

RESEARCH

Open Access



Rs-net: Residual Sharp U-Net architecture for pavement crack segmentation and severity assessment

Luqman Ali^{1,3,4†}, Hamad AlJassmi^{2,3†}, Mohammed Swavaf^{1,4}, Wasif Khan¹ and Fady Alnajjar^{1,4*}

[†]Luqman Ali and Hamad AlJassmi contributed equally to this work.

*Correspondence: fady.alnajjar@uaeu.ac.ae

¹ Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain 15551, United Arab Emirates

² Department of Civil and Environmental Engineering, United Arab Emirates University, Al Ain 15551, United Arab Emirates

³ Emirates Center for Mobility Research (ECMR), United Arab Emirates University, Al Ain 15551, United Arab Emirates

⁴ AI and Robotics Lab, College of IT, United Arab Emirates University, Al Ain 15551, United Arab Emirates

Abstract

U-net, a fully convolutional network-based image segmentation method, has demonstrated widespread adaptability in the crack segmentation task. The combination of the semantically dissimilar features of the encoder (shallow layers) and the decoder (deep layers) in the skip connections leads to blurry features map and leads to undesirable over- or under-segmentation of target regions. Additionally, the shallow architecture of the U-Net model prevents the extraction of more discriminatory information from input images. This paper proposes a Residual Sharp U-Net (RS-Net) architecture for crack segmentation and severity assessment in pavement surfaces to address these limitations. The proposed architecture uses residual block in the U-Net model to extract a more insightful representation of features. In addition to that, a sharpening kernel filter is used instead of plain skip connections to generate a fine-tuned encoder features map before combining it with decoder features maps to reduce the dissimilarity between them and smoothes artifacts in the network layers during early training. The proposed architecture is also integrated with various morphological operations to assess the severity of cracks and categorize them into hairline, medium, and severe labels. Experiments results demonstrated that the RS-Net model has promising segmentation performance, outperforming earlier U-Net variations on testing data for crack segmentation and severity assessment, with a promising accuracy (>0.97)

Keywords: Pavement deterioration, Crack detection, Improved U-Net, U-Net, Deep Learning, Severity Assessment

Introduction

Surface cracks are the initial indicators of pavement structure deterioration, which affects performance and health, and pose a potential threat to the safety of vehicles. Structural assessment is crucial for effective maintenance and predicting potential failures. The pavement condition data can be collected by subjective human experts (manually) by visually inspecting and evaluating the road, or automatically. Manual Inspection techniques are laborious, time-consuming, inspector-dependent, inconsistent [1, 2], and easily vulnerable to the perspicacity of the inspector. Inadequate inspection and condition assessment can result in various accidents [3, 4]. The accuracy and efficiency standards of pavement management systems cannot be matched by manual

pavement condition evaluation techniques. In recent years, there has been a significant increase in the deployment of automated vision-based pavement detection systems, which focus on the characterization of surface cracks and quantitative modeling based on collected data. These systems frequently utilize vehicle-mounted equipment to efficiently cover extensive road networks. Developing a reliable pavement condition assessment system remains a challenging task due to the various complexities associated with the images acquired from pavement surfaces. The factors include the pavement surface crack's asymmetrical size and shape, varying intensities within the image, the existence of different textures, the presence of shadows, and the similarity between the crack and pavement surfaces. Machine learning techniques have been successfully applied across various applications to solve various real-world problems [5–7]. In crack detection field, the earlier classical image processing-based approaches focused on using various filters [8–11], thresholding [12, 13], and edge detection techniques [14, 15] for crack detection in paved surfaces. However, their widespread adoption has been stymied by factors such as requiring a lot of human intervention, being affected by lighting, and having no continuity or contrast between neighboring crack pixels. The shortcomings of conventional approaches can be addressed by feature extraction-based machine learning techniques which involve extraction of handcrafted features and their classification [16], but these approaches lack robustness if ineffective representations of surface cracks are extracted from the input images.

Recent advances in technology have led to widespread adoption of Deep Learning (DL) techniques, for the classification [17–24], localization [23, 25–27], and segmentation [24, 28–34] of cracks in civil infrastructures. The DL models learn the efficient features from the input images automatically, reducing the necessity for hand-crafted feature extraction required for traditional ML approaches. Liu et al. [24] developed a dataset for asphalt pavement fatigue crack classification using visible, infrared, and fused images, and evaluated thirteen CNN models. They applied Grad-CAM and Guided Grad-CAM for model interpretation and investigated the impact of image types on classification accuracy, highlighting the applicability of infrared thermography in crack detection. In DL-based segmentation models, U-Net [35] has been extensively used, particularly for pavement crack segmentation due to its simple architecture, high detection speed, and accuracy. The U-Net architecture was utilized for the first time by Cheng et al. [36] to segment road cracks, and it has since been proven effective in a variety of supervised crack segmentation methods [30, 36–39]. Despite its benefits, the method still has some major limitations that must be addressed, including the merging of incompatible features from the encoder and decoder, vanishing, and exploding gradients, and others. The performance of the U-Net can be enhanced by incorporating a pre-trained image classification models [40–42] and residual blocks [43, 44] as the encoders. Liu et al., [42] proposed a UNet-based model by replacing its encoder with typical CNN models to have different computational and model complexity and also integrates visual explanations to interpret the model. The deeper U-Net architectures based on residual encoders exhibit better pixel-wise crack detection performance than models having pre trained encoders [45]. The deep networks can resolve challenges such as the disappearance and exploding gradient by introducing the concept of residual connections [46]. In these networks the output of a layer is added to its input before being passed on to the next

