Iraqi Journal of Intelligent Computing and Informatics (IJICI)
Vol. 4, 1, June 2025, pp. 77~89
ISSN: 2791-2868, DOI: 10.52940/ijici.v4i1.97

Asphalt Crack Detection and Segmentation Using Deep Learning

Shemeam T. Muhey [1], Sinan A. Naji [2]

[1] Informatics Institute for Postgraduate Studies, University of Information Technology and Communications, Iraq
[2] Dept. of Postgraduate Studies, University of Information Technology and Communications, Iraq

Article Info

Article history:

Received February 16, 2025
Revised March 26, 2025
Accepted April 5, 2025

Keywords:

Asphalt crack detection
Deep learning
Semantic segmentation
Unet 3+ model
YOLO model

ABSTRACT

Semantic segmentation is a computer vision task that utilizes deep learning algorithms to recognize a collection of pixels that form a distinct class. This technique could be used to early recognize road pavement cracks and reduce maintinence cost and inssuring safety for all road users. This research study presents an interesting semantic segmentation model for detecting asphalt cracks in roads based on deep learning techniques that combine object detection and semantic segmentation through three steps: In the first step, preprocessing the source images, then YOLOv10 model had been used for the crack detection framework. Finally, the UNet 3+ model was employed as a semantic segmentation model in which pixel-level segmentation is carried out. The geometric properties of the cracks are quantified to assess the damage in the road. The system has been trained, evaluated, and tested using two datasets: The SUT-Crack dataset and the IRD-Crack dataset. The proposed system shows excellent performance across different metrics such as Recall, Precision, Accuracy, mAP, Confidence score, and Dice coefficient. The accuracy reached up to 99.06%, demonstrating its ability to be applied in real-world environments.

Corresponding Author:

Shamem T. Muhey
Informatics Institute for Postgraduate Studies, University of Information Technology and Communications.
Email: ms202220727@iips.edu.iq

1. INTRODUCTION

Ensuring the safety and reliability of road infrastructure is a highly important procedure that requires continuous monitoring and analyzing of asphalt cracks [1] [2]. This issue involves automatically detecting and quantifying various types of asphalt cracks, which reflect the overall condition and safety of the pavement. Consequently, the early detection of these cracks can substantially support the maintenance planning process, which in turn prevents pavement degradation, reduces maintenance expenses, and enhances safety strategy that aims to protect road users and workers from potential accidents that may cause injury or even death [3] [4]. Generally, road networks and other transportation methods positively affected the economic growth of the country, communication, and social opportunities [4] [5].

Due to several environmental factors, usually roads degrade in many ways. The degradation can be recognized by the presence of cracks [6]. Generally, cracks may lead to the deterioration of the pavement if not maintained in due time. In fact, many types of cracks are there, each of which refer to the major cause of pavement degradation [7].

For defining automatic crack detection: Automatic crack detection uses an arbitrary image to identify whether or not there are any cracks in the asphalt and, if so, to return the location and size of each crack. The challenges associated with crack detection can be attributed to the following factors: lighting conditions, weather conditions, presence or absence of structural components, image orientation, shadows, oil stains, scaling, resolution, etc.

Generally, asphalt crack detection is usually done either by the labor manual visual inspection or by the automatic image-based methods [8]. The automatic image-based methods are usually divided into two classes:

the non-learning-based techniques and the learning-based techniques [8] [9]. The non-learning-based techniques mostly use different image processing techniques (e.g., edge detection, image segmentation, thresholding, morphological operation, etc.) [10] [11] [12].

On the other hand, One of the most significant developments in the fields of pattern recognition and classification is the deep learning methods based on deep Convolutional Neural Networks (CNNs). In general, some deep CNNs such as GoogLeNet [13], VGG-16 [14], AlexNet [15], ResNet [16], U-Net [17], DeepLab [18], YOLO [19], RetinaNet [20], etc., have become popular standards that are being integrated into several applications at the moment. Many authors have applied these CNN architectures in crack detection and classification models.

This study proposes an interesting semantic segmentation model based on deep CNN models for asphalt crack detection and quantification. The proposed model combines the YOLOv10 model and UNet 3+ structure with pixel feature extractor networks, which assign a class label for every pixel in the input image to improve the general performing of the CNNs. The full-scale skip connections in UNet 3+ combine low-level details with high-level semantics, which form feature maps in various scales. The proposed model was pre-trained on real images along with the corresponding ground truth images.

The rest of this research is arranged as in the following. Section 2 gives a short description of related work in automatic crack detection. Section 3 derives our proposed CNNs model. Section 4 shows the experimental results using two different datasets, and Section 5 concludes this paper with discussions and future work.

2. RELATED WORK

Numerous methods have been proposed to detect asphalt cracks in a single image of intensity or color images. Among these methodsthose that use learning algorithms have garnered a lot of interest lately and produced outstanding outcomes. Li et al. introduced an interesting version of the road crack detection model called RDD-YOLO [21]. The model integrates a simple attention mechanism (SimAM) into the backbone network to draw attention to important details in the input image. The neck structure is improved by substituting traditional convolution modules with GhostConv. As a result, there is less redundant data, fewer parameters, and less computing complexity, and this will achieve more lightweight and effective performance in the task of damage recognition. Finally, by substituting more precise bilinear interpolation for the nearest interpolation, the upsampling algorithm in the neck is enhanced. This finer interpolation technique enhances the detection results' accuracy and more effectively restores the image's subtle features.

Deng et al. proposed an integrated framework for automatic detection, segmentation, and measurement of road surface [22]. Three distinct computer vision algorithms are creatively merged in the suggested framework: At first, the real-time object detection algorithm YOLOv5 is used in order to detect cracks on the object-level, It attains a 91% mean average precision.. Secondly, a modified ResNet is created by inserting an attention gate module to higher accuracy segment the cracks at the pixel level and attains 87% intersection over union (IoU) on crack pixels segmention. Lastly, a novel surface feature quantification approach is created to more precisely determine both the width and length of segmental road cracks, achieving a 95% identification accuracy.

