Supplementary Material Table 1

#	Paper	Year	Algorithm	Results	Additional Information
1	(Ashraf et al., 2023)	2023	Custom YOLOv7	The YOLOv7 model has a notable accuracy of 92 percent on the RDD2022 dataset and 88 percent on their own dataset. Precision and recall for Custom YOLOv7 in RDD2022 were 0.9523 and 0.9545, respectively, while the corresponding values for the custom dataset were 0.93 and 0.9158.	Extended Efficient Layer Aggregation, Model Scaling Techniques, Re-parameterization Planning, and Auxiliary Head Coarse-to-Fine are the major reasons why YOLOv7 was chosen for this study. Sample size Customized yolov7's parameter value is 20.
2	(Yang et al., 2023)	2023	To get over the inaccurate boundary position between crack and non-crack pixels, an end-to-end deep convolutional neural network (AttentionCrack) is presented.	F1 have an average score of 0.71 on the Crack500 and over 0.70 on the CrackSegNet.	An attention mechanism is added into the multi-scale convolutional feature of the AttentionCrack network, which is based on U-Net encoder-decoder architecture, to improve the detection of crack regions. The encoder decoder architecture also includes a dilated convolution module to lessen the loss of crack detail caused by the pooling operation in the encoder network.
3	(Kim et al., 2023b)	2023	Rsef based on U-net namely Rsef-Edge	Top 5 avg. Acc. 97.36%	In this research, they propose an edge-accelerated crack detection algorithm.
4	(Lee et al., 2023)	2023	Meta model using stacking ensemble learning	IoU=0.74	For 64.4 percent of the images, the stacking ensemble model displayed an IoU of 0.5 or above. By training UNet, DeepLabV3, DeepLabV3+, DANet, and FCN-8s, they examined the crack-detection performance. As a result of the FCN- 8s' inadequate crack segmentation performance, stacking ensemble learning was carried out using the remaining four models. For the test dataset, each model produced an intersection over union (IoU) score that varied between around 0.4 and 0.6.
5	(Zhang et al., 2023b)	2023	Yolo v4	The suggested crack detection method's	The model could produce 140.2 frames per second and only used

Supplementary Material Table 1

				precision, recall, and F1 score are 93.96 percent, 90.12 percent, and 92 percent, respectively.	23.4 MB of storage space. The suggested method demonstrated advantages in terms of accuracy, speed, and model size over existing bridge fracture detection techniques.
6	(Zhao et al., 2023b)	2023	CrackNet model which is an encoder–decoder architecture.	On three datasets, namely UAVRoad-Crack, CRKWH100, and CrackLS315, the average accuracy (AP) scores of the CrackNet network were 0.665, 0.942, and 0.895, respectively.	
7	(Maslan & Cicmanec, 2023)	2023	Yolo v2	Average precision (AP) = 0.89	
8	(Lv et al., 2023)	2023	Mask region-based convolutional neural network (Mask R-CNN)	After testing and evaluating the created model using various datasets, the best accuracy is 99 percent, while the lowest accuracy is 95 percent.	
9	(Wang, 2023)	2023	CrackSN an automatic crack detecting method built on deep learning, was introduced. The Adam-SqueezeNet architecture served as the foundation for this suggested deep learning model.	The improved CrackSN system properly classifies 97.3 percent of the cracked patches in the image dataset, outperforming state-of-the-art models in recent re- search.	
10	(Gooda et al.)	2023	There are two methods that are suggested for segmenting and identifying objects. For segmentation, the YOLO v5 algorithm is recommended, while the EfficientNet with residual U- Net approach is put forth for crack detection.	99.35% accuracy	
11	(Kapadia et al., 2023)	2023	Inceptionv3 model	While 97.67% accuracy and 7.69% cross-entropy are averaged during model validation, 97.49% accuracy and 7.38% cross-entropy are averaged during training. Precision, recall, and F-score aver- ages come out to 0.88, 0.98, and 0.93 respectively.	With 5,000 epochs and a batch size of 100, the training was conducted. During the performance evaluation of fracture detection on concrete cubes, an average accuracy of 98% was attained.
12	(Ngo et al., 2023)	2023	Deep learning	Accuracy: 95.19%.	
13	(Chu et al., 2023)	2023	CNN with a K-fold cross validation technique with a K value of 10.	98 percent precision, 97 percent recall, and accuracy were tested on unseen images.	