layer, thus allowing gradients to flow more efficiently through the network and leading to improved training and performance. Yang et al., [47] proposed Residual U-Net architecture in which the encoder uses residual structure to enhance feature learning capabilities of the traditional U-Net. Huan [45] proposed ResCrack U-Net architecture consisting of seven residual units for Pixelwise asphalt concrete pavement crack detection. Similarly, Yu et al., [48] also proposed RUC-Net, a novel approach to demonstrate the effectiveness of Residual U-Net in crack detection tasks, and show that the architecture can effectively learn features and patterns in pavement images that are indicative of cracks. Additionally, Liu et al. [49] proposed a Multi-Scale Residual Encoding Network for concrete crack segmentation, integrating a residual structure and attention mechanism to improve feature extraction and crack detection accuracy [3]. These studies collectively highlight the effectiveness of residual-based U-Net models in advancing crack detection capabilities in infrastructure maintenance and safety. The performance of the U-Net can be enhanced by integrating attention strategy into the skip connections and blocks containing convolutional and attention layers [38, 40, 50–56]. Zou et al. [57] proposed a new encoder-decoder model resembling Seg-Net structure, which outperforms basic U-Net and Seg-Net by fusing output feature maps from multiple scales. Similarly, the authors in Liu et al. [58] proposed an encoder-upscale method as an alternative to the traditional encoder-decoder structure by fusing upscaled multi-scale and multi-level feature maps. Skip connections in encoder-decoder networks combine shallow and deep features from the encoder and decoder respectively to improve accuracy in dense prediction tasks such as image segmentation. However, this feature fusion process may not always match effectively. The encoder-decoder network's mismatch is a result of the difference in the features computed by the encoder (fine-grained and low-level) and decoder (high-level, coarse-grained, and semantic). This leads to the fusion of features that are semantically dissimilar and results in a blurred feature map, which in turn negatively impacts the output segmentation map by under- and/or over-segmenting ROIs. By optimizing and redesigning skip connections in U-Net, the model improves its connection between feature maps from various layers, thereby improving its overall performance [59, 60]. Huang [60] and Zhang [61] uses dense skip connections to combine the features from both the encoder and decoder nodes, resulting in the learning of more extensive feature maps through deep supervision. Khaledyan et al. [62] introduced a technique that employs sharpening filters to merge encoder and decoder features, enhancing the accuracy of breast ultrasound segmentation. Additionally, Sharp Dense U-Net, an improved U-Net architecture has been utilized for nucleus segmentation from histopathology images in various studies [63, 64]. Xie et al., [65] integrate a sharpening skip-connection layer with the Swin Transformer-based U-Net structure in a cascaded manner for unsupervised EM image registration. Moreover, a novel network structure named UNet-sharp (UNet#) combines dense and full-scale skip connections to aggregate feature maps at different scales, improving segmentation accuracy for organs and lesions across various modalities and dimensions [66]. Zhou et al. [67] presented a Wide U-Net architecture named U-Net++ in which the number of filters are increased on both encoder and decoder ends. U-Net++ is a combination of U-Nets with different depths and decoders connected through modified skip connections. Despite its effectiveness, it is complicated and has extra blocks for certain tasks and requires more learnable parameters.

Zioulis et al., [68] proposed hybrid skip connections that provides balanced exchange of information between low and high frequency features from the encoder and decoder maintaining the sharpness of edges while reducing the minimizing texture transfer artifacts. To address the aforementioned issues with the U-Net architecture and leverage the strengths of both Sharp and Residual U-Net, this paper proposes a Residual Sharp U-Net architecture, which combines the advantages of both Sharp and Residual U-Net models. The residual block helps enable a smoother flow of gradients through the network, while the use of a sharpening kernel in the skip connection through depth-wise convolution helps to bridge the semantic gap between the encoder and decoder features. Using depth-wise convolutions and a sharpening filter, the network can emphasize fine details in the early-level features and improve feature fusion, ultimately leading to better representation learning and more accurate segmentation results. The proposed Residual Sharp U-Net architecture outperforms previous U-Net models in crack segmentation and severity assessment, according to the results of experiments conducted on two publicly available datasets. The model achieved improved performance in various evaluation metrics including dice coefficient, Jaccard index, mIoU, precision, recall, and accuracy. The proposed approach has potential applications in automated pavement management systems for detecting cracks of varying types and severity. The main contribution of the article are as follows:

1. We proposed a novel architecture named RS-Net that combines the strengths of Residual and Sharp U-Net architectures to improve feature representation in both the encoder and decoder blocks, reducing the semantic gap between them through the use of skip connections.
2. We compared the performance of the proposed RS-Net architecture with the traditional U-Net, Sharp U-Net, and Residual U-Net on two publicly available crack segmentation datasets demonstrating that RS-Net consistently outperforms these models across all evaluation metrics, thus validating the effectiveness of our proposed approach.
3. We investigated the impact of various loss functions on the performance of RS-Net. The findings show that RS-Net with Binary Cross Entropy (BCE) Loss achieved higher accuracy, while RS-Net with Binary Focal Loss (BFL) excelled in Jaccard, mIoU, and Dice coefficient metrics, underscoring the importance of loss function selection in model performance.
4. For crack severity assessment, we utilized the crack segmentation mapping results from RS-Net in conjunction with morphological operations to categorize cracks into severity levels, such as severe, moderate, and hairline. This demonstrates the practical applicability of our approach in real-world scenarios.

The remaining of the article is organized as follows. Section [System overview](#) explain the overview of the proposed system. The experiments and results are discussed in section [Experiments and results](#) followed by discussion and conclusion in the fourth section. The last section is about the future directions.

System overview

The proposed crack detection and assessment system consists of four distinct modules as depicted in Fig. 1. The first module is the datasets utilization, which involves using two publicly available crack segmentation datasets as input to the crack segmentation models. The second module is the implementation of multiple segmentation models, including U-Net, Residual U-Net, Sharp U-Net, and Residual Sharp U-Net followed by the post-processing module. The post-processing module is responsible for the preprocessing of the segmented data. Finally, the crack measurement and severity assessment module quantifies important crack characteristics, such as length, width, and severity.

Datasets

The proposed work utilized two publicly available datasets D1 [69] and D2 [70]. These datasets are made by combination and refinement of a diverse range of existing crack segmentation datasets such as DeepCrack [58], CrackTree [71], Crack500 [72], CFD(Crack Forest Dataset) [73] and others. The datasets are standardized by ensuring consistency in the image resolution. The dataset D1 [69] consists of 11.29K images, each with a resolution of 440*440 while the dataset D2 [70] consists of 9.49K images having a resolution of 400*400 pixels. These images in the dataset have a diverse range of backgrounds, types of cracks, surfaces, and ground truth annotations, providing a rich and varied source of information for the training and evaluation of the crack detection and assessment system. A sample of the dataset’s images is shown in Fig. 2.

