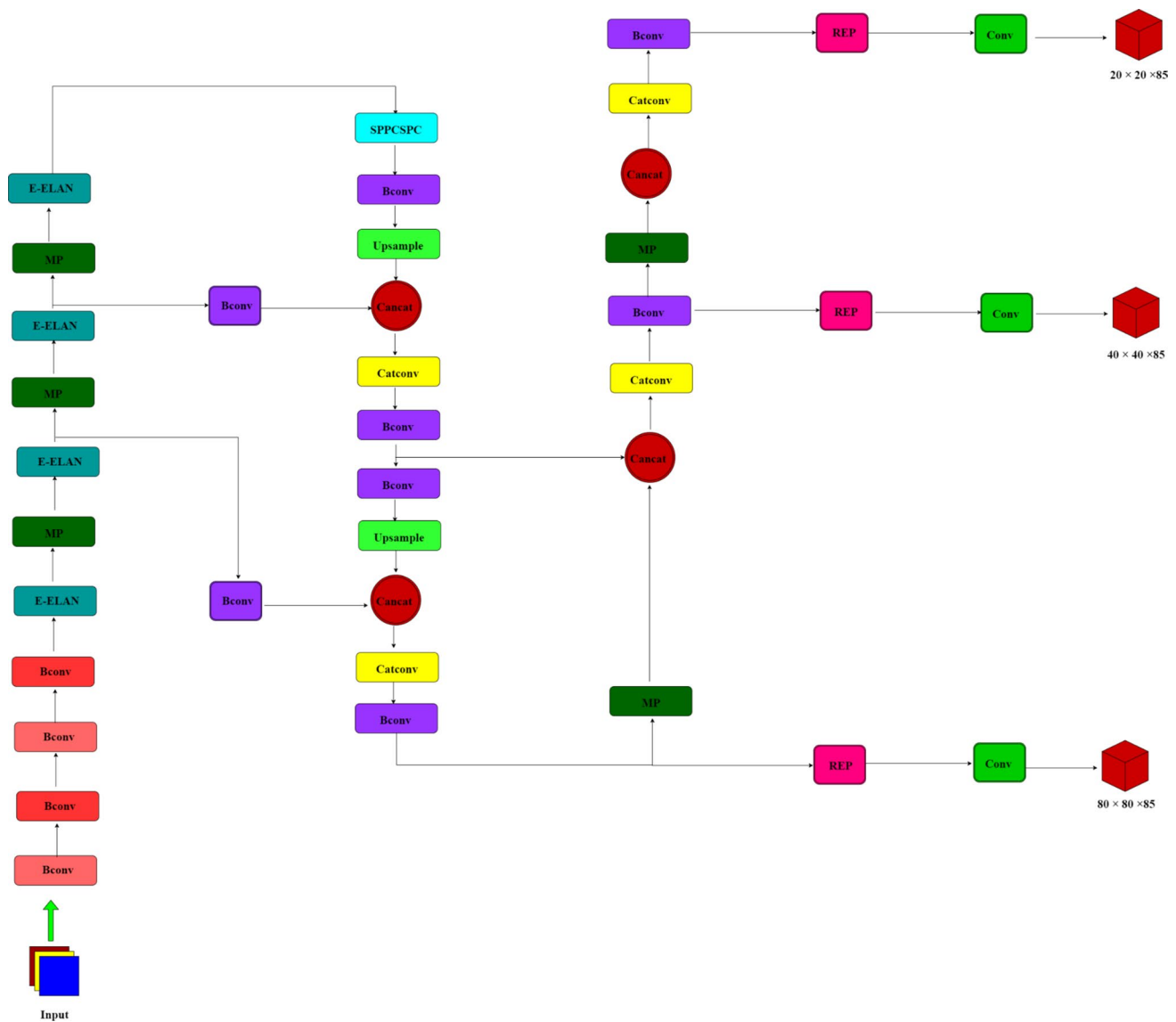


Fig. 5 YOLOv7 network structure

CSPDarknet network: the middle layer is CSP1_3, the lower intermediate layer known as CSP1_3, and the bottom layer known as CSP2_1. The FPN layer network can assist in feature extraction and feature fusion by various feature layers after acquiring three efficient feature layers. As seen in Fig. 4, PANet is a network design where P and N stand in for several layers of feature data. YOLOv5 employs the PANet design in three functional feature layers—CSP1_3, CSP1_3, and CSP2_1 to achieve the fusion of feature information from the three scale feature layers.

YOLOv7 network architecture

In this study, YOLOv7 is used for more precise crack detection by processing images of the concrete crack surface with an increased degree of speed and accuracy than synchronous techniques. Figure 5 illustrates the YOLOv7 network structure (localization and techniques 2022).

The preprocessing process for YOLOv7 is similar to that of YOLOv5, including mosaics, adaptive image scaling, and adaptive anchor boxes. YOLOv7 uses the Mosaic data improvement approach enhanced by the CutMix data improvement technique in the training process. Unlike Mosaic's data augmentation technique, which uses four randomly resized, arranged, and cropped, images, CutMix only requires two images. By combining several images into one,

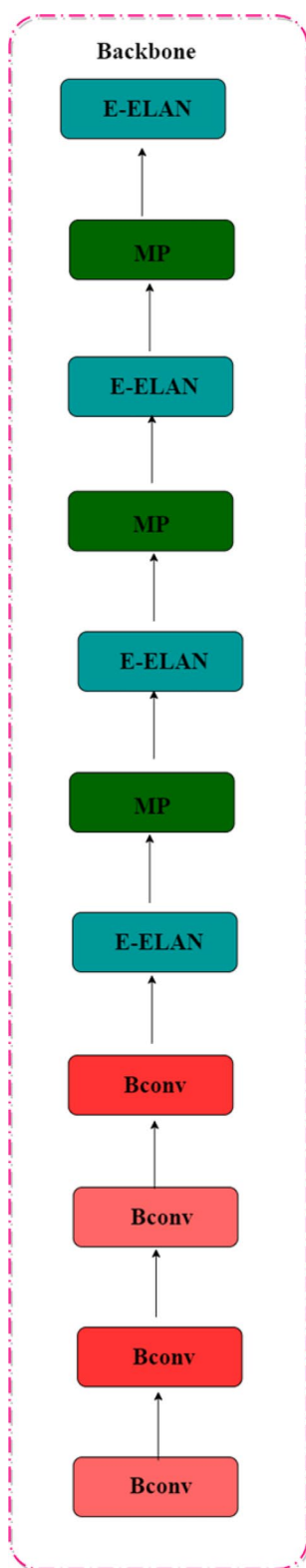


Fig. 6 Structure of the YOLOv7 backbone

this enhancement technique reduces network training time and memory usage while increasing dataset variety and network detection accuracy. YOLOv7's backbone layer consists of MP, E-ELAN, and Bconv layers in Fig. 6, while the convolution layer, BN layer, and activation function comprise the BConv layer. Figure 7 shows the schematic design of the BConv layer.