Shu et al. proposed a pavement crack recognition model that combines the street view image data source, which is a low-cost method, and the YOLOv5 target detection network [23]. The result shows that this network can effectively detect cracks with mAP of over 70%.

An et al. combines deep learning recognition with a concrete surface fracture identification and size estimation method. The authors named their system the Crack Identification Network (CIN) [24]. The accuracy rate attained 99%, which is capable of efficiently classifying concrete into cracks/non-cracks and clustering segmentation based on improved K-means and morphological methods.

Zhang Z. et al. proposed the ResUnet, a semantic segmentation neural network, which extracts road areas from high-resolution remote sensing images by combining the advantages of U-Net and residual learning. [25]. This model has two advantages: first, residual units make deep network training easier. Second, the network's rich skip connections could facilitate information propagation that enables the construction of networks with fewer parameters but higher performance. The result of the proposed method, which are defined as breakeven points, is 0.9187.

Zhang Q. et al. proposed an improved U-net network for crack detection and segmentation with a complex background [26]. To improve the recognition accuracy of narrow cracks in the road surface, the VGG16 and novel Up_Conv module are introduced as the backbone network. Furthermore, the Ca (Channel Attention) mechanism was added in U-net’s jump connection to distinguish cracks and background noise at the same time. The DG_Conv (Depthwise GSConv Convolution) module and UnetUp (Unet Upsampling)

modules are added in the decoding section, in order to extract richer information through more convolutional layers in the network. The results of the proposed system show a precision reached up to 87.4%.

He and Lau proposed an interesting model known as CrackHAM. This model is an encoder-decoder network based on the U-Net architecture and an innovative model network named HASP module, which is added to overcome the issue of deteriorating spatial data [27]. Furthermore, the channel attention module was used to capture abundant contextual information for high-level features and spatial attention for low-level features to extract rich edge information. The Multi-Fusion U-Net architecture is proposed to aggregate contextual information from feature maps of various sizes through the downsampling. The system achieves a precision of 86.41%.

Zhang et al. use a novel approach to recognize multi-type cracks using ResNet model integrated with a Convolutional Block Attention Module (CBAM) [28]. The spatial attention mechanism AM is produced by utilizing the features' inter-spatial relationship. By using average-pooling and max-pooling, an efficient feature descriptor is produced. The proposed model achieves a precision of 92.9%. For a more detailed literature review, interested researchers can refer to Refs. [2] and [7].

3. THE PROPOSED SYSTEM

The suggested system architecture for asphalt crack detection encompasses three phases: image pre-processing, semantic segmentation with the UNet 3+ model, and object detection using YOLOv10. Figure 1 illustrates the flowchart of the suggested CNN network architecture. A brief explanation of each of these models is given in the following sections.

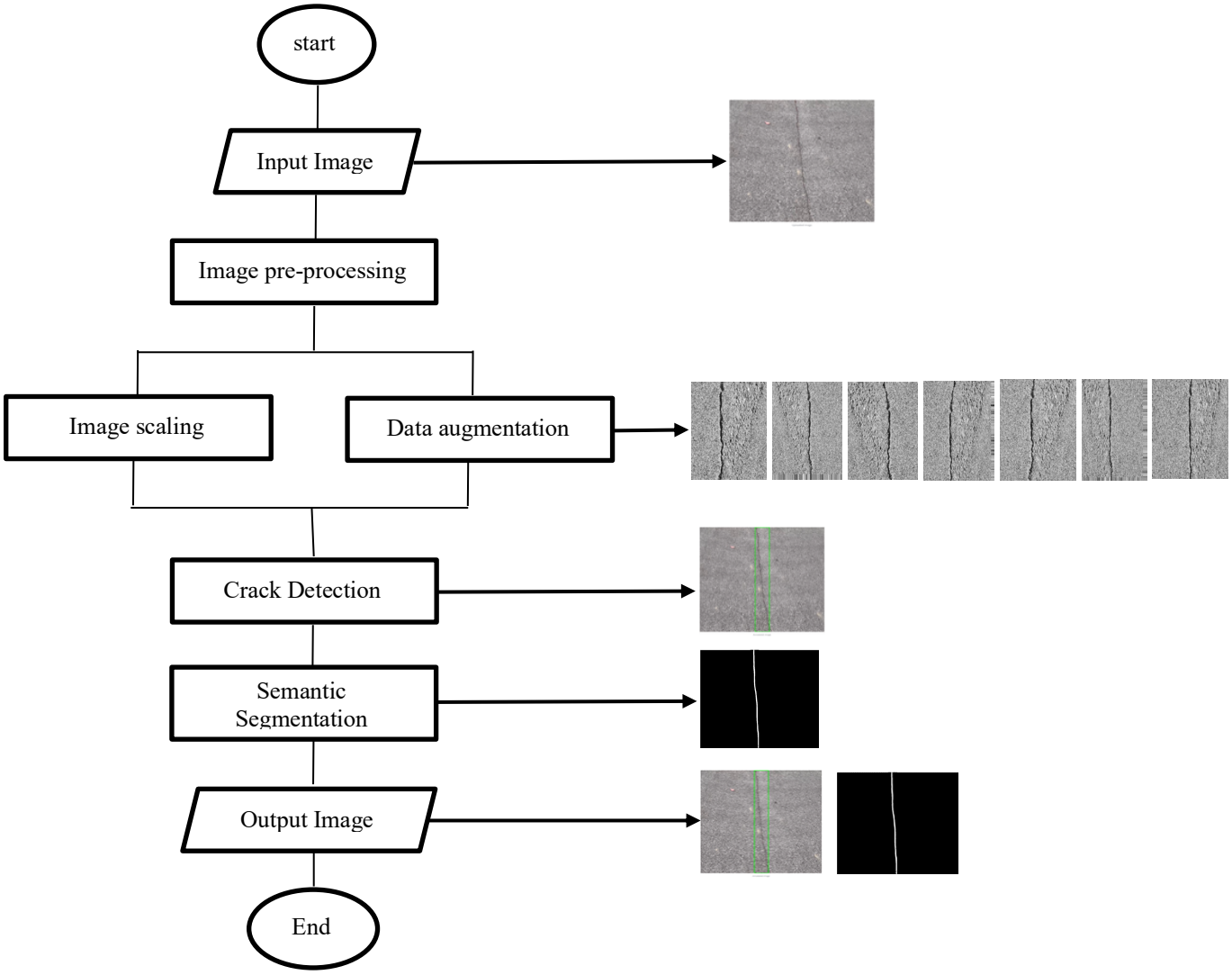


Figure 1. The general CNN network architecture of the asphalt crack detetion system