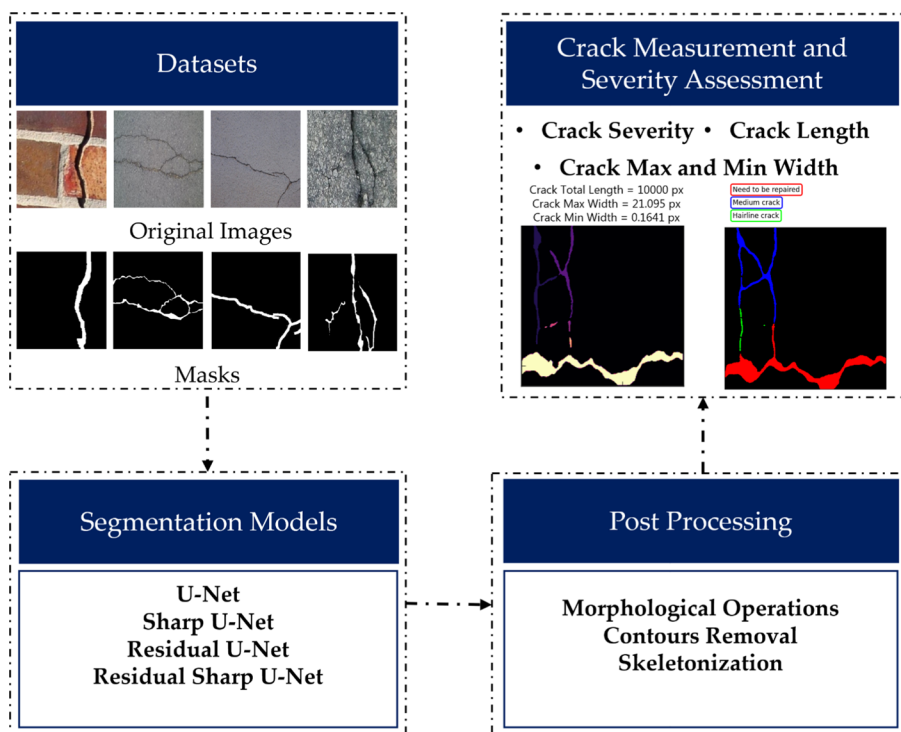


Fig. 1 Overview of the proposed system

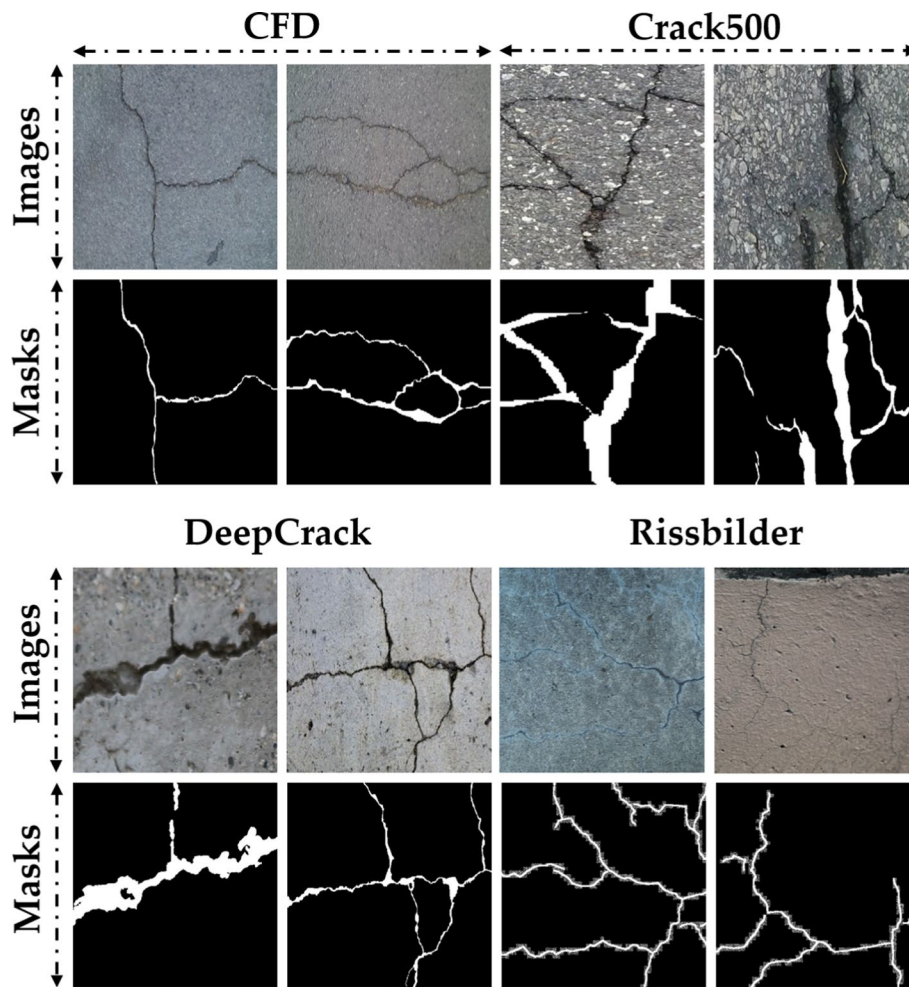


Fig. 2 Sample of the dataset's images

Segmentation models

The two publicly datasets are used to train various segmentation models i.e., U-Net, Sharp U-Net, Residual U-Net and the proposed Residual Sharp U-Net (RS-Net). Each model is explained in detail below.