Different kernels' convolution layers are represented by different-colored Bconv modules. In the first BConv module, k and s are 1, Second BConv module has $k=3$ and $s=1$, whereas the third BConv module has $k=3$ and $s=2$. It should be emphasized that the colored Bconv modules primarily discriminate between k and s , but not between input and output. In YOLOv7, an extended ELAN is proposed based on ELAN to regulate gradients' shortest and longest paths, improving deep learning and convergence. By extending, shuffling, and merging cardinality, the YOLOv7 E-ELAN increases the network's learning capacity without obliterating its gradient path.

During the construction of E-ELAN, the structure of the block is modified on its own, while the architecture of the transition layer remains unchanged. A group convolution is applied to all computing blocks in the computing layer, including the number of channels, and expands their channel and cardinality. Figure 8 depicts the E-ELAN layer, which is similarly made up of several convolutions. A feature map is produced by each computational block, which is then subjected to the following procedures: connect the g groups that were created by randomly rearranging the g group parameters of the feature map. Each set of feature maps has the same number of channels as initially designed, and to integrate the cardinality, the g group's feature maps need to be added.

Figure 8 illustrates the different convolutions that make up the E-ELAN layer.

Figure 9 illustrates the structure of the MP layer. This branch has the same input and output channels, and the output's size (length and breadth) is half the input's length and width. Using maximum pooling, the top branch is first cutting the length and breadth in half, then cutting the channel in half with BConv. The bottom unit splits the channel in half using the first BConv, and the length and width are divided in half using the second BConv.

In YOLOv7, the head segment is similar to YOLOv5, the down sampling module is moved to the MPConv layer, and the CSP module is replaced with E-ELAN. The head layer consists of SPPCPC layers, numerous BConv layers, numerous MPConv layers, numerous Catconv layers, and RepVGG block layers that output three heads sequentially. The diagram of YOLOv7's Head section is shown in Fig. 10.

The SPPCSPC layer module is created using the pyramid pooling method and the CSP structure, and it still has numerous branches. The input will be split into three halves and

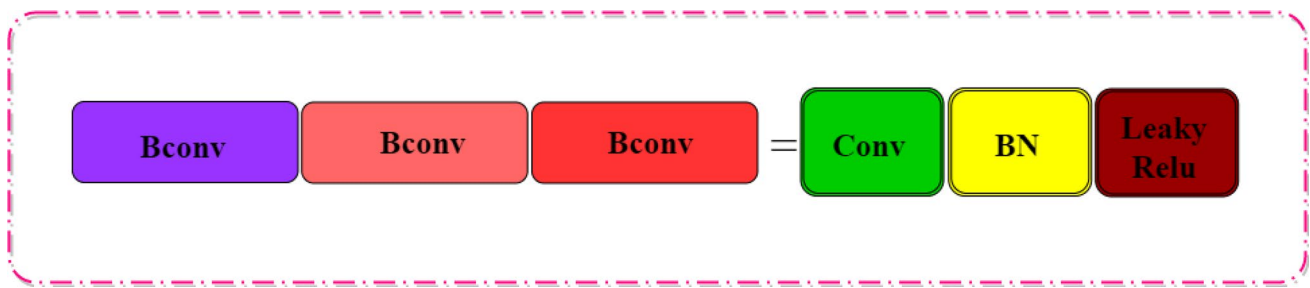


Fig. 7 Structure of the BConv layer

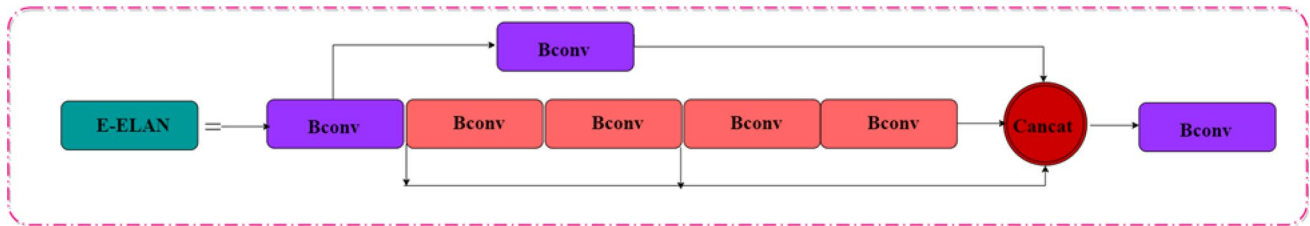


Fig. 8 E-ELAN layer structure

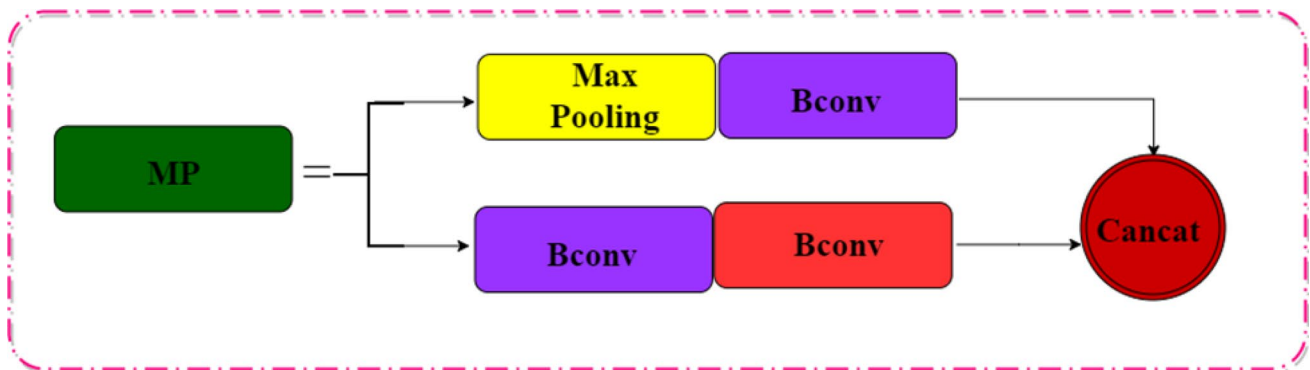


Fig. 9 MP lay structure

distributed across several units, and the output will be the concatenated information. Figure 11 shows the architecture of the SPCCSPC layer.