3.1. Image Pre-Processing

In order to enhance the quality and quantity of the dataset images used for training the system and to get a more effective learning model, pre-processing the raw input images is an essential step in deep learning strategy. A variety of pre-processing methods such as noise removal, color enhancement, color-space transformation, contrast adjustment, flipping, cropping, and rotation [29]. In this research, several processes were conducted on the SUT-Crack and IRD-Crack Datasets, as illustrated in the following sections.

3.1.1. Image Scaling

Practically, the input images of asphalt cracks may be collected from different sources and dissimilar datasets that have different sizes, resolutions, lighting conditions, etc. Therefore, the first step in building CNNs models is to resize the input images to a standard and fixed size [29]. In this research study, all input images were scaled to (640×640×3) and (320×320×3) to be compatible with the inputs used by YOLOv10 and UNet 3+ models, respectively. Image resizing provides a standard input size for the model and ensures computational efficiency.

3.1.2. Image generation (Augmentation)

Generally, deep learning with CNN-based methodologies requires large datasets in the training phase to enhance the model's ability to learn additional image patterns and make precise predictions [29] [30]. By employing numerous image transformations, the augmentation procedure enhances the training dataset. These transformations may include rotation, shifting, shearing, zooming, flipping, and reflecting [29] [30]. The dataset images of asphalt cracks are augmented to produce new images, thereby preventing the acquisition of undesired features, mitigating overfitting, and improving overall performance. Table 1 shows different transformation types and their corresponding parameters.

Table 1: Dataset augmentation with different transformations.

Transformation Type	Corresponding Values
Range of Rotation	30 degrees
Range of Width-Shift	10%
Range of Height-Shift	10%
Range of Shear	10%
Range of Zoom	[70% - 100%]
Horizontal-Flip	‘True’
Fill Mode Reflection	'Nearest'

3.1.3 Splitting the Dataset

Splitting the dataset into smaller sub-datasets is a common procedure used in machine learning, data mining, pattern recognition, etc. In this research study, the datasets SUT-Crack and IRD-Crack had been divided into three subsets: 70% for training, 20% for validation, and 10% for testing.

3.2. Crack Detection Using YOLO

Practically, object detection techniques aim to locate certain objects in the source image, typically by drawing a bounding box around each detected object [25]. Recently, the YOLO model (which stands for You Only Look Once) is considered one of the best real-time object detection architectures. It has a significant and wide-ranging impact on numerous computer vision projects [21] [31]. In this research study, the YOLOv10 had been used for the asphalt crack detection task, The YOLOv10 model exhibits the following attractive features [21] [31]:

- Extremely fast, accurate, and strong performance.
- Proposes the self-attention module.

- The transformer-based module is added to enhance the feature extraction.
- Using the label assignment method.
- Non-Maximum Suppression (NMS)-free is applied to eleminate redundant detections.

This research study uses YOLO for asphalt crack detection as follows: The source image is preprocessed and typically resized to a fixed size so it can be fed into the CNN model (i.e., 320 × 320 pixels). The image is divided into a (s × s) grid. Each grid cell is responsible for detecting cracks whose center falls within that cell. Then, for each grid cell, the model predicts certain bounding boxes and their probabilities (i.e., a score indicating how confident the model is that the box contains a crack).. Figure (2) shows the general architecture used in the step. Consistent Dual Assignments during training is applied, allowing the model to learn from rich supervision while eliminating the need for computationally expensive non-maximum suppression (NMS) during inference.
The key component of the dual label assignment strategy is the consistent matching metric used to evaluate the concordance between predictions and ground truth instances. The formula of which is as in equation 1.

$$m(\alpha,\beta)=S\ .\rho^{\wedge}\alpha.IoU(b\hat{},b)^{\wedge}\beta \quad (1)$$

where p is the classification score, bˆ and b denote the bounding box of prediction and ground truth, respectively. S represents the spatial prior indicating whether the anchor point of prediction is within the instance.

3.3 SEMANTIC SEGMENTATION BASED ON UNet 3+

Practically, the UNet 3+ network model is to take advantage of deep supervision along with full-scale skip connections. While deep supervision enhances accuracy by learning hierarchical representation from full-scale aggregated feature maps. The basic architecture of the UNet 3+ is composed of two main parts: Encoder and Decoder. The encoder implies a chain of convolutional layers that capture high-level features. Each decoder layer in UNet includes both smaller- and same-scale feature maps from the encoder and larger-scale feature maps from the decoder, which capture fine-grained features and coarse-grained semantics in complete sizes. Skip connections are also used in the basic architecture. The main idea of skip connections is that as the encoder reduces the spatial resolution, which can cause a loss of fine details, the skip connections help to preserve spatial details by passing them directly to the decoder. The key benefit of UNet 3+ is that it can be trained effectively with relatively small datasets. Figure (3) shows the general architecture of the proposed UNet 3+ for the automatic crack detection model inspired by [32]. Its primary goal is to assign a class label to each pixel in an image (i.e., crack or non-crack). As shown in this figure, the encoder reduces spatial dimensions through multiple convolutional layers; while the decoder resamples the feature maps back to the original resolution. For accurate segmentation, skip connections maintain details that may be lost. Skip connection formulated as in equation 2 in which i represents the down-sampling layer in the encoding layer and N represents number of encoding layers. The feature map X^{i}_{De} [32] is as in equation:

$$D^{i}_{De}=\begin{cases} X^{i}_{En,}\ i=N \\ H\left(\left[\underbrace{C(D(X^{k}_{En}))^{i-1}_{k=1},}_{Scales:\ 1^{th}-i^{th}}\ c(X^{i}_{En}),\ \underbrace{C(u(X^{k}_{De}))^{N}_{k=i+1}}_{Scales:\ (i+1)^{th}-N^{th}}\right]\right) \end{cases} \quad (2)$$

C(·) referes the convolutional operator, while H(·) uses convolution, batch normalization, and ReLU activation functions to implement feature aggregation mechanisms. D (·) stands for downsampling operator and U(·) stands for the upsampling operator. [·] stands for concatenation.