Supplementary Material Table 1

14	(Bai et al., 2023)	2023	ResNet and ResNet+UNet	Accuracy: 67.6%	The first step is to classify several classes in eight SDD activities, including scene levels, damage levels, and material types, using a 152-layer residual network (ResNet). Even though it is im- possible to pinpoint the locations of the dam- age, the proposed ResNet achieved good accuracy for each assignment. Second, a new pipeline called cascaded networks is created by combining the already-existing ResNet and a segmentation network (U-Net) for identifying and locating structural damage. According to the findings, deploying a segmentation network alone would not have greatly increased the accuracy of damage detection. In order to directly identify cracks and spalling in image collections of recent big earth- quakes, end-to-end networks are designed and tested as a new approach.
15	(Kolappan Geetha et al., 2023)	2023	Deep learning. As a post-processor to DL tracks missing tiny cracks on segments identified as cracks, iterative differential sliding- window-based local image processing is used.	Time for testing/inference: roughly 0.02 seconds per image (60 images/s) (excluding image pre- and post-processing) Ap- proximately 0.1–0.2 seconds per image (5–10 images/s), including DL testing. 120–180 s/image on average (includes DL testing, picture preparation, and iterative postprocessing)	The suggested method uses image processing as a 1D DL model's pre- and post-processor. The model detects possible crack candidate sites. Image-processing-assisted DL as a prelude to DL eliminates labor-intensive labeling and the flat structural background without any identifiable features during DL training and testing.
16	(Panta et al., 2023)	2023	Novel encoder-decoder-based fully convolutional neural network to detect cracks from levee images at a pixel level automatically. They propose that the feature learning be strengthened using the decoder and bottleneck feature maps by concatenating them back to the encoder blocks. The addition	Achieving an increment of Intersection over Union (IoU) by 10.32% on aver- age for a 10-Fold CrossValidation (FCV) compared to the baseline U-Net model. In addition, IterLUNet has at least 63% fewer parameters to be trained than the baseline model.	

Supplementary Material Table 1

			reinforcement in the U-Net-like architecture results in a loop-like structure to exploit all the feature maps from encoders, bottlenecks, and decoders. The proposed architecture is, Iterative Loop UNet (IterLUNet).		
17	(Philip et al., 2023)	2023	VGG16, VGG19, ResNet50, MobileNet, and Xception	Recognized accuracy values for the ResNet50-based classifier were 99.91% for training and 99.88% for testing. The least performant architecture was Xception, with training and test accuracy of 99.64% and 98.82%, respectively.	
18	(Qayyum et al., 2023)	2023	GoogLeNet, MobileNet-V2, Inception V3, ResNet18, ResNet50, ResNet101, and ShuffleNet	Images are divided into four categories: uncracked (UC), horizontal (HC), diagonal (DC), and vertical (VC). With classification accuracy rates of 96%, 94%, 92%, and 96% for DC, HC, UC, and VC, respectively, Inception-V3 beats all other models. The training time for ResNet18 is 32 minutes, whereas ResNet101's is 171 minutes.	32,000 photos, evenly distributed among each class, are used to train each architecture. 100 photos from each category are used to test the trained models, and the outcomes are compared.
19	(Ina´cio et al., 2023)	2023	Multi-class CNN	This method, which is automated and has a short processing time, produced promising findings in the examination of high- way pavements.	
20	(Deng et al., 2023)	2023	After using YOLOv5 for crack detection and Res- UNet for precise segmentation, a new crack surface feature quantification algorithm is created to calculate the width and length of the crack in pixels. This method evaluated against DeepLabv3+ and You Only Look at CoefficienTs ++ (YOLACT++). The modified Res-UNet achieves 87 percent intersection over union (IoU) when segmenting crack pixels, 6.7 per- cent higher than the original Res-UNet.	The YOLOv5 achieves a mean average precision of 91 percent; the developed crack surface feature algorithm achieves 95 percent accuracy in identifying the crack length and a root mean square error of 2.1 pixels in identifying the crack width, with the accuracy being 3 percent higher in length measurement than that of the traditional method.	
21	(Tan & Dong, 2023)	2023	Pyramidal residual network based on encoder decoder using Omni-Dimensional Dynamic Convolution.	99.05 percent accuracy and a mIoU of 87.0 percent.	According to the experimental findings, SegNet, DeeplabV3+, and Swin-unet are inferior to