Catconv, another layer enabling deeper networks to learn and converge more effectively, operates similar to E-ELAN, and Fig. 12 illustrates how it works.

The REP layer's structure differs during deployment and training. Training adds a 1×1 convolution branch based on a 3×3 convolution. When adding a BN branch, three output branches are added if the input and output channels and h and w are the same. The unit's parameters are re-parameterized and assigned to the main branch, and the 3×3 convolution output is used in Fig. 13.

Performance metrics

A variety of metrics, including precision, recall F1 score, and mean average precision (mAP), is used for evaluating Deep learning models. These metrics provide insights into the performance of deep learning models and can be used to compare different models. Precision is a measure of the accuracy of the classifier when it predicts positive examples. Recall is a measure of the ability of the classifier to find all positive samples. The F1 score is a weighted average of precision and recall. The mAP is an overall measure of performance that considers both precision and recall models that can be tuned for specific applications by optimizing for these metrics. For example, if high precision is desired, then the model should be

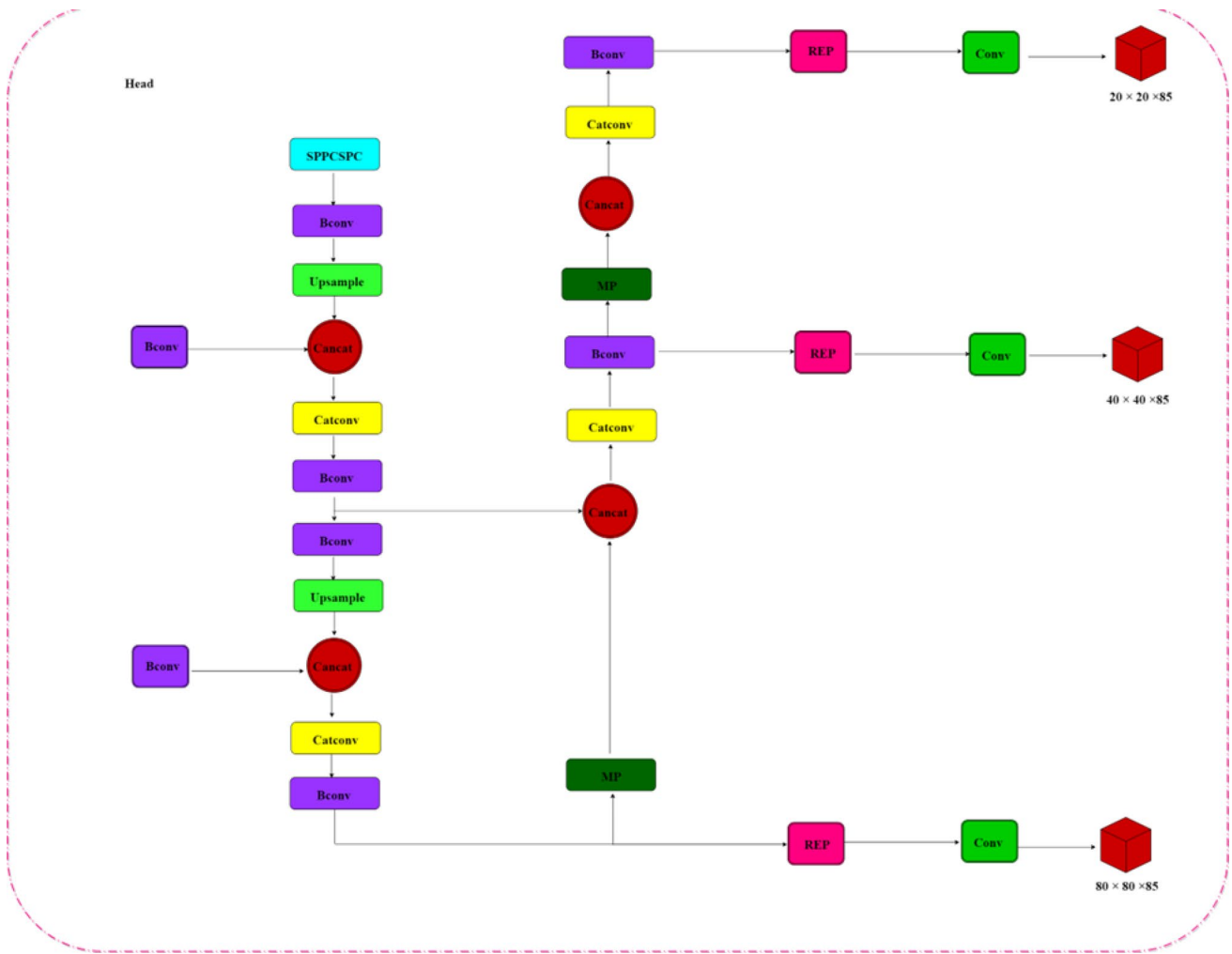


Fig. 10 YOLOv7 head structure

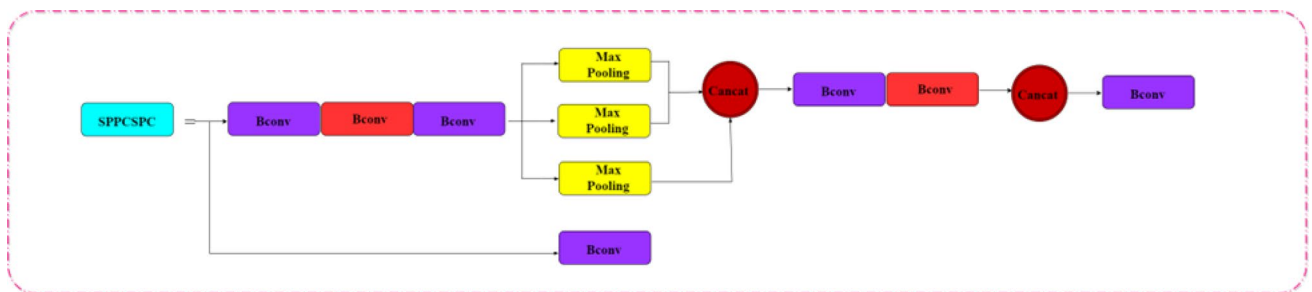


Fig. 11 SPPCPC layer structure

tuned to achieve high precision. Conversely, if a high recall is desired, then the model should be tuned to achieve high recall. In general, the F1 score is a better metric than either precision or recall alone. However, each metric has its strengths and weaknesses. For example, precision is more important than recall in applications where false positives are more costly

than false negatives. Conversely, recall is more important than precision in applications where false negatives are more costly than false. The true positive, true negative, false positive, and false negative (FN) characteristics are used to establish the precision, recall, and F1 score. When the model properly detects the positive class, the result is TP.TN is the result when the