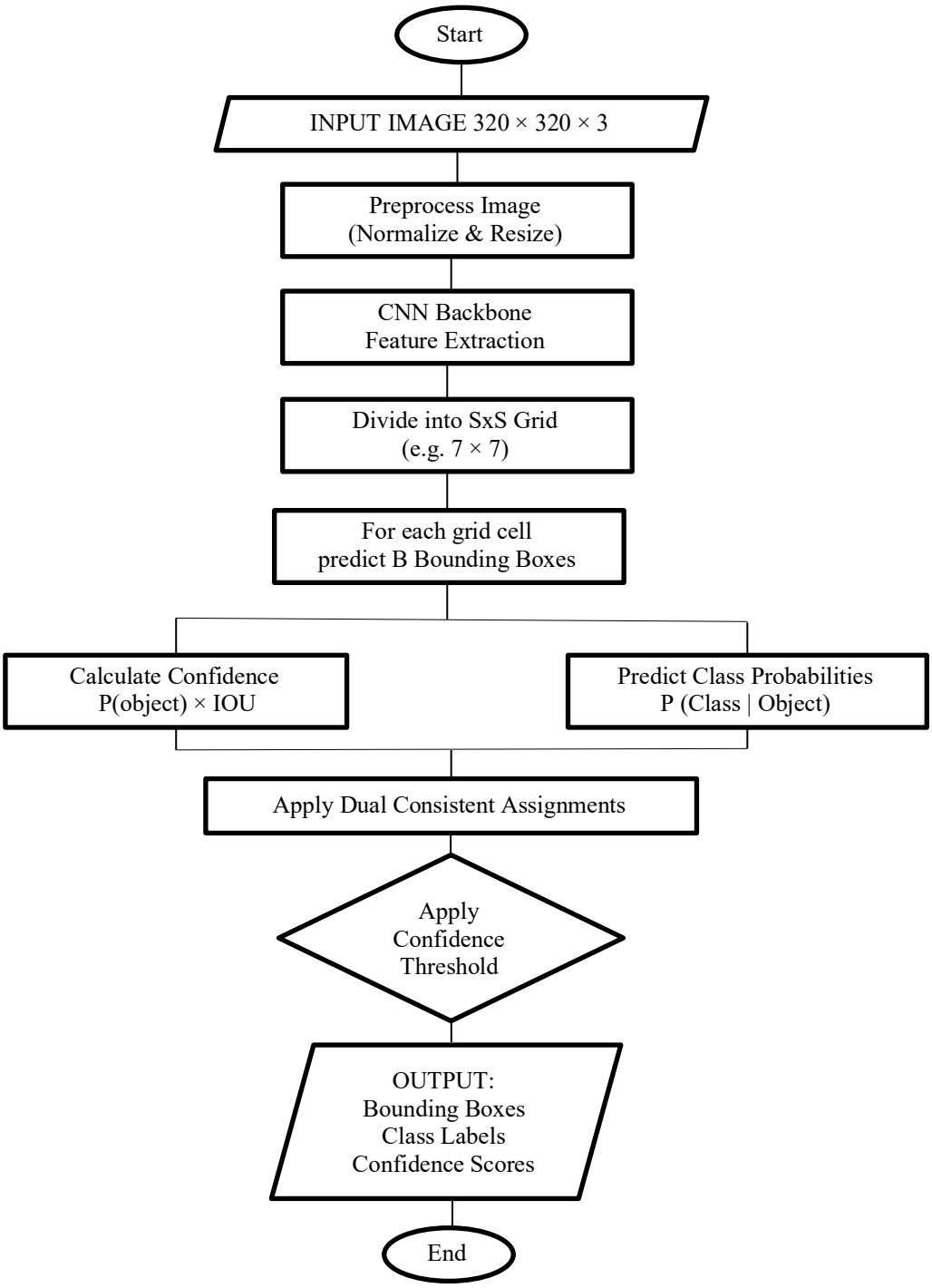


Figure 2. Crack detection in images based on YOLOv10

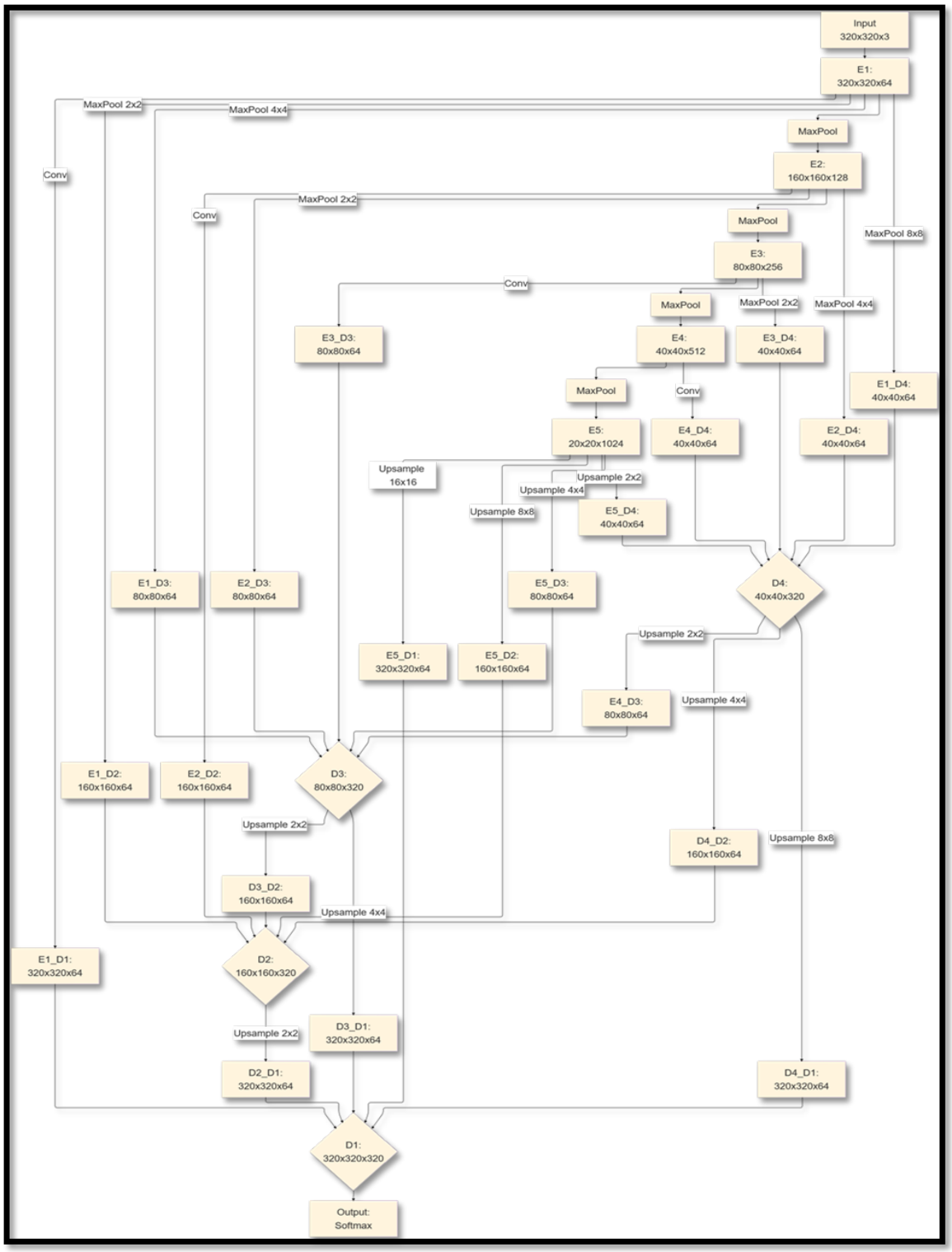


Figure 3. The general architecture of the proposed UNet 3+ for automatic cracks detection model.

Figure (3) illustrates how to create the feature maps in different layers, where the E_i refers to encoder feature maps and the D_i refers to decoder feature maps. As an example, E_3, the decoder receives the feature map from E_2 , then carries out the required convolutional filters and delivers to the next layer of the encoder E_4 directly. Unlike the classical UNet, a chain of intra-decoder skip connections transmits the high-level semantic information from the larger scale. For example, the low-level detailed information from the smaller-scale encoder layer E_1 and E_2 is delivered by a set of inter-encoder-decode skip connections. In this research study,

five feature maps had been generated. Further processing is needed to standardize the number of channels and eliminate unnecessary data. By trial and error, we revealed that the convolution with 64 (3 × 3) filters is a good option. We additionally apply a feature aggregation process, comprising 320 filters of size (3 × 3), batch normalization, and a ReLU activation function, on the concatenated feature map from five scales in order to smoothly combine the shallow exquisite information with deep semantic information.