U-Net

U-Net is one of the most well-liked networks for crack segmentation tasks in the field of structural health monitoring [30, 74]. The main advantages are its ability to capture fine-grained details, handle varying input sizes, and high accuracy using a minimal amount of training data. The U-Net architecture consists of the encoder (compressed path) and decoder networks (extended path) as depicted in Fig. 3. Encoder is the adoptable and customizable part in which the number and size of layers are optimized to achieve optimal feature extraction performance. The encoder path consists of four blocks with two convolutional layers and one max-pooling layer each. The convolutional layer increases the number of channels and generates

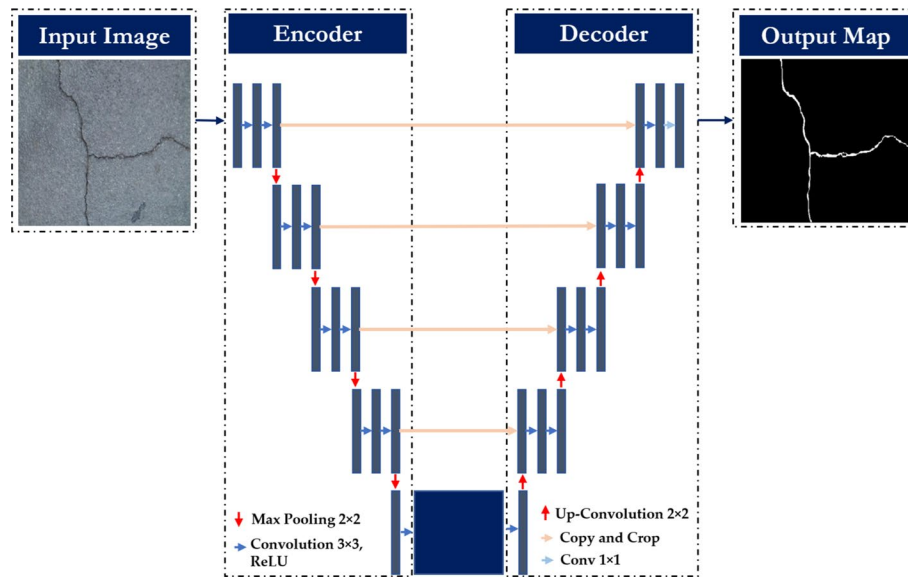


Fig. 3 U-Net architecture

a group of feature maps that contain details regarding the existence and positioning of distinct features within the input image. The Max pooling layer downsizes feature maps by half and doubles their number in each coding block. The extended path has two convolutional layers and one deconvolutional layer in its decoding blocks. The convolutional layers in the decoding blocks shrinks the channels, and the deconvolutional layer enlarges the feature map size, reducing the feature maps by half and doubling the size for each encoding block. Skip connections retain important semantic information by fusing the low-level convolutional features of an encoding block with the input of its corresponding high-level convolutional features of the decoding block at the same level. This, in turn, enhances the precision of segmentation.

Residual U-Net

Residual U-Net is an enhanced version of the U-Net architecture that integrates Residual Neural Networks and U-Net model strengths, resulting in better performance in image segmentation tasks [75]. By integrating skip connections within the residual unit, information can be propagated without degradation, resulting in a simpler network training process and fewer training parameters required. Despite having fewer parameters, the Residual U-Net architecture can achieve comparable or even better performance on semantic segmentation tasks. The Residual U-Net architecture is like the conventional U-Net architecture, consisting of an encoding path, a decoding path, and a bridge path constructed using residual units as shown in Fig. 4. The residual unit is incorporated into each of these paths. Each residual unit in the Residual U-Net architecture is composed of an identity mapping, and two convolution blocks with a size of 3×3 , each with BN, ReLU, and convolutional layers. The identity mapping connects the input and output of the unit.