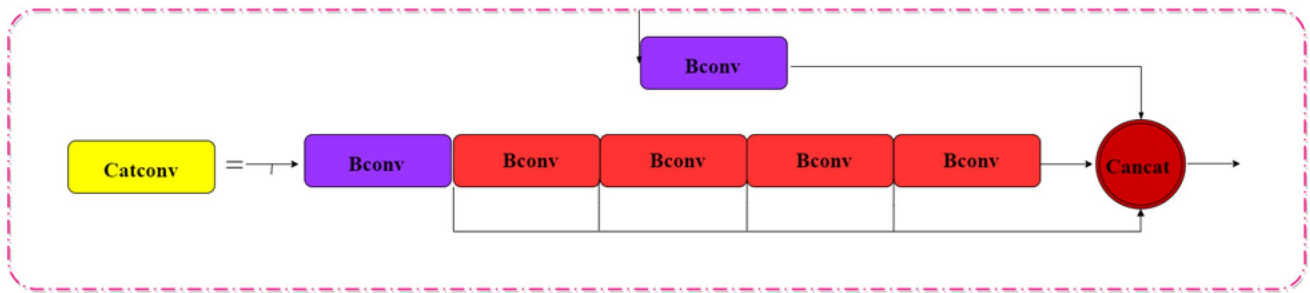


Fig. 12 Catconv layer structure

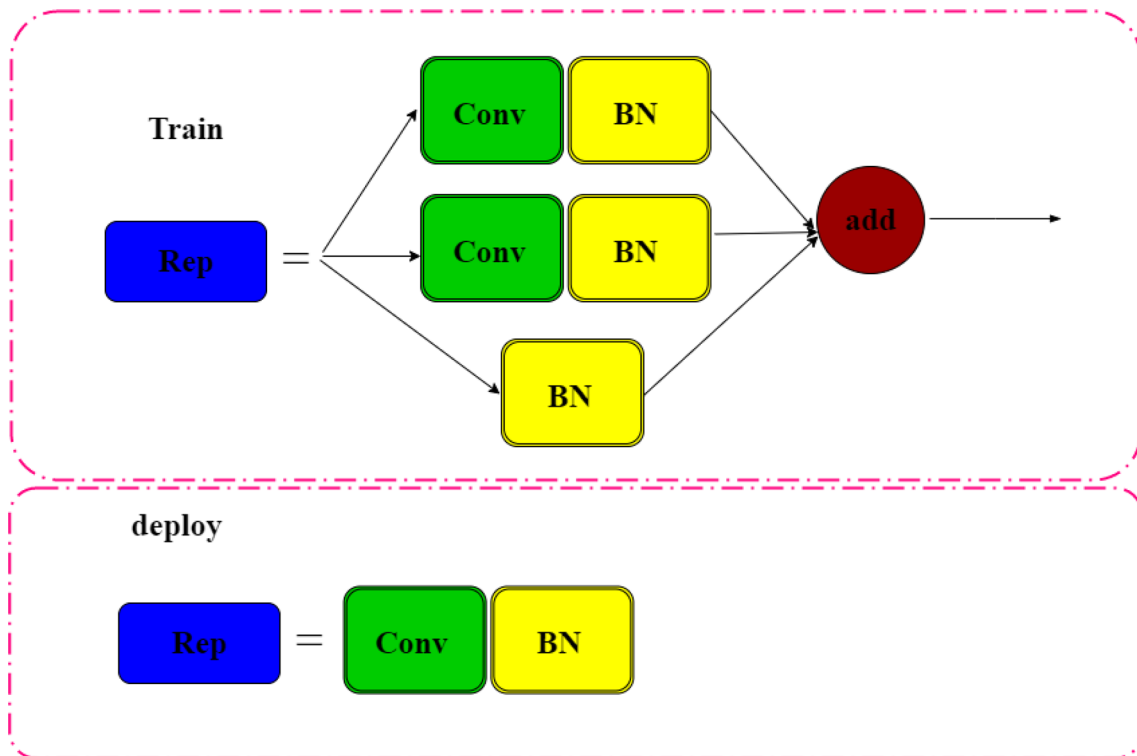


Fig. 13 Rep layer structure

model properly detects the negative class. The model returns FP when uncracked concrete is detected as cracked. The result is FN when the model detects a crack as uncracked concrete. As indicated in Eq. 1, precision is the percentage of correctly identified cracks to total detected cracks.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

The recall is the percentage of properly identified cracks about the total number of cracks as observed in Eqs. 2 and 3 defines the dice coefficient.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{3}$$