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the suggested system is demonstrated by the experimental findings in this section. The data collection concerning the crack detection task is conducted using two distinct datasets. The dataset involves two subsets: one for the original images and a second for their corresponding ground truth annotations. These are:

- **Set 1**: The "SUT-Crack Dataset" [33] of the "Sharif University of Technology" implies a high-resolution original images depicting asphalt road cracks with dimensions of (3024 × 4032) pixels. The dataset is publicly available to enable crack detection through the use of many deep learning methods.
- **Set 2**: The "IRD-Crack Dataset", which represents our local dataset. It consists of asphalt crack images that were collected in cooperation with the directorate of highways and bridges in Diyala governorate. It includes various types of images that present various problems for crack detection, such as shadows and stains of oil. A fixed height of one meter, directly above the pavement, was used to capture the high-quality photos. using a digital camera type (Canon RP + 18-135mm), with a resolution of (6240 × 4160). All pictures were captured during morning hours to ensure clarity and similar lighting conditions.

(a) (b) (a) (b)

Figure 4. Dataset images; (a) The source image; (b) The ground truth images.

Figure (5) shows samples of the crack detection results using our proposed YOLOv10, while Figure (6) shows the precision-confidence curve. The semantic segmentation of the cracks using the proposed UNet 3+ is shown in Figure (7).