Supplementary Material Table 1

					the concrete crack segmentation approach.
22	(Paramanandham et al., 2023)	2023	Using a novel technique called the pixel-intensity resemblance measurement (PIRM) rule, noisy and noise-less images are classified. After that, a median filter was used to re- move the noise. Using the VGG-16, ResNet-50, and InceptionResNet-V2 models, the cracks were found.	For the VGG-16 model, the suggested method produced improvements of 6% without PIRM and 10% with PIRM. Similar results were obtained with ResNet-50 (3 and 10%), Inception ResNet (2 and 3%) and the Xception model (9 and 10%). The ResNet-50 model for Gaussian noise, the Inception ResNet v2 model for Pois- son noise, and the Xception model for speckle noise each achieved accuracy of 95.6 percent when the images were affected by a single noise source alone.	
23	(Yu & Zhou, 2023)	2023	For crack detection, a network named YOLOv5- CBoT—an upgraded YOLOv5 network coupled with a Bottleneck Transformer—is proposed.	According to the experimental findings, our system's F1 score is up 1.3%, its mAP0.5 is up 2.1%, its mAP0.5:0.95 is up 3.1%, and its inference speed is 1.4 times faster than the original YOLOv5 method.	
24	(Zhang et al., 2023a)	2023	CTCD-Net: A Cross-layer Transmission network for tiny road Crack Detection.	86.59% precision on DeepCrack537 dataset.	Accurately detecting road cracks has two main drawbacks: (1) Small cracks are often neglected because they lack distinguishing characteristics and are more susceptible to sounds; (2) the majority of current extraction techniques extract cracks with coarse and thicker edges, indicating room for development. To make up for the drawback of the unnoticeable features of tiny fractures, they first suggest a cross-layer information transmission module based on an attention mechanism. This module emphasizes the feature representation of small fracture locations by layer-by-layer transmitting the feature information from upper layers to

Supplementary Material Table 1

					the next one in order to enrich the information. To further increase the accuracy of crack boundary placements, they construct a boundary refinement block. This block refines borders by learning the residuals between the label pictures and the intermediate coarse maps.
25	(de Leo´n et al., 2023)	2023	An innovative crack segmentation algorithm based on the notion of minimal path selection and a region-based strategy was developed by segmenting texture information that had been retrieved using Gabor filters.	F1 score is 0.839.	The proposed method does not require any prior knowledge, however it is possible that the statistical parameters will need to be changed depending on the specific scenario and case.
26	(Inam et al., 2023)	2023	YOLOv5 for crack detection and U-Net for segmentation.	The test set for the YOLOv5 s, m, and l models had mean average precision (mAP) values of 97.8%, 99.3%, and 99.1%, respectively. This indicates that the YOLOv5 m model performed better than its two counterparts.	
27	(Liu et al., 2023)	2023	Utilizing image processing technologies, a novel crack identification and feature recognition method is presented. While a deep convolutional network (Single Shot MultiBox Detector (SSD)) is proposed for object detection in complicated images based on deep learning, this method completely takes into account the peculiarities of tunnel images and the coupling of these qualities with deep learning.	In the classification comparison test, the test set accuracy and training set accuracy for the support vector machine (SVM) are up to 88 and 87.8 percent, respectively. In contrast, the test set accuracy and training set accuracy for Alexnet's deep convolutional neural network-based classification and identification are up to 96.7 and 97.5 percent, respectively.	
28	(Guo et al., 2023)	2023	VGG16 + Focal Loss.	F1 score is 0.613.	By dividing the pixel count by the pixel size, it is possible to determine the precise crack dimension. When compared to the conventional method, the proposed method may estimate flaws dispersed throughout a complex structure. On a chosen 1500 mm 1500 mm concrete road stretch, a pilot case study was carried out on a concrete pathway with fractures spread across it.

Supplementary Material Table 1

					Overall, 10 out of 88 photos are chosen for validation; average errors for small cracks ¡ 5 mm ranged from 0.26 mm to 0.71 mm, showing a promising outcome of the intended study.
29	(Li et al., 2023b)	2023	The YOLOv7 with attention mechanism	The results demonstrate that the enhanced model using the SimAM attention mechanism can achieve 100% precision, 75% recall rate, 96.89% AP, and a processing time of 10 s for 100 images.	
30	(Kim et al., 2023a)	2023	AlexNet, VGG-16 and ResNet152	Precision (%) and Recall (%) for AlexNet 87.74, 77.77 and for VGG-16 88.76, 87.14, and for ResNet152, 87.59, 85.47.	
31	(Tse et al., 2023)	2023	Improved YOLOv4 with an attention module	90.02% mean average precision (mAP) and performs 5.23 percent better than the YOLOv4-original.	Navigation trajectories do not impose any limitations on it.
32	(Kao et al., 2023)	2023	YOLOv4	Accuracy: 92%	Crack-containing images were converted to grayscale and then binary images using local thresholding for quantitative crack analysis. Crack edges were extracted using Canny and morphological edge detectors, creating two types of edge images. Crack edge size was determined using the planar marker and total station measurement methods.
33	(Lee & Yoo, 2023)	2023	A fast encoder-decoder network that pays care to scaling. By eliminating encoder-decoder pairs and implementing an Atrous Spatial Pyramid Pooling (ASPP) layer, they concentrate on a low-level feature map and increase the detection precision of minute cracks.	Only 1.2% separates the top score from the suggested model's mDice score, al- though there are twice as less FLOPs in the proposed model.	Reduced crackling noise is one of the difficulties. In order to suppress irrelevant regions, this proposes a novel scaling attention, AG+. How- ever, employing merely enhanced segmentation networks makes it difficult to remove cracklike noise, such as grooving.
34	(Zhao et al., 2023a)	2023	U-Net	According to the experimental findings, the average ODS, OIS, AIU, sODS, and sOIS are,	First, multi-scale characteristics are extracted from the crack image using a U-Net network. Second, the extracted crack