Fig. 14 Examples of images with labels and bounding boxes

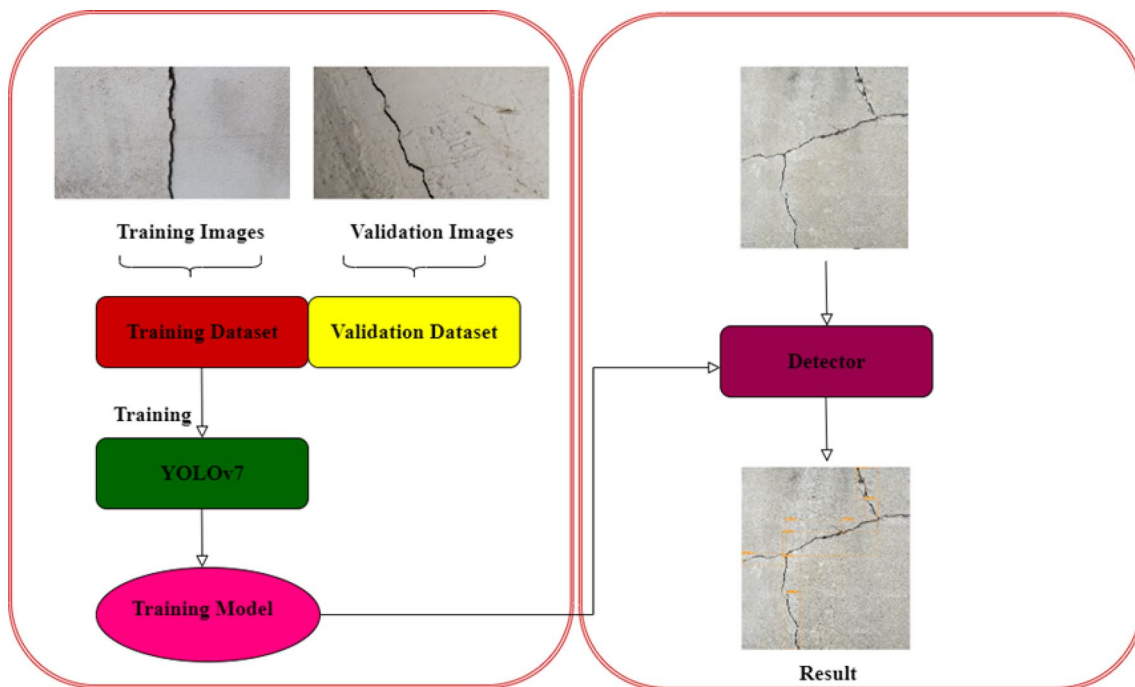


Fig. 15 Process of the research design

Defect detection accuracy is measured using precision and recall. The average precision is defined by the area under a precision-recall curve and the (average precision) AP values take accuracy and recall into account and are given as

the area under the P-R curve and mAP is equal to the AP value for each class.

$$AP = \int_0^1 P(R)dR$$

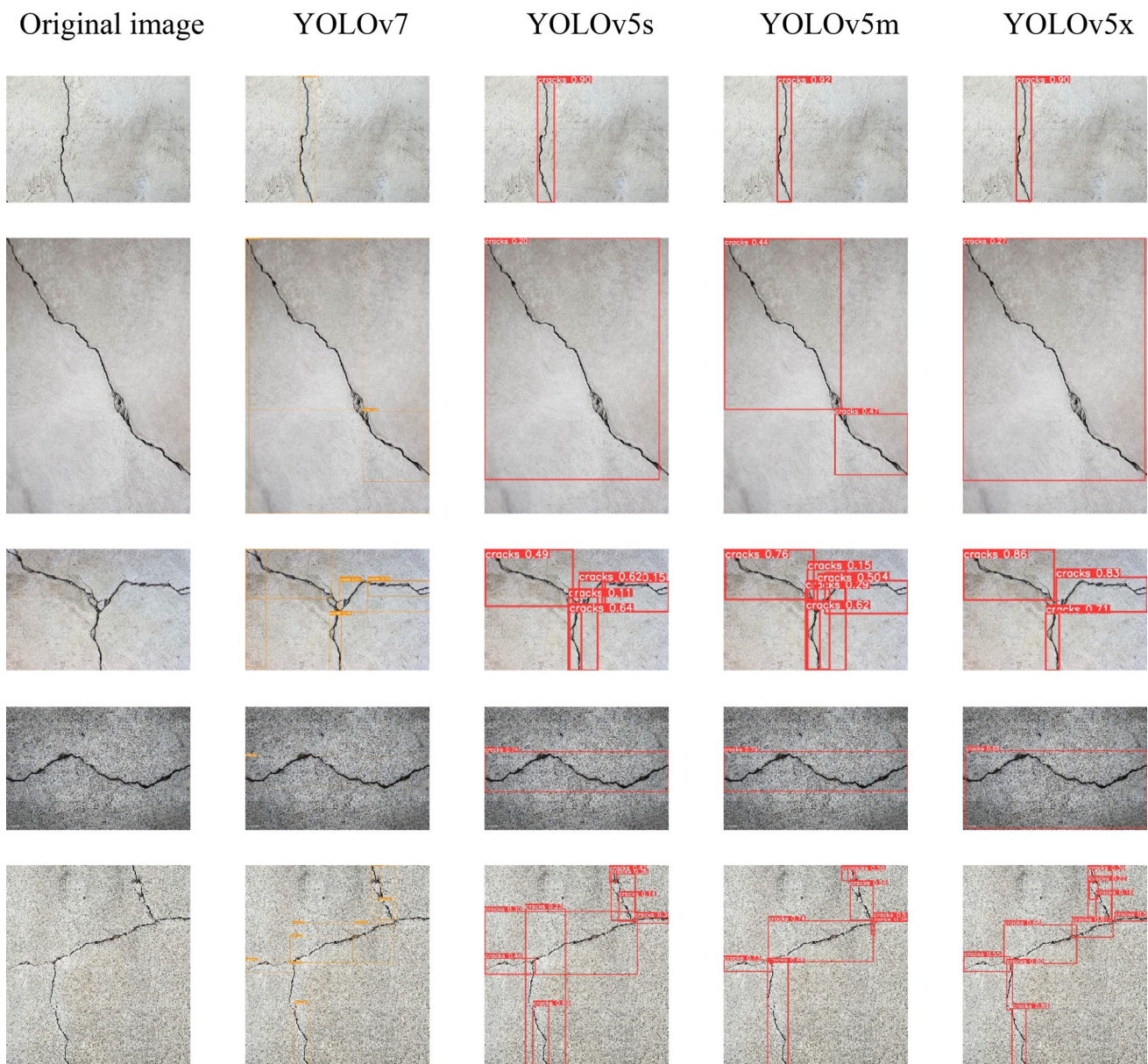


Fig. 16 Examples of concrete crack detection for trained models

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i.$$

Dataset and initial evaluation

A collection of 1600 inspection images of concrete surfaces from the robflow website was used for this study Hyegeun (2022). All of the training images contained cracks and concrete backgrounds. To evaluate the performance of the network, data were divided into training and validation data.

Annotated images with cracks are shown in Fig. 14. For crack detection, each image’s label was kept in a text file with four coordinates that described a rectangle box.