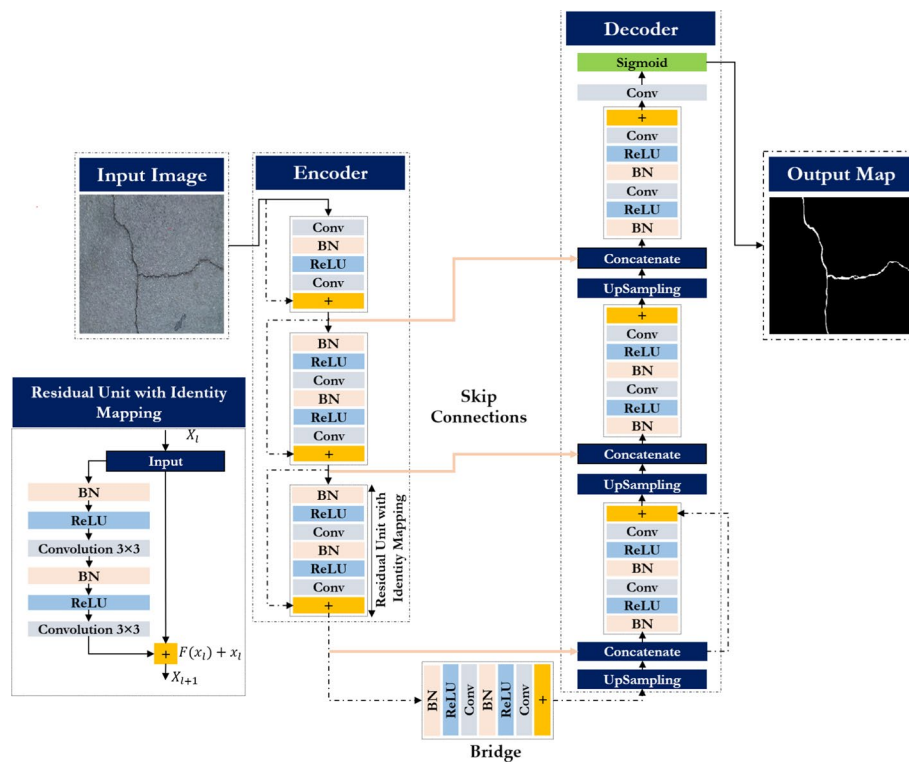


Fig. 4 Residual U-Net architecture

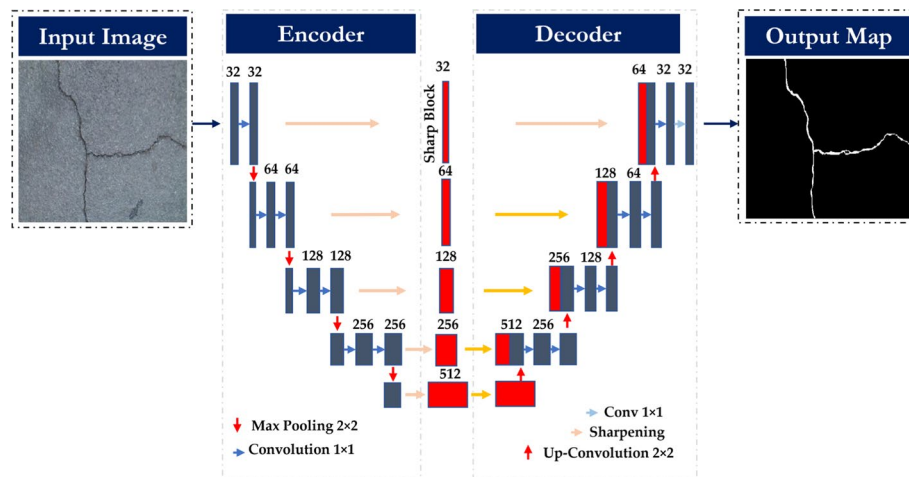


Fig. 5 Sharp U-Net architecture

Sharp U-Net

The Sharp U-Net is a newly proposed architecture by Zunair et al. [76], which is tailored for biomedical image segmentation. The architecture utilizes a depth-wise convolution operation with a sharpening spatial kernel on the encoder features before they are combined with the decoder features as shown in Fig. 5. This is done to

minimize any potential mismatch between the encoder and decoder subnetworks, thereby improving the accuracy of the segmentation process.

The Sharp block is a part of the Sharp U-Net framework, which applies a sharpening spatial kernel to each channel of the encoder features before fusing them with the decoder features. This involves using M filters to convolve each input channel separately with the kernel K , yielding a feature map of size $W \times H \times 1$. By applying the filter to each channel of the encoder features, the Sharp block not only enhances the fusion of semantically similar features but also reduces high-frequency noise during early training stage. The feature fusion process involves padding the encoder features to match the dimensions of the decoder features. The padded encoder and decoder features are then combined and passed through a depth wise convolution layer to generate the final output, which is a 3D tensor with dimensions $W \times H \times M$. Where W, H , and M represents the width, height, and number of encoder feature maps. Figure 6 illustrates the sequence of operations involved in the sharp block.

Residual Sharp U-Net (RS-Net)

The Residual Sharp U-Net (RS-Net) architecture proposed in this work is a novel modification of the original U-Net as depicted in Fig. 7.

The RS-Net architecture incorporates skip connections with sharp blocks to allow the network to capture both low-level and high-level features. The sharp block output can be mathematically represented in Eq. 1:

$$Y_L = LaplaceConv(X_L, K) \tag{1}$$

where X_L is the input feature map for the L^{th} residual block, Y_L is the output feature map after sharpening, and K is the Laplacian filter kernel. After applying the Laplacian filter kernel K to each input channel separately, the resulting feature maps are stacked together to obtain the final output feature map Y of size $W \times H \times M$. The Laplace operator is highly sensitive to intensity variations, making it adept at highlighting edges and fine details in images. By emphasizing high-frequency components, the Laplace operator enhances texture information within the encoder features, which is essential for the accurate identification and delineation of cracks in pavement images. The use of M filters in each sharp block allows the network to learn a richer set of features and capture more

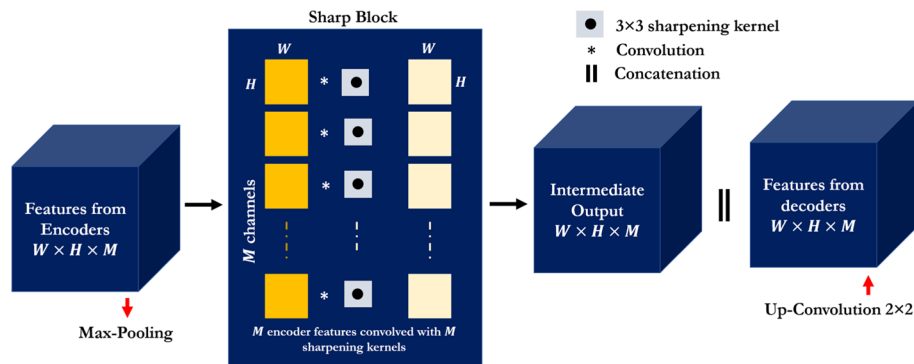


Fig. 6 Sequence of operations involved in the sharp block

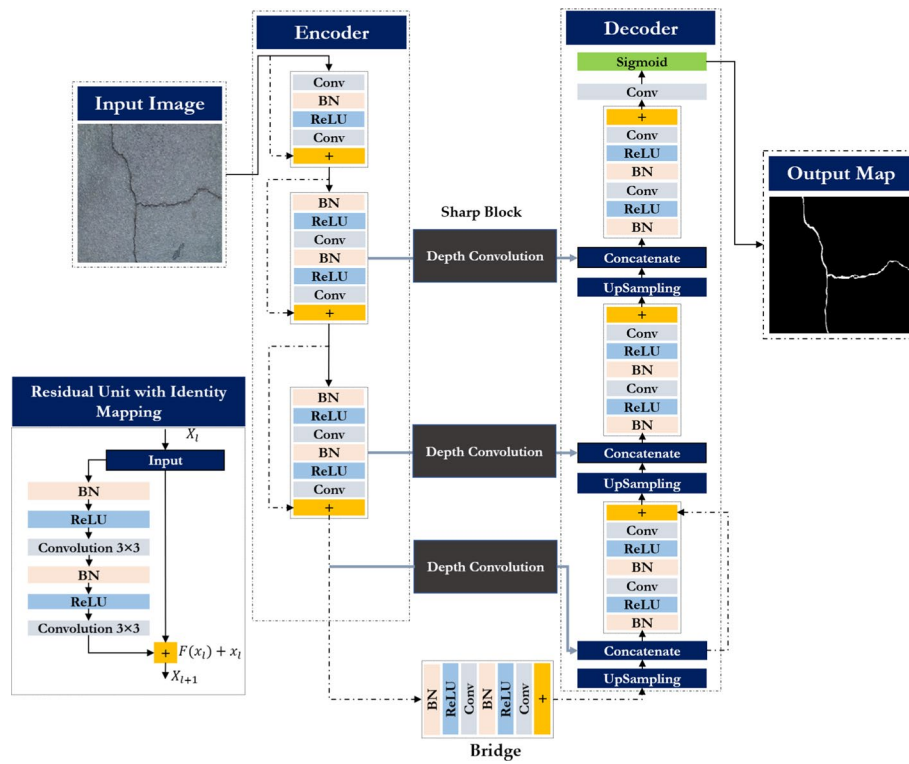


Fig. 7 Residual Sharp U-Net architecture

complex patterns in the input data. Moreover, in the RS-Net architecture, the residual blocks help to preserve important features and gradients and can be represented by Eq. 2.