Supplementary Material Table 1

				respectively, 75.7%, 73.9%, 36.4%, 52.4%, and 52.2%.	features are further morphologically processed by a white-top hat trans- form and a black bottom hat transform to remove the impact of polarized light on the cracks under various illuminations.
35	(Shim et al., 2023)	2023	Deep Learning with an adversarial learning-based balanced ensemble discriminator network.	An F1 score of 82.91 percent and a mean intersection-over-union of 84.53 percent.	
36	(Li et al., 2023a)	2023	Two stage transfer learning. Segmentation using the UNet model ResNet50. Additionally, multilayer parallel residual attention (MPR) is used to sharpen the focus on important information for more accurate fracture edge segmentation.	mIoU of 88.3 percent and mPA of 92.7 percent, respectively. The suggested strategy enhances mIoU and mPA by 4.6	
37	(Popli et al., 2023)	2023	Deep learning and robots. The proposed approach combines deep learning techniques for crack detection with the ability of a robot vision system to collect high-quality data about the road.	Among the algorithms put to the test, Xception stands out as the most precise and prescient model, with a validation accuracy of over 90% and a mean square error of just 0.03.	DenseNet201, MobileNetV2, VGG16, and VGG19 models have been improved using the SDNET2018 dataset, which contains photos of concrete surfaces with various degrees of fractures.
38	(Xu et al., 2022)	2022	Fast RCNN, Mask RCNN, YOLO.	In comparison to Mask RCNN, Fast-RCNN produces better and more com- prehensive results. When compared to YOLO, Fast RCNN also performs better.	It is clear that Faster R-CNN and Mask R-CNN can outperform YOLOv3 and finish the crack detection task when trained with only 130+ images.
39	(Jayaraju et al., 2022)	2022	CNN	99 % accuracy	
40	(Zhang et al., 2022)	2022	FPN-vgg16	F1 score is 84.40% and the best IoU score is 73.11%	This study is the first of its type to use the concept of transfer learning, use a mixed-crack picture dataset for training three deep learning models, and propose a deep learning method to investigate the cracks on earthen historical sites at the pixel- level.
41	(Wang et al., 2022b)	2022	An effective mobile-attentionX-network (MA-Xnet) for crack detection is proposed, which is based on the dual-attention network (DANet) and U-Net, two semantic segmentation models.	The important metrics of the F1-Score and the mean intersection of union(mIoU) are, respectively, 90.53 per- cent and 81.32 percent.	Deep learning-based modern crack detection algorithms face challenges like a dearth of context information and an abundance of parameters in the