Results and discussion

This study compares the performance of the YOLOv7 model with the other state-of-the-art object detection model YOLOv5. The YOLO models are trained to detect cracks in photos of concrete cracked surfaces. A computer with the following configuration was used for these tests: The

Fig. 17 Video test result

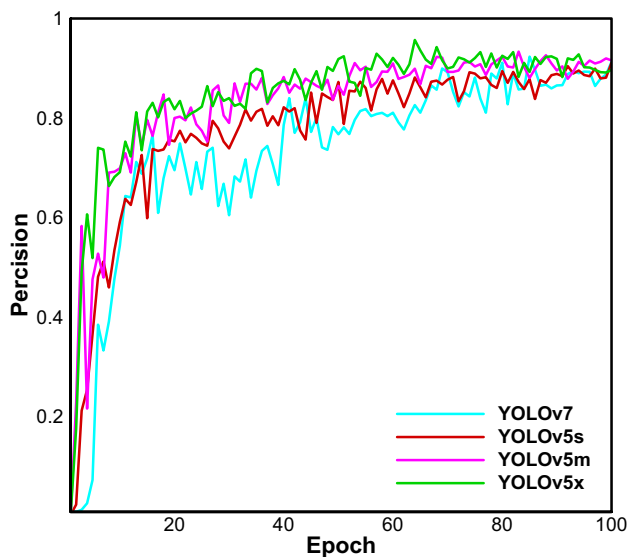


Fig. 18 Precision curve

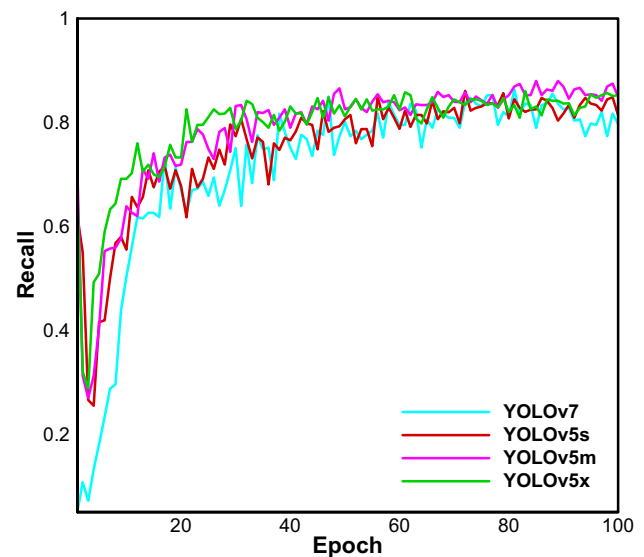


Fig. 19 Recall curve

technique was built using the PyTorch library, two Intel Xeon@2.3 GHz CPU, 64 GB access memory; and a Tesla k80 12 GB GPU. A total of 100 photos of cracked surfaces are used to test the model's accuracy. To evaluate the performance of the YOLO network, three different weights of the YOLO version 5 network and YOLOv7 were assessed on their performance using a dataset of 1600 concrete surface images. Prediction outcomes of the models applied to new data were examined to confirm their generalizability for use in selecting machines in the future which is shown in Fig. 15.

The results are impressive, with the model detecting cracks with high accuracy. Figure 16 depicts some of the

suggested network's testing outcomes. The model can identify cracks of different sizes and locations. The video on the concrete cracks runs for 10 s in MP4 format using YOLOv7. The result of the video detection for concrete cracks is shown in Fig. 17. The proposed model can detect cracks in real-time, without human intervention. This makes the proposed approach ideal for portable devices such as smartphones and tablets.

The precision curves of several models are shown in Fig. 18. The comparison curve shows that during the first training step, all four models increase rapidly as training increases. The YOLOv7 model's curve increases gradually with visible fluctuation, and the amplitude variation is

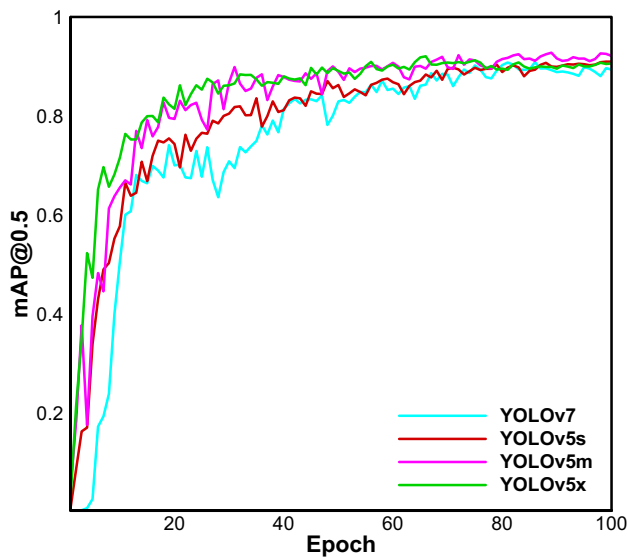


Fig. 20 Map@0.5 curve

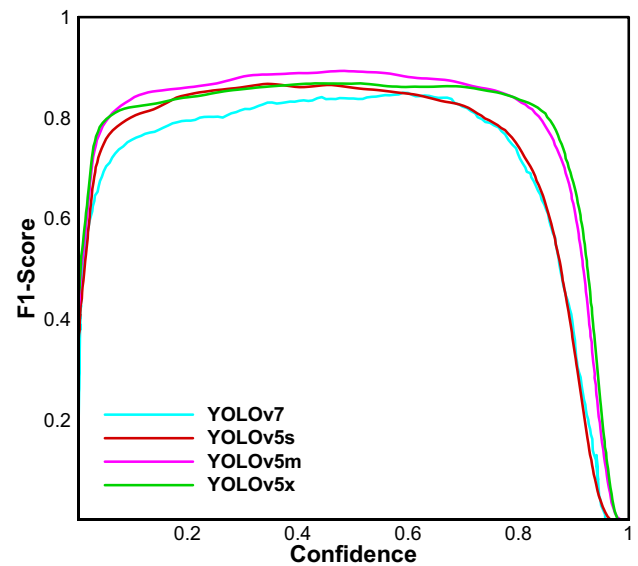


Fig. 22 F1-score curve

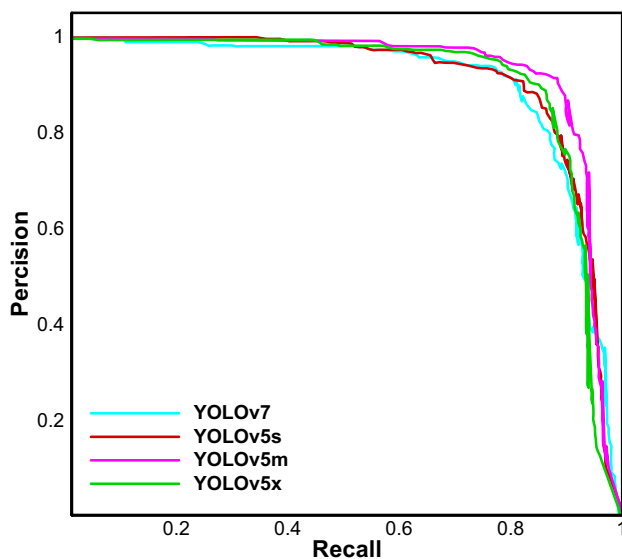


Fig. 21 Precision–recall curve

noticeable. The precision of YOLOv5m and YOLOv5x with extensive parameters rise the fastest.