$$Z_L = ReLU(BN(Z_{L-1} + Y_L)) \tag{2}$$

Where Z_L , BN , and Y_L represent the output feature map of the L^{th} residual block, batch normalization, and the sharpened output map respectively. The output of the last residual block in the RS-Net architecture can be represented in Eq. 3 as.

$$Y = LaplaceConv(Z_N + K') \tag{3}$$

The term Z_N in Eq. 3 represents the last residual block output while K' is the Laplacian filter kernel of the final sharp block. The final output feature map after concatenation with the decoder features are then passed through a softmax activation function to obtain the probability map for semantic segmentation. In the proposed work, the performance of the models is compared under different loss function to determine which loss function works better for crack segmentation task based on the datasets used in the proposed work.

Post processing operations

After performing the crack segmentation, post processing operation are employed to analyze the crack regions. The post processing operations includes skeletonization and crack measurement operations.

Crack skeletonization, measurement and severity assessment

Crack skeletonization simplifies the analysis and interpretation of crack patterns and provides a useful reference for structural health monitoring and maintenance. This is done by representing the topology of cracks using a single pixel i.e., crack skeletonization. By regularly comparing the crack skeleton to new data, changes in the structure's condition can be identified over time, enabling better decisions about repairs and maintenance to be made to prolong the structure's lifespan and prevent further damage. In the proposed work, crack skeletonization is performed using medial axis algorithm [77, 78]. It is possible to calculate the length of cracks when they are converted into a single-pixel-wide representation through the process of skeletonization as depicted in Eq. 4.

$$Crack_{length} = \int_c f(a, b) dl \cong \sum f(a, b) dl \quad (4)$$

In the above equation, the parameters geometric calibration index ($f(a, b)$), a measure used to calibrate the displacements of pixels in crack image and the finite length of the skeleton elements (dl) are used to calculate the length of the cracks in skeletonized images. The proposed method assumes that the input images have no geometric distortion. In such cases, the calculation can be simplified by directly counting the pixels of the skeletons to obtain the length of the cracks without any geometric calibration adjustments. However, the calculation of crack average crack width involves the geometric calibration index, the finite area of the crack elements, and the length of the cracks as show in Eq. 5.

$$Crack_{Average-width} = \frac{\int_s f^2(a, b) dS}{Crack_{length}} \cong \frac{\sum f^2(a, b) dS}{\sum f(a, b) dl} \quad (5)$$

The term $Crack_{length}$ and dS represent the length of the cracks and the finite area of elements of cracks. Based on the crack width, the cracks are classified into large, medium, and hairline cracks. The cracks are classified as either “needs to be repaired”, “medium crack”, or “hairline crack” based on their widths. Cracks with widths greater than 10 pixels are classified as “need to be repaired”, cracks with widths between 5 and 10 pixels are classified as “medium crack”, and cracks with widths between 0 and 5 pixels are classified as “hairline crack”.

Experiments and results

The following section provides a comprehensive overview of the experimental methodology employed to investigate the proposed approach, as well as a detailed analysis of the results obtained.

Experimental setup

In the proposed work, the models are trained using an NVIDIA DGX-1 system, which is a high-performance computing platform that is designed specifically for deep learning research. The system is equipped with dual 20-core Intel XEON E5-2698 v4 2.2 GHz CPUs, 8 Tesla V100 GPUs, 256 GB GPU memory and 40,960 NVIDIA CUDA cores, making it one of the most powerful systems. PyTorch library is used along with Windows 10 and Python 3.8. The models are trained with a batch size of 5, a learning

rate of 0.001, and the Adam optimizer. The performance of the models is evaluated using various evaluation metrics such as Accuracy, Loss, Jaccard, Precision, Recall, mIoU and Dice coef. These metrics are discussed in detail in the next section.

Evaluation metrics

In the crack segmentation task, accuracy compares the proportion of correctly identified pixels (cracks and non-cracks) in the segmentation output to the ground truth as depicted in Eq. 6.

$$Accuracy = \frac{T_P + T_N}{T_P + F_N + T_N + F_P} \quad (6)$$

where T_P , T_N , F_P , and F_N represent the true positive, true negative, false positive, and false negative, respectively. Recall refers to the proportion of correctly identified cracks to all the actual crack (correctly identified) and missed actual cracks (actual crack pixels that are missed by the segmentation algorithm) as shown in Eq. 7.

$$Recall = \frac{T_P}{T_P + F_N} \quad (7)$$

Precision is the proportion of correctly identified cracks to actual cracks (correctly identified) and false cracks (non-crack pixels that are incorrectly identified as cracks) and can be mathematically represented as (Eq. 8).

$$Precision = \frac{T_P}{T_P + F_P} \quad (8)$$

By dividing the intersection of the two segmentation masks by their union, as shown in Eq. 9, the Jaccard Index or Intersection over Union (IoU) metric determines how comparable the predicted and ground truth segmentation masks are.

$$IoU(JaccardIndex) = \frac{T_P}{T_P + F_P + F_N} \quad (9)$$

The *JaccardIndex(IoU)* average for all images in the dataset is determined by using the *mIoU(meanIntersectionoverUnion)* measure as depicted in Eq. 10.

$$mIoU = \frac{IoU_1 + IoU_2 + IoU_3 + \dots + IoU_n}{n} \quad (10)$$

where n represent the types of cracks in the images. The dice coefficient is the overlap between predicted and ground truth segmentation while loss represents the difference between the two masks and can be mathematically represented as (Eq. 12).