Supplementary Material Table 1

					generated models. subsequently improved pro- posed MA-Xnet to enable real-time road-crack picture segmentation on small mobile devices
42	(Li et al., 2022b)	2022	By adding the convolutional block attention module (CBAM) to re- duce background noise, the model increases the saliency of pavement dam- age and investigates the impact of the CBAM mod- ule's embedding position on the detection accuracy. In order to increase the target recognition accuracy and create the high- performance pavement crack detection model known as YOLOv4-3, the anchor box parameters were optimized using the Kmeans++ algorithm.	mAP = 82.95% The upgraded YOLOv4- 3 network's mAP (mean average precision), according to the results, was 2.96% higher than it was before the change.	
43	(Yang et al., 2022)	2022	Yolov5s	F1-score is 86.79%	
44	(Lee et al., 2022)	2022	CNN	Recall: 75% and, Precision: 71%	
45	(Mo et al., 2022)	2022	Fast R-CNN-based deep learning module and the gap hazard evaluation method (GHEM)	Accuracy is 89.06%	
46	(Yong & Wang, 2022)	2022	An end-to-end real-time pavement crack segmentation network (RIIAnet).	The outcomes demonstrate that the suggested model achieves the highest MIOU, MPA, Recall, and F1 scores, with values of 0.7705, 0.9868, 0.8047, and 0.8485, respectively.	More crucially, the suggested model's parameter size is drastically reduced and is just 0.04 times that of U-Net.
47	(Li et al., 2022c)	2022	Edge extraction and deep supervision are carried out by the SoUNet: U-Net and side-output portion that is added to the U-Net decoder.	According to the results, the mean intersection over union for their dataset is 69.32, while it is 61.05 for a different pavement dataset group that did not take part in training. When compared to earlier semantic segmentation models, the suggested strategy can raise the MIoU value by up to 5.55 and the MPA value by up to 10.41.	
48	(Islam et al., 2022)	2022	VGG16, ResNet18, DenseNet161, and AlexNet with pre trained (trained on ImageNet) weights.	With a testing accuracy of 99.90%, precision of 99.92%, recall of 99.80%, and F1-score of 99.86% for crack class, AlexNet beats existing models. The ac- curacy of the	

Supplementary Material Table 1

				models VGG16, ResNet18, DenseNet161, and AlexNet was 99.90%, 99.60%, 99.80%, and 99.90%, respectively, using BWCI.	
49	(Ha et al., 2022)	2022	SqueezeNet, U-Net, and Mobilenet-SSD models together.	The crack type and severity assessments both have a 91.2 percent accuracy.	Since SqueezeNet performs classification with an accuracy of roughly 99.6%, the remaining 0.4% of test images are incorrectly typed and passed to the incorrect U-Net for segmentation.
50	(Loverdos & Sarhosis, 2022)	2022	U-Net, DeepLabV3+, U-Net (SM), LinkNet (SM), and FPN (SM)	High accuracy was attained using deep learning techniques (the crack detection model had an F1-Score of 79.6% and the block detection model had a validation accuracy of 96.86%).	
51	(Ali et al., 2022)	2022	By combining the vision-transformer (ViT) classifier with the sliding-window and tubularity-flow-field (TuFF) algorithms, it was evaluated to improve the performance of crack classification, localization, and segmentation.	Scores for accuracy, precision, recall, and F1 were 0.960, 0.971, 0.950, and 0.960, respectively.	To create a crack-localization map from big pictures, the trained ViT was combined with the sliding-window (SW) method. The SW-based ViT classifier was then combined with the TuFF method, and in the final stage, the undesirable regions were suppressed to obtain an effective crack-mapping.
52	(Wibowo et al., 2022)	2022	Transfer learning based on feature extraction using VGG16 and RestNET50, paired with an ANN and kNN for the classifier, and evaluated the performance of the predictions.	Average AUCROC: Greyscale: VGG16-ANN: 0.92, ResNet50-ANN: 0.96 VGG16-KNN: 0.84 ResNet50 KNN: 0.92 RGB: VGG16-ANN: 0.94 ResNet50-ANN: 0.95 VGG16-KNN: 0.85 ResNet50-KNN: 0.89	The overall detection accuracy was able to increase by 3.7 percent for RGB and 4.8 percent for the greyscale dataset after switching from the VGG-ANN to ResNet50-ANN. Additionally, switching from an ANN to a KNN classifier showed a more notable impact of 9.2% and 10.7% for the RGB and greyscale datasets, respectively.
53	(Pu et al., 2022)	2022	Deep convolutional neural network (DCNN) and an encoder–decoder module	Accuracy is 91.62%	
54	(Munawar et al., 2022)	2022	To anticipate pixel-wise segmentation in an end-to- end manner, a modified	Global Accuracy (GA); Class Average Accuracy (CAC); mean	To improve the accuracy of the predictions, they have applied