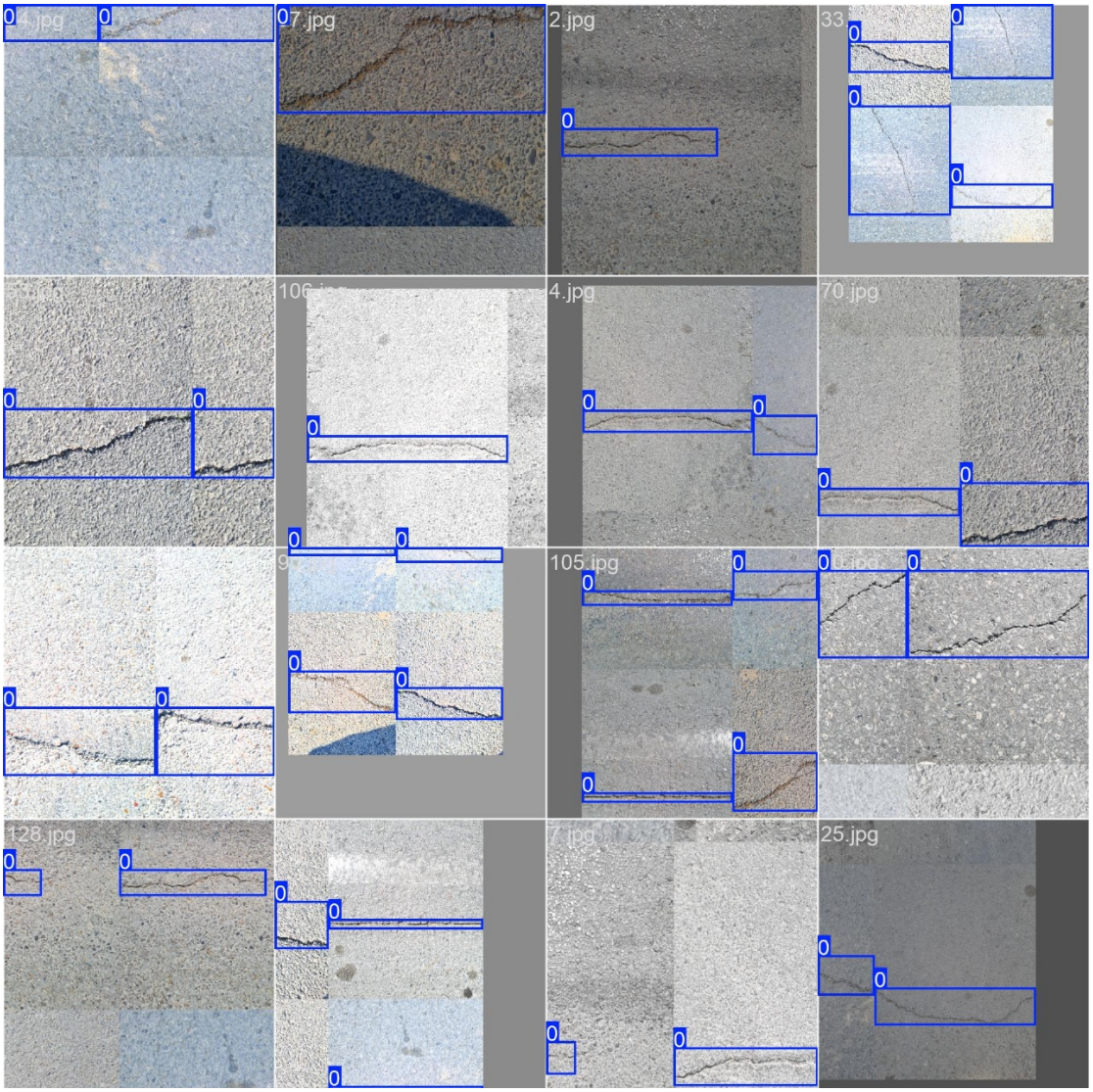


Figure 5. Crack detection results using the proposed YOLOv10.

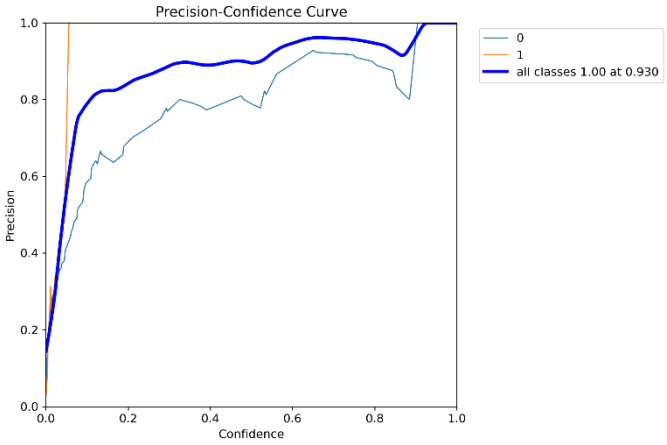
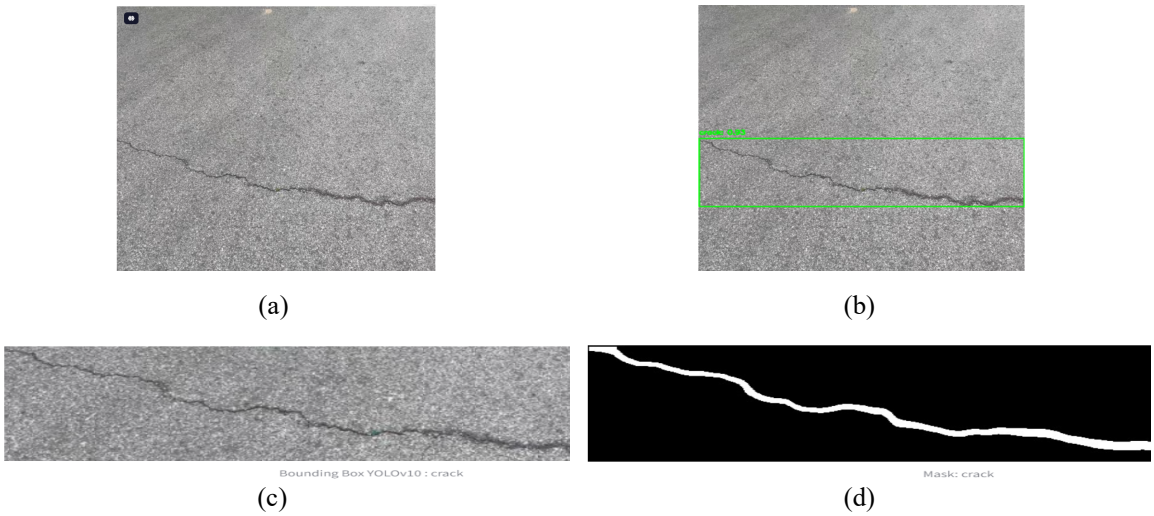


Figure 6. Precision-confidence curve of YOLOv10.



(a) (b) (c) (d)

Figure 7. Semantic segmentation results of the proposed UNet 3+; (a) Source image; (b) Crack detection; (c) Bounding box; (d) Semantic segmentation.

For quantitative evaluation, numerous metrics, such as Accuracy (ACC), Precision (Pr), Recall (Re), F1-score, mean Average Precision (mAP), Intersection over Union (IoU), and Dice Coefficient (DC), were assessed in order to quantitatively analyze the experimental results. Table (3) shows a sample of the crack detection results of the proposed YOLOv10 model using 200 epochs. Table (4) illustraits the final semantic segmentation findings of the suggested UNet 3+ model. Figure (8) shows the performance results during the training and validation phases of the UNet 3+ Model, showing the Accuracy, Precision and Recall using 100 epochs.