$$dice = \frac{2 \times precision \times Recall}{Precision \times Recall} = \frac{2 \times T_P}{2 \times T_P \times F_P \times F_N} \quad (11)$$

In the proposed work two various loss functions named binary cross entropy and binary focal loss has been used and is formulated in Eqs. 9 and 10 as.

$$\text{BinaryEntropyLoss} = \frac{-1 \times \sum (GT \times \log(PM) + (1 - GT) \times \log(1 - PM))}{N} \quad (12)$$

$$\begin{aligned} \text{BinaryFocalLoss} = & -1 \times \sum (GT \times \log(1 - PM)^\beta \times GT \times \log(PM)) \\ & + (1 - \alpha) \times (PM^\beta \times (1 - GT) \times \log(1 - PM))/N \end{aligned}$$

where GT , PM , and N represent the ground truth mask, the predicted mask, and the number of pixels in the image. α is a balancing parameter that controls the contribution of the easy and hard examples to the loss, and β is a focusing parameter and is used to adjust the weight given to each pixel based on its difficulty level. Each metric serves a specific purpose in quantifying different aspects of model performance. Accuracy measures the proportion of correctly classified pixels, providing an overall view of model performance. Precision and Recall evaluate the model's ability to correctly detect positive instances (crack pixels) and avoid false negatives, respectively. The Jaccard Index (Intersection over Union) and Dice Coefficient assess the spatial overlap between predicted and ground truth masks, offering insights into segmentation accuracy and boundary alignment. Mean Intersection over Union (mIoU) provides a holistic measure by considering both the overlap and total area covered by predicted and ground truth masks, essential for comparing segmentation models objectively and guiding their optimization for accurate crack detection in civil infrastructure. To further contextualize this, we have referenced various works that employ accuracy, precision, and recall metrics. These studies illustrate the use of these metrics in similar contexts, supporting the evaluation of our network's performance [79, 80].

Quantitative results of the crack segmentation models

In the proposed work, all the segmentation models are evaluated on two publicly available datasets, D1 [69] and D2 [70], which consist of 11.29K images and 9.49K images with resolutions of 440×440 and 400×400 , respectively. The details of the datasets, including their sources and the number of images, are summarized in Table 1. The data split is as follows: 20% of the data is reserved for testing, and the remaining 80% is split into 80% for training and 20% for validation. This ensures that the model is tested on data samples that were not used in the training and validation phases, providing a robust evaluation of its performance.

The number of parameters and training time of each model is shown in Table 2. The U-Net and sharp U-Net model consists of 7.76 million parameters and the training time for both models is around 100–104 min. On the other hand, the Residual U-Net and the proposed RS-Net architecture have fewer parameters i.e., 4.73 million while the training time is 170 min. The increase in the training time is due to an increase in the number of layers in the residual block of the model. The models used in the proposed work are trained using Binary Cross Entropy (BCE) and Binary Focal Loss (BFL) to evaluate their performance on various loss functions. The result of the models using dataset D1 is shown in Table 3.

All the models achieved promising accuracy (> 0.97) indicating that they can perform crack segmentation task effectively. The U-Net model achieved relatively good accuracy (> 0.97), precision (> 0.72) and recall ($> 61\%$) however, the Jaccard, mIoU

Table 1 Details of Dataset D1 and D2 used in the proposed work

Dataset name	D1	No of images	D2	No of images
CFD [73]	✓	117	✓	118
Crack500 [72]	✓	3369	✓	3126
DeepCrack [81]	✓	521	✓	443
Eugen Muller [72]	✓	55	–	–
Masonry [82]	–	–	✓	240
Ceramic [83]	–	–	✓	100
SDNET2018 [84]	–	–	✓	1411
Rissbilder [72]	✓	3822	✓	2736
Sylvie-Chambon [85]	✓	185	–	–
GAPS384 [86]	✓	383	✓	383
CrackTree200 [86]	✓	206	✓	175
Forest [87]	✓	118	–	–
Volker dataset [88]	✓	990	✓	427
Non-Crack [72]	✓	1411	–	–
Total Images		11,176		9159

Table 2 Models parameters and training time

Model	Number of parameters	Training time
U-NET	7.76	1:40
Sharp U-NET	7.76	1:44
Residual U-NET	4.73	2:51
Residual Sharp U-NET	4.73	2:50

Table 3 Results of the models on Dataset D1 (11.2K images)

Loss function	Accuracy	Loss	Jaccard	Precision	Recall	mIoU	Dice coefficient
U-Net							
BCE	0.9785	0.0645	0.4158	0.7206	0.6141	0.4938	0.5855
BFL	0.9785	0.0181	0.2184	0.7203	0.6128	0.4928	0.3572
Sharp U-Net							
BCE	0.9807	0.0638	0.4499	0.7682	0.6300	0.5195	0.6148
BFL	0.9789	0.0208	0.2305	0.7362	0.6025	0.4885	0.3698
Residual U-Net							
BCE	0.9808	0.0728	0.4494	0.7859	0.6087	0.5129	0.6137
BFL	0.9810	0.0208	0.2977	0.7879	0.6123	0.5152	0.4537
RS-Net							
BCE	0.9892	0.0644	0.4681	0.7883	0.6486	0.5409	0.6331
BFL	0.9842	0.0161	0.2910	0.7899	0.6380	0.5404	0.4657

Bold values represents the highest values of the evaluation metrics

BCE: Binary Cross Entropy; BFL: Binary Focal Loss; Acc: Accuracy

and dice coefficient values are lower which indicates the limited ability of the model to capture the shape and details of the crack in the segmentation task. When compared to the U-Net model, the Sharp U-Net model has higher Jaccard, mIoU, and Dice coefficient values, indicating better performance in capturing the shape and

features of the fractures. It also has a greater precision and recall (> 0.63), meaning that it can detect more crack pixels properly. However, the performance improvement is minimum when compared to the U-Net model. The Residual U-Net model has higher accuracy and precision (> 0.78) than the other models, but lower recall and Jaccard coefficient. This implies that the model can properly identify cracks but may overlook some crack locations.