As shown in Fig. 19, the recall curves of YOLO models are trained for 100 epochs. During the first 20 epochs, accuracy increases dramatically with epoch number. The YOLOv5m and YOLOv5x models have the highest recall. The YOLOv5m and YOLOv5x models had the highest recall. However, all the models performed well overall, and any of them could be used depending on the user's specific needs.

Table 1 Identification results for concrete crack detector

Model	mAP @0.5	mAP @0.5:0.95	Precision	Recall	F1-score
YOLOv7	0.89	0.69	0.91	0.79	0.84
YOLOv5s	0.90	0.67	0.91	0.81	0.85
YOLOv5m	0.92	0.73	0.91	0.85	0.87
YOLOv5x	0.90	0.74	0.89	0.85	0.86

Figure 20 shows the accuracy results of the YOLOv5 models at (a) mAP@0.5. The accuracy increases rapidly with the epoch. The mAP value of the YOLOv5m model is greater than the other models, as shown in Fig. 20.

Figure 21 depicts the trained model's precision–recall curve. The PR curves show how accuracy and recall are related. mAP@50 was calculated using the area under the curve. Therefore, it is better if the PR curve is close to the upper right corner. It is found that YOLOv5m is located closer to the top right corner, indicating a high accuracy and recall.

Figure 22 depicts the F1-score of the YOLO models. In object detection, the confidence threshold specifies the likelihood that an estimated bounding box includes an object. The F1 score curve detection findings show that the F1 score values of the YOLOv5m and YOLOv5x are greater than those of the other models.

Table 1 displays the proposed model's results for concrete crack detection. The greatest mAP@0.5 and mAP @0.5:0.95 values are found in the YOLOv5m model, suggesting the maximum accuracy.

The results showed that YOLOv5m and YOLOv5x had F1-scores of 0.87 and 0.86, respectively.

The precision of the YOLOv7, YOLOv5s, and YOLOv5m is 0.91, and the accuracy of the YOLOv5x models is 0.89. The recall of the YOLOv7 and YOLOv5s is 0.79 and 0.81, respectively, and in the YOLOv5m and YOLOv5x models, recall is the highest. Results show that YOLOv5 outperforms YOLOv7 models in terms of both speed and accuracy, making it a promising solution for real-time object detection applications.

Conclusions

Concrete is a widely used material in the infrastructure system. Thus, detecting concrete defects is crucial for reducing maintenance costs. Despite this, elements such as shrinkage and temperature changes, and structural cracking quickly affected by concrete's inherent properties. Cracks are often the most intuitive and reliable indicator when assessing the structural performance. This study proposes an object detection approach for detecting numerous concrete defects faster and more straightforwardly. Four versions based on YOLO object detection were trained and tested to detect cracks. For training, variety of concrete images labeled with concrete cracks are employed. It was found that the YOLOv5m and YOLOv5x models outperformed other models in terms of F-1 score and mAP@50. These models offer more accurate and reliable detection of objects in images and videos, making them the perfect choice for security and surveillance applications.

Acknowledgements The authors would like to acknowledge the electronic access to the web-based information service through the library of SVNIT, Surat without the library support, this article would have been much difficult to refer the literature. Moreover, the authors would like to thank the unknown referees for their useful remarks and critical suggestions in order to improve the quality of the article.

Author contributions EN: contributing significantly to the conception of article, study design, execution, acquisition of data, analysis, and interpretation. Writing an original draft, have written and substantially revised the article. AP: made a significant contribution to conception, investigation, methodology, study design, execution, acquisition of data, analysis, and interpretation. Writing original draft, reviewing, and editing an original draft, preparation, supervision, and formal analysis, and finalizing the article.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Andrushia, A., Diana, N. A., Lubloy, E., & Prince Arulraj, G. (2021). Deep learning based thermal crack detection on structural concrete exposed to elevated temperature. *Advances in Structural Engineering*, 24(9), 1896–1909. <https://doi.org/10.1177/1369433220986637>.
- Arbaoui, A., Ouahabi, A., Jacques, S., & Hamiane, M. (2021). Concrete cracks detection and monitoring using deep learning-based multiresolution analysis. *Electronics*, 10(15), 1772. <https://doi.org/10.3390/electronics10151772>.
- Bang, H., Min, J., & Jeon, H. (2021). Deep learning-based concrete surface damage monitoring method using structured lights and depth camera. *Sensors*, 21(8), 2759. <https://doi.org/10.3390/S21082759>.
- Bhowmick, S., Nagarajaiah, S., & Veeraraghavan, A. (2020). Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos. *Sensors*, 20(21), 6299. <https://doi.org/10.3390/S20216299>.
- Chow, J. K., Su, Z., Wu, J., Tan, P. S., Mao, X., & Wang, Y. H. (2020). Anomaly detection of defects on concrete structures with the convolutional autoencoder. *Advanced Engineering Informatics*. <https://doi.org/10.1016/j.aei.2020.101105>.
- Cui, X., Wang, Q., Dai, J., Zhang, R., & Li, S. (2021). Intelligent recognition of erosion damage to concrete based on improved YOLO-V3. *Materials Letter*, 302(November), 130363. <https://doi.org/10.1016/J.MATLET.2021.130363>.
- Deng, J., Lu, Y., & Lee, V.C.S. (2020). Imaging-based crack detection on concrete surfaces using you only look once network. 20(2), 484–499. <https://doi.org/10.1177/1475921720938486>.
- Du, Y., Pan, N., Xu, Z., Deng, F., Shen, Y., & Kang, H. (2021). Pavement distress detection and classification based on YOLO network. *International Journal of Pavement Engineering*, 22(13), 1659–1672. <https://doi.org/10.1080/10298436.2020.1714047>.
- Han, X., Zhao, Z., Chen, L., Hu, X., Tian, Y., Zhai, C., Wang, L., & Huang, X. (2022). Structural damage-causing concrete cracking detection based on a deep-learning method. *Construction and Building Materials*, 337(June), 127562. <https://doi.org/10.1016/J.CONBUILDMAT.2022.127562>.
- Hyegeun (2022). Open source dataset, *Roboflow universe*, 4, <https://universe.roboflow.com/hyegeun/-rxh4q>.
- Joshi, D., Singh, T. P., & Sharma, G. (2022). Automatic surface crack detection using segmentation-based deep-learning approach. *Engineering Fracture Mechanics*, 268(June), 108467. <https://doi.org/10.1016/j.engfracmech.2022.108467>.
- Kang, D., Benipal, S. S., Gopal, D. L., & Cha, Y. J. (2020). Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2020.103291>.
- Kang, D. H., Benipal, S., & Cha, Y. J. (2022). Hybrid concrete crack segmentation and quantification across complex backgrounds without a large training dataset. *Conference Proceedings of the Society for Experimental Mechanics Series*. https://doi.org/10.1007/978-3-030-76004-5_14/cover.
- Kaveh, A. (2016). Applications of metaheuristic optimization algorithms in civil engineering. *Applications of Metaheuristic Optimization Algorithms in Civil Engineering*. <https://doi.org/10.1007/978-3-319-48012-1>.
- Kaveh, A., & Dadras, A. (2018). Structural damage identification using an enhanced thermal exchange optimization algorithm. *Engineering Optimization*, 50(3), 430–451. <https://doi.org/10.1080/0305215X.2017.1318872>.
- Kaveh, A., & Maniat, M. (2015). Damage detection based on mcss and pso using modal data. *Smart Structures and Systems*, 15(5), 1253–1270. <https://doi.org/10.12989/sss.2015.15.5.1253>.