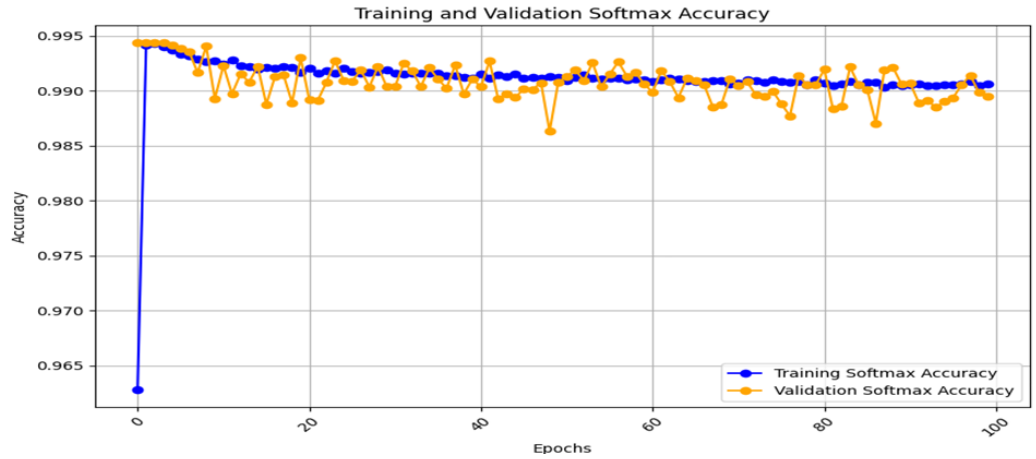
Table (3): Crack detection results of the proposed YOLOv10 model.

	Precision	**Recall**	**mAP@0.5**	**mAP@0.5 : 0.95**
YOLOv10 (Ours)	99.831	0.91	0.68901	0.54582

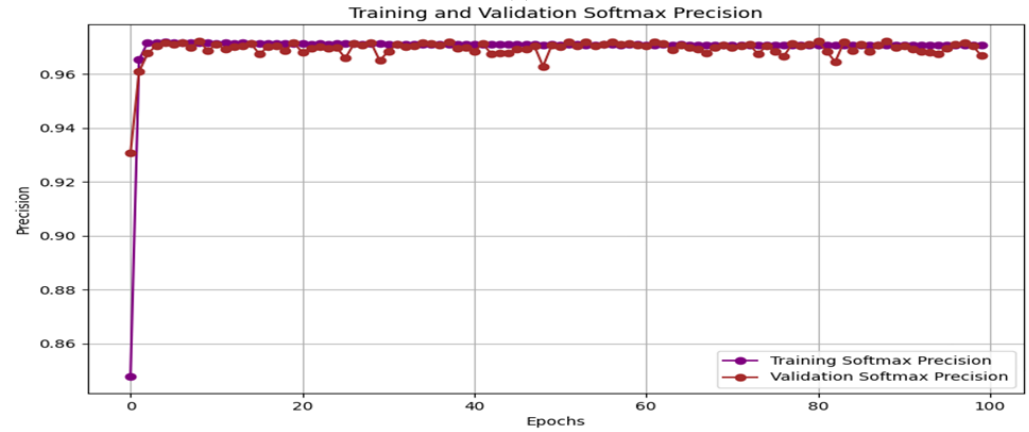
Table (4): Semantic segmentation results of the proposed UNet 3+ model.

	Loss	**IoU**	**ACC.**	**Pre**
UNet 3+ (ours)	0.499050081	0.045811992	0.98949939	0.9667
	Re	**Confid.**	**DC**	**Map**
	0.9864	0.5	0.97646	0.96701

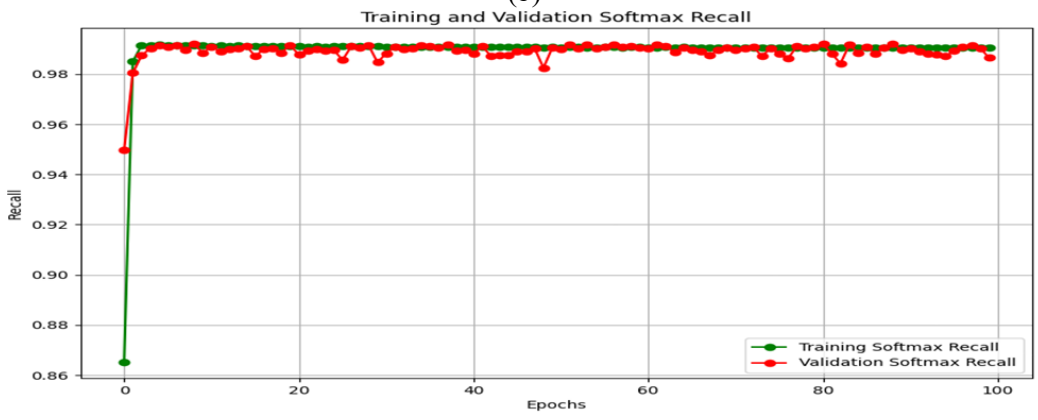
As shown in Table 4, the maximum accuracy, precision, and recall scores achieved during training were 0.9906, 0.9706, and 0.9905, respectively, whereas for the validation dataset, they were 0.9894, 0.9667, and 0.9864, respectively.



(a)



(b)



(c)

Figure 8. Performance results during the training and validation phases of the UNet 3+ Model; (a) Accuracy; (b) Precision; (c) Recall.