The RS-Net model achieved the highest values for most of the evaluation metrics, including jaccard (0.468), mIoU (0.540), dice coefficient (0.633), and accuracy ($>98.4\%$). This indicates that it can accurately segment the cracks, capturing their shape and details effectively. The model also has good precision and recall (>0.63), indicating its ability to correctly identify crack pixels. While the quantitative results show that the RS-Net model trained using binary cross entropy loss function achieves the highest accuracy, Jaccard index, precision, recall, mIoU, and Dice coefficient for crack segmentation, these results are further validated by the superior segmentation results observed in visual inspection of the model's output below. The training and validation accuracy and loss curves of the RS-Net model are depicted in Fig. 8, which demonstrate less divergence between the two curves, indicating that the model is not subjected to overfitting.

The results achieved by the models employed in the proposed research, when tested on dataset 2, have been collated in Table 4. Using the BCE loss function, the U-Net model achieved a high accuracy of 0.9824 and a low loss of 0.0545. The model also performed well in terms of jaccard and mIoU, although with a lower dice coefficient when compared to other models. BFL loss obtained a very low loss of 0.0171 but at the expense of large reductions in jaccard, precision, recall, and dice coefficient. The Sharp U-Net model has reduced jaccard, precision, recall, mIoU, and dice coefficient values using BCE loss function while maintaining comparable accuracy and loss. The model achieved a very low loss of 0.0168 with decent jaccard, precision, recall, and mIoU values using BF loss function, but the model's dice coefficient was relatively lower compared to other models. It is worth mentioning that the performance difference between the U-Net and Sharp U-Net is very small. The Residual U-Net model performed better than the U-Net and Sharp U-Net models in terms of all evaluation metrics for both loss functions, i.e., BCE and BFL.

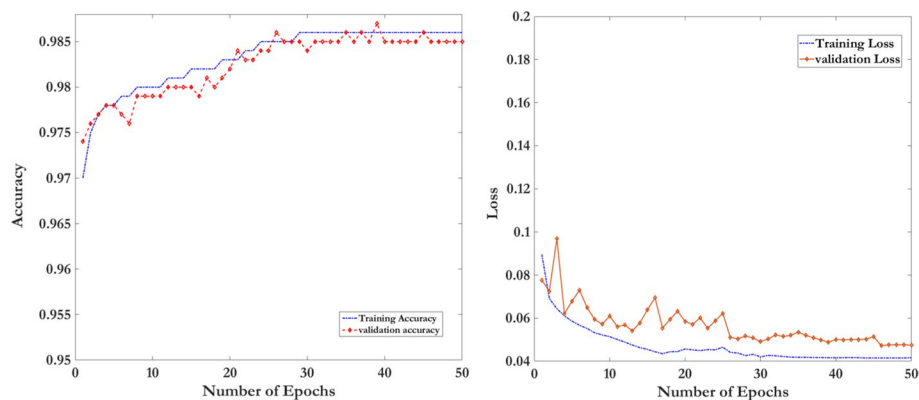


Fig. 8 Training and validation accuracy and loss curves of the RS-Net model

Table 4 Results of the models on Dataset D2 (9K images)

Loss function	Accuracy	Loss	Jaccard	Precision	Recall	mIoU	Dice coefficient
U-Net							
BCE	0.9824	0.0545	0.4300	0.7246	0.6290	0.5031	0.5992
BFL	0.9786	0.0171	0.0987	0.7300	0.4092	0.3489	0.1789
Sharp U-Net							
BCE	0.9821	0.0523	0.3858	0.7301	0.6019	0.4792	0.5500
BFL	0.9844	0.0168	0.2522	0.7628	0.6636	0.5311	0.3946
Residual U-Net							
BCE	0.9849	0.0574	0.4749	0.7858	0.6516	0.5355	0.6363
BFL	0.9849	0.0162	0.2964	0.7903	0.6456	0.5321	0.4498
RS-Net							
BCE	0.9853	0.0471	0.4995	0.7980	0.6695	0.5459	0.6518
BFL	0.9850	0.0184	0.3401	0.7837	0.6603	0.5414	0.5000

Bold values indicate the highest values of the evaluation metrics

BCE: Binary Cross Entropy; BFL: Binary Focal Loss; Acc: Accuracy

The Residual U-Net has a higher accuracy of 0.9849, jaccard of 0.4749, mIoU of 0.5355, and dice coefficient of 0.6363, indicating better overall segmentation performance. Furthermore, Residual U-Net has greater precision (0.7858) and recall (0.6516) values, showing that it can catch both foreground (crack) and background (non-crack) regions in the images more effectively during the crack segmentation. RSNet achieved the highest accuracy score of 0.9853 for BCE loss and 0.9850 for BFL loss. It also achieved the highest Jaccard score of 0.4995, Precision of 0.7980, Recall of 0.6695, mIoU of 0.5459, and Dice Coefficient of 0.6518 for BCE loss function, indicating better segmentation performance compared to the other models. While the numerical improvements in accuracy or mIoU metrics over comparable models may appear modest, our emphasis extends beyond mere metric superiority. The RS-Net architecture incorporates novel features like residual connections and sharp blocks, specifically tailored to enhance crack segmentation efficacy in civil infrastructure. The finding of datasets D1 and D2 shows that the performance of all the models is largely similar however, the overall performance of D1 is better than D2 which may be owing to the increased number of samples in the D1. The fact that RS-Net outperformed the other models across all evaluation criteria shows RS-Net model can capture both high-level and low-level features and preserve the spatial details, resulting in improved segmentation performance.

Qualitative results of the crack segmentation models

Qualitative analysis was performed on four segmentation models using test data that was not utilized during the training and validation phase. Sample images of the RS-Net model on dataset D1 are shown in Figs. 9 and 10, with yellow and red boxes representing False positives and False negatives, respectively.

The U-Net model showed many false positives and false negatives, leading to over-segmentation and under-segmentation. It can also be seen from Figs. 9 and 10 that the incorporation of the sharp block into the U-Net architecture reduced the number of FPs and FNs, resulting in fewer over-segmented regions when using the BCE loss