Supplementary Material Table 1

			version of deep hierarchical CNN architecture based on 16 convolutional layers and cycle generative adversarial network (CycleGAN) is used.	Intersection of the Union (IOU); Precision (P); Recall (R); and F score values of 0.989, 0.931, 0.878, 0.849, 0.818 and 0.833, respectively.	guided filtering and Conditional Random Fields (CRFs) techniques.
55	(Ma et al., 2022b)	2022	YOLO v3, YOLO v4s-mish, and YOLO v5s.	The YOLO v3 model's loss function is 0.026, recall rate is 91.64 percent, and mAP value is 0.955, which are better than the YOLO v4s-mish and YOLO v5s models. However, the model's maximum weight is 118 MB, minimum FPS value is 71.43, and maximum inference time is 21 ms, indicating that while it performs well overall, training is slow.	
56	(Wan et al., 2022)	2022	Single shot multibox detector (SSD) and the eight- neighborhood algorithm.	Precision of at least 95% and recall of at least 75%.	The SSD algorithm's deep learning was applied to the training data in order to create the detection model. Sliding window technology was incorporated to detect test set flaws. The eight-neighborhood technique was also used to fix any remaining crack detection issues.
57	(Ren et al., 2022)	2022	YOLOV5.	The precision of YOLOV5-CoordAtt is 95.27 percent. It performed better than other traditional and deep learning techniques.	
58	(Kang & Cha, 2022)	2022	A novel semantic transformer representation network (STRNet). A focused Tversky loss function, a squeeze and excitation attention-based encoder, a multi head attention-based decoder, coarse upsampling, and a learnable swish activation function make up STRNet.	In terms of precision, recall, F1 score, and mIoU, it obtains 91.7%, 92.7%, 92.2%, and 92.6%, respectively.	The performance of STRNet is compared to that of recently created advanced networks (Attention U-net, CrackSegNet, DeeplabV3+, FPHBN, and Unet++), with STRNet achieving the fastest processing at 49.2 frames per second and displaying the best performance overall.
59	(Siriborvornratanakul, 2022)	2022	DeepLabV3-ResNet101	F-measure is 84.49% on CrackTree260, 80.29% on CRKWH100, 72.55% on CrackLS315, and 75.72% on Stone331.	
60	(Elghaish et al., 2022)	2022	AlexNet, VGG16, VGG19, GoogleNet	The GoogleNet model is the most accurate in this case, with an	

Supplementary Material Table 1

				accuracy of 89.08%, 1.26 percentage points greater than AlexNet's. While employing Adam's optimization approach, the newly generated deep learning CNN model's computed accuracy outperformed all previously trained models by obtaining 97.62% at a learning rate of 0.001.	
61	(Wu et al., 2022)	2022	FCN-8s, FCN-16s FCN-32s	Correct crack detection: 0.7585	
62	(Liu et al., 2022)	2022	Deep Domain Adaptation-based Crack Detection Net- work (DDACDN)	On CFD dataset: Accuracy: 96.8% On CQU-BPMDD: 82%	DDACDN, which only has access to image-level labels in the target domain, learns domain-invariant features by utilizing the source domain's knowledge to forecast the multi-category crack location information. In particular, DDACDN uses a two-branch weights-shared backbone network to collect crack characteristics from the source and target domains first.
63	(Nomura et al., 2022)	2022	YOLOv2 + VGG16	The recall of damage detection at the pixel level was discovered to be 0.7 to 0.9 by morphological processing.	Using images of a monorail's running surface captured by a vehicle-mounted camera, certain experiments are carried out to demonstrate the applicability of the technology established in this work.
64	(Yu et al., 2022)	2022	Modified DeepLabV3+	Comparing the suggested model to other methodologies, Acc, mAcc, MioU, and FWIoU increased by 0.1%, 1.2%, 2.9%, and 0.5%, and by 0.1%, 0.8%, 2.0%, and 0.4% when compared to the original DeepLabV3+.	
65	(Yu & Zhou, 2023)	2022	YOLOv5	The model achieved 92.0 percent precision, 97.5 percent recall, and 98.7 percent mAP@0.5 for the testing set during the detection stage.	More specifically, the ratio filter is used to remove speckle linear noises, while the mask filter is used to remove handwriting marks. They proposed a unique fusion method to integrate these bounding boxes when a single