5. CONCLUSION

In this research, an interesting deep learning model has been proposed for automatic asphalt crack detection in image(s). As demonstrated in experiments, the suggested model shows excellent results regarding precision in different imaging conditions. It demonstrates its suitability for use by various government agencies and civil departments. that are interested in the detection of pavement cracks and implementing the suitable repair to reduce cost, speed up the work, and increase the safety of works. The YOLOv10 model is used in the first step to locate the crack regions in the source images, and the UNet 3+ is used in the second phase to perform a pixel-level segmentation process. Furthermore, the IRD-Crack (Iraqi-Roads Dataset) is our own dataset that consists of highly diverse asphalt cracks and contains different kinds of real-world image

environments with different lighting conditions. This dataset can be used publicly by other researchers in future research studies.

List of abbreviations

Abbreviations	Full Form
IoU	Intersection over union
RDD	Road Damage Detection
CIN	Crack Identification Network
YOLO	You Only Look Once
CNN	Convolutional Neural Networks
AI	Artificial intelligence
ANN	Artificial Neural Networks
BN	Batch Normalization
DL	Deep Learning
P	Precision
R	Recall
CBAM	Convolutional Block Attention Module
SimAM	simple attention mechanism
ReLU	Rectified Linear Unit
Resnet	Residual Neural Network
NMS	Non-Maximum Suppression
IRD	Iraqi-Roads Dataset
mAP	Mean Average Precision
TN	True Negative
TP	True Positive
FN	False Negative
FP	False Positive

REFERENCES

[1] L. Pauly, D. Hogg, R. Fuentes, and H. Peel, "Deeper networks for pavement crack detection," in *Proceedings of the 34th ISARC*, 2017, pp. 479-485.

[2] S. D. Nguyen, T. S. Tran, V. P. Tran, H. J. Lee, M. J. Piran, and V. P. Le, "Deep learning-based crack detection: A survey," *International Journal of Pavement Research and Technology,* vol. 16, pp. 943-967, 2023.

[3] M. Salman, S. Mathavan, K. Kamal, and M. Rahman, "Pavement crack detection using the Gabor filter," in *16th international IEEE conference on intelligent transportation systems (ITSC 2013)*, 2013, pp. 2039-2044.

[4] M. Iacono and D. Levinson, "Mutual causality in road network growth and economic development," *Transport Policy,* vol. 45, pp. 209-217, 2016.

[5] A. Dhamija and V. Dhaka, "A novel cryptographic and steganographic approach for secure cloud data migration," presented at the 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 2015.

[6] V. Mandal, L. Uong, and Y. Adu-Gyamfi, "Automated road crack detection using deep convolutional neural networks," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 5212-5215.

[7] Y. Hamishebahar, H. Guan, S. So, and J. Jo, "A comprehensive review of deep learning-based crack detection approaches," *Applied Sciences,* vol. 12, p. 1374, 2022.

[8] Y. Liu, "Evaluation and optimization of image segmentation models in pavement crack detection," in *Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024)*, 2025, pp. 509-514.

[9] M. Abdellatif, H. Peel, A. G. Cohn, and R. Fuentes, "Pavement crack detection from hyperspectral images using a novel asphalt crack index," *Remote sensing,* vol. 12, p. 3084, 2020.

[10] I. Abdel-Qader, O. Abudayyeh, and M. E. Kelly, "Analysis of Edge-Detection Techniques for Crack Identification in Bridges," *Journal of Computing in Civil Engineering,* vol. 17, 2003.

[11] H. Zhao, G. Qin, and X. Wang, "Improvement of canny algorithm based on pavement edge detection," *IEEE, 3rd international congress on image and signal processing,* 2010.

[12] Z. Yang, C. Ni, L. Li, W. Luo, and Y. Qin, "Three-stage pavement crack localization and segmentation algorithm based on digital image processing and deep learning techniques," *Sensors,* vol. 22, p. 8459, 2022.
[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov*, et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
[14] K. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 2015, pp. 234-241.
[18] L.-C. Chen, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587,* 2017.
[19] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
[20] T. Lin, "Focal Loss for Dense Object Detection," *arXiv preprint arXiv:1708.02002,* 2017.
[21] Y. Li, C. Yin, Y. Lei, J. Zhang, and Y. Yan, "RDD-YOLO: Road Damage Detection Algorithm Based on Improved You Only Look Once Version 8," *Applied Sciences,* vol. 14, p. 3360, 2024.
[22] L. Deng, A. Zhang, J. Guo, and Y. Liu, "An integrated method for road crack segmentation and surface feature quantification under complex backgrounds," *Remote Sensing,* vol. 15, p. 1530, 2023.
[23] Z. Shu, Z. Yan, and X. Xu, "Pavement crack detection method of street view images based on deep learning," in *Journal of Physics: Conference Series*, 2021, p. 022043.
[24] Q. An, X. Chen, X. Du, J. Yang, S. Wu, and Y. Ban, "Semantic Recognition and Location of Cracks by Fusing Cracks Segmentation and Deep Learning," *Complexity,* vol. 2021, p. 3159968, 2021.
[25] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters,* vol. 15, pp. 749-753, 2018.
[26] Q. Zhang, S. Chen, Y. Wu, Z. Ji, F. Yan, S. Huang*, et al.*, "Improved U-net network asphalt pavement crack detection method," *Plos one,* vol. 19, p. e0300679, 2024.
[27] M. He and T. L. Lau, "CrackHAM: A novel automatic crack detection network based on U-Net for asphalt pavement," *IEEE Access,* 2024.
[28] Z. Zhang, K. Yan, X. Zhang, X. Rong, D. Feng, and S. Yang, "Automated highway pavement crack recognition under complex environment," *Heliyon,* vol. 10, 2024.
[29] M. M. Islam, M. B. Hossain, M. N. Akhtar, M. A. Moni, and K. F. Hasan, "CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack," *Algorithms,* vol. 15, p. 287, 2022.
[30] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data,* vol. 6, pp. 1-48, 2019.
[31] C.-Y. Wang and H.-Y. M. Liao, "YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems," *APSIPA Transactions on Signal and Information Processing,* vol. 13, 2024.
[32] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto*, et al.*, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2020, pp. 1055-1059.
[33] M. Sabouri and A. Sepidbar, "SUT-Crack: A comprehensive dataset for pavement crack detection across all methods," *Data in Brief,* vol. 51, p. 109642, 2023.