- Kaveh, A., & Mahdavi, V. R. (2016). Damage identification of truss structures using CBO and ECBO algorithms. *Asian Journal of Civil Engineering*, 17(1), 75–89.
- Kaveh, A., & Zolghadr, A. (2015). An improved css for damage detection of truss structures using changes in natural frequencies and mode shapes. *Advances in Engineering Software*, 80(C), 93–100. <https://doi.org/10.1016/j.advengsoft.2014.09.010>
- Kaveh, A., & Zolghadr, A. (2017). Cyclical parthenogenesis algorithm for guided modal strain energy based structural damage detection. *Applied Soft Computing Journal*, 57, 250–264. <https://doi.org/10.1016/j.asoc.2017.04.010>
- Kim, B., & Cho, S. (2018). Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. *Sensors (switzerland)*. <https://doi.org/10.3390/s18103452>
- Kumar, P., Batchu, S., Narasimha, S. S., & Kota, S. R. (2021). Real-time concrete damage detection using deep learning for high rise structures. *IEEE Access*, 9, 112312–112331. <https://doi.org/10.1109/ACCESS.2021.3102647>
- Li, X., Sun, H., Song, T., Zhang, T., & Meng, Q. (2022a). A method of underwater bridge structure damage detection method based on a lightweight deep convolutional network. *IET Image Processing*, 16(14), 3893–3909. <https://doi.org/10.1049/ipr2.12602>
- Li, Y., Yu, M., Wu, D., Li, R., Xu, K., & Cheng, L. (2022b). Automatic pixel-level detection method for concrete crack with channel-spatial attention convolution neural network. *Structural Health Monitoring*. <https://doi.org/10.1177/14759217221109496>
- Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2019). Computer vision-based concrete crack detection using u-net fully convolutional networks. *Automation in Construction*, 104(August), 129–139. <https://doi.org/10.1016/j.autcon.2019.04.005>
- Liu, Y., Yeoh, J. K. W., & Chua, D. K. H. (2020). Deep learning-based enhancement of motion blurred uav concrete crack images. *Journal of Computing in Civil Engineering*, 34(5), 1–14. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000907](https://doi.org/10.1061/(asce)cp.1943-5487.0000907)
- Lu, Y., Qiu, Z., Liao, C., Zhou, Z., Li, T., & Wu, Z. (2022). A GIS partial discharge defect identification method based on YOLOv5. *Applied Sciences (switzerland)*. <https://doi.org/10.3390/app12168360>
- Mishra, M., Jain, V., Singh, S. K., & Maity, D. (2022). Two-stage method based on the you only look once framework and image segmentation for crack detection in concrete structures. *Architecture, Structures and Construction*. <https://doi.org/10.1007/s44150-022-00060-x>
- Park, S. E., Eem, S. H., & Jeon, H. (2020). Concrete crack detection and quantification using deep learning and structured light. *Construction and Building Materials*, 252(August), 119096. <https://doi.org/10.1016/j.conbuildmat.2020.119096>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition 2016-December* (pp. 779–788). <https://doi.org/10.1109/cvpr.2016.91>.
- Ren, Y., Huang, J., Hong, Z., Lu, W., Yin, J., Zou, L., & Shen, X. (2020). Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Construction and Building Materials*, 234, 117367. <https://doi.org/10.1016/j.conbuildmat.2019.117367>
- Sun, Y., Yang, Y., Yao, G., Wei, F., & Wong, M. (2021). Autonomous crack and bughole detection for concrete surface image based on deep learning. *IEEE Access*, 9, 85709–85720. <https://doi.org/10.1109/ACCESS.2021.3088292>
- Teng, S., Liu, Z., Chen, G., & Cheng, L. (2021). Concrete crack detection based on well-known feature extractor model and the YOLO_v2 network. *Applied Sciences*, 11(2), 813. <https://doi.org/10.3390/APP11020813>
- Wan, C., Xiong, X., Wen, Bo., Gao, S., Fang, Da., Yang, C., & Songtao, X. (2022). Crack detection for concrete bridges with imaged based deep learning. *Science Progress*, 105(4), 003685042211284. https://doi.org/10.1177/00368504221128487/asset/images/large/10.1177_00368504221128487-fig2.jpeg
- Yang, C., Chen, J., Li, Z., & Huang, Y. (2021). Structural crack detection and recognition based on deep learning. *Applied Sciences (switzerland)*. <https://doi.org/10.3390/app11062868>
- Yang, Z., Ni, C., Li, L., Luo, W., & Qin, Y. (2022). Three-stage pavement crack localization and segmentation algorithm based on digital image processing and deep learning techniques. *Sensors*, 22(21), 8459. <https://doi.org/10.3390/s22218459>
- Ye, W., Deng, S., Ren, J., Xu, X., Zhang, K., & Du, W. (2022). Deep learning-based fast detection of apparent concrete crack in slab tracks with dilated convolution. *Construction and Building Materials*, 329(April), 127157. <https://doi.org/10.1016/J.CONBUILD-MAT.2022.127157>
- Yu, Z. (2022). YOLO V5s-based deep learning approach for concrete cracks detection. *SHS Web of Conferences*, 144, 03015. <https://doi.org/10.1051/SHSCONF/202214403015>
- Zhao, S., Kang, F., & Li, J. (2022). Concrete dam damage detection and localisation based on yolov5s-hsc and photogrammetric 3D reconstruction. *Automation in Construction*, 143, 104555.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.