Supplementary Material Table 1

					crack is detected by several bounding boxes.
66	(Munawar et al.)	2022	CNN and a cycle generative adversarial network (CycleGAN).	As shown by the global accuracy (0.990), class average accuracy (0.939), IoU (0.879), precision (0.838), recall (0.879), and F score (0.8581) values, the Guided Filtering (GF) approach performed better than all other methods.	The use of guided filtering (GF) and conditional random fields (CRFs) to improve the anticipated outputs and produce accurate findings is a crucial part of the proposed CNN design.
67	(Kun et al., 2022)	2022	Deep bridge crack classification (DBCC)-Net as a classification-based deep learning network.	The proposed method has 19 frames per second (FPS) and 0.79 Miou at the actual bridge photos of 25602560 pixels, according to experimental findings. The Miou value is 7.8% greater than other approaches despite the decreased FPS.	Using image slice classification, DBCC-Net first achieves the coarse extraction of crack position. The entire crack morphology is extracted in the second stage from the site that the semantic seg- mentation network has recommended.
68	(Mohammed et al., 2022)	2022	A modified U-Net, which has half the parameters of the original U-Net network to detect surface cracks.	The suggested semi-supervised learning strategy achieved relatively close accuracies to the established fully supervised models utilizing different accuracy in- dices, however the necessity for the labeled data drops to 40%, according to the results. The modified U-Net network improves accuracy by up to 20% while utilizing just 15% of the training time of the traditional U-Net network, according to a comparison using 20 epochs.	
69	(Hammouch et al., 2022)	2022	CNN and also, Transfer learning is also applied by testing a pre-trained VGG- 19 model.	The results revealed that the longitudinal crack F1-score results of 88.53% for CNN and 86.24% for VGG-19 are smaller than the F1-score results of alligator cracks, which are 93.45% for CNN and 89.34% for VGG-19.	
70	(Lee & Huh, 2022)	2022	Real-time RGB and IR images were obtained by in- stalling a multi-sensor system on the mobile mapping system (MMS), and related feature points were then located using the geometric constraint method in order to spatially register these images.	-	

Supplementary Material Table 1

71	(Lu et al., 2022)	2022	A novel multi-scale crack detection network, called MSCNet. To enhance the cracks' capacity to express depth information, they de- ploy Res2Net as the back- bone network.	Inference speed of more than 63 FPS, recall of 94.2 percent, and precision of 93.5 percent.	They apply a texture enhancement module based on group attention to capture the fine details of cracks in low-level features in order to fully utilize this visual characteristic because the edge attribute of bridge cracks is noticeable.
72	(Kim et al., 2022)	2022	Conv2D ResNet	By contrasting the results of Xception, VGG19, DenseNet, ResNet, and Conv2D ResNet Exponential, the model's performance was verified. The experimental results demonstrate that it is outperformed by the Conv2D ResNet model with an exponential activation layer, which has an F-score value of 0.9978 and may be a vi- able alternative for categorizing different wall faults.	
73	(Kou et al., 2022)	2022	Deep Learning	Defect Accuracy: 96.4% Cracks Accuracy: 87.83%	If the system is expanded to include the use of high-frequency cameras, it can attain rapid detection speeds. It is a cost-effective, time-saving, and ecologically responsible technology for detecting future rail surfaces.
74	(Zhao et al., 2022)	2022	Fast R-CNN	The outcome reveals a 5% improvement in mean average precision (mAP) over the conventional approach.	
75	(Jing et al., 2022)	2022	AR-UNet within the encoder and decoder of U-Net	Overall accuracy for the DeepCrack dataset is 87.2%, while precision and re- call are 88.9% and 85.7%, respectively.	To efficiently extract global and local detail information, AR-UNet adds a convolutional block attention module (CBAM) within the encoder and decoder of U-Net. The model's input and out- put CBAM features are linked to lengthen the transmission path of the features. In order to pre- vent network deterioration brought on by gradient disappearance and network layer growth, the BasicBlock is utilized to replace

Supplementary Material Table 1

					the convolutional layer of the original network.
76	(Gharehbaghi et al., 2022)	2022	FastCrackNet	Overall average F1: Nrm 0.95 SP 0.94 MB 0.95 SH 0.90	A crack-detection method with high computational efficiency is FastCrackNet. This method uses an efficient, fully connected network in place of the computationally expensive convolutional neural network (CNN), which is combined with a 2D wavelet image transform for analysis and a locality sensitive discriminant analysis (LSDA) for feature reduction.
77	(Yuan et al., 2022)	2022	An improved generative adversarial network	The experimental results show that the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) of the reconstructed images of the self-built pavement crack dataset achieve 29.21 dB and 0.854, respectively. Additionally, the im-pacts of image reconstruction on detection and segmentation are verified by using Faster RCNN and a Fully Convolutional Network (FCN). According to the results, when compared to state-of-the- art techniques, the segmentation results' F1 is boosted by 0.012 to 0.737 and the detection results' confidence is increased by 0.031 to 0.9102.	To address the enormous discrepancies in image quality caused by varying equipment and lighting circumstances in the image-collecting stage of intelligent pavement recognition, a super-resolution reconstruction strategy based on an upgraded generative adversarial network is described. First, the generator's nonlinear network is enhanced, and the Residual Dense Block (RDB), which will act as Batch Normalization (BN), is developed. The RDB, Gated Recurrent Unit (GRU), and Conv Layer are then brought together to create the Attention Module. Finally, the original loss function is replaced with one based on the L1 norm.
78	(Paramanandham et al., 2022)	2022	Alexnet, VGG16, VGG19 and ResNet-50.	In comparison to the VGG16 and VGG19 models, the Alexnet model performs faster (about 3 hours) but has the lowest test accuracy (98.03%). In comparison to all the other models used for comparison, ResNet-50 has a testing accuracy of 99.92% and trains samples in a significantly shorter amount of time (just 14 minutes). VGG16 takes	Both destructive and non-destructive testing (NDT) methods can be used to find the cracks. This article offers state-of-the-art deep learning- based image-based NDT methods for identifying concrete fractures. NDT is a procedure for examining materials, parts,

Supplementary Material Table 1

				around 14 hours to train and has the highest test accuracy of 99.98.	structures, etc. without endangering them.
79	(Quqa et al., 2022)	2022	CNN	AUC: 0.804	
80	(Ji et al., 2022)	2022	U-Net. They employed ResUNet, VGGU-Net, EfficientU-Net, and U-Net-based neural networks for efficient localization and crack detection.	In comparison to VGGU-Net (67.71%) and EfficientU-Net (68.07%), ResU-Net (68.47%) obtains the highest MIoU with a small number of parameters, according to the results of the integrated dataset. In addition to its performance, ResUNet had the fastest true positive rate in the pixel wise recognition test, at 45.0%, and the shortest test runtime, at 40 milliseconds for each individual image.	Mean pixel accuracy (MPA), MIoU, and confusion matrix were some of the assessment measures used to assess the models through five-fold cross-validation.
81	(Chen et al., 2022)	2022	Graph network branch	In the comparison experiments, the proposed method gets the highest F1 and IoU. In addition, the graph branch addition outperforms U-Net by 0.06 on IoU and 0.08 on F1.	In this method, first, the image splitting is used to create the graph's nodes and the nodes' attributions. Based on the Gaussian distribution, the graph's edges are chosen. The image's graph is then entered into the graph branch. The image feature map from the encoder and the graph feature map from the graph branch output are fused before entering the decoder to recover the image resolution and produce the crack segmentation result.
82	(Yadav et al., 2022)	2022	Multi-scale feature fusion (3SCNet+LBP+SLIC). 3SCNet (3ScaleNetwork), a revolutionary deep convolutional neural network, is used to identify cracks. The crack image's texture pattern is discovered using the LBP (Local Binary Pattern) segmentation technique and the SLIC (Simple Linear Iterative Clustering) segmentation technique. 3SCNet receives the SLIC, LBP, and grayscale pictures to create a feature vector pool.	The values for sensitivity, specificity, and accuracy were 99.47 percent, 99.75 per- cent, and 99.69 percent, respectively.	

Supplementary Material Table 1

83	(Golding et al., 2022)	2022	Deep learning-based autonomous crack detection method using CNN technique.	The results of the study showed that the grayscale models (F1 score for 10 epochs: 99.331%, 20 epochs: 99.549%) performed similarly to the RGB models (F1 score for 10 epochs: 99.432%, 20 epochs: 99.533%), with the rate of performance improvement increasing with training. The performance of the thresholding and edge-detection models was inferior to that of the RGB models (20-epoch F1 score versus RGB: thresholding -0.723%, edge detection -0.402%).	The three methods (grayscale, thresholding, and edge detection) were used because they have all been successful at detecting cracks in image processing (IP), but not in deep learning (DL). Based on this study, color may not be a factor in DL crack detection.
84	(Li et al., 2022a)	2022	A novel crack detection network called the dense boundary refinement network (DBR-Net), which combines the benefits of the refinement network and STDC-Net (short- term dense concatenate network).	The detection accuracy is 97.54 percent. Furthermore, the detection rate has in- creased to 37.0 images per second (IPS).	Through the elimination of superfluous structures and the optimization of the detailed information using binary cross-entropy loss and dice loss, STDC-Net primarily increases the detection rate. In order to forecast the segmentation outcomes, the detail aggregation module integrates the shallow geographical data with the deep semantic in- formation.
85	(Wang et al., 2022a)	2022	Pyramid Scene Parsing Network (PSPNet), Fully Convolutional Network (FCN), Global Convolutional Net- work (GCN), UPerNet, and DeepLabv3+. The VGG, ResNet, and DenseNet are used as the backbones.	The DeepLabv3+ with the ResNet101 backbone obtained the highest IoU of 0.6298, the highest recall of 0.6834, and the highest F1 score of 0.7732 based on the comparison of test set metrics. According on a comparison of predicted photos, UperNet with ResNet101 as the backbone performs best for shading-containing images, whereas DeepLabv3+ with ResNet101 as the backbone per- forms best for blemish-containing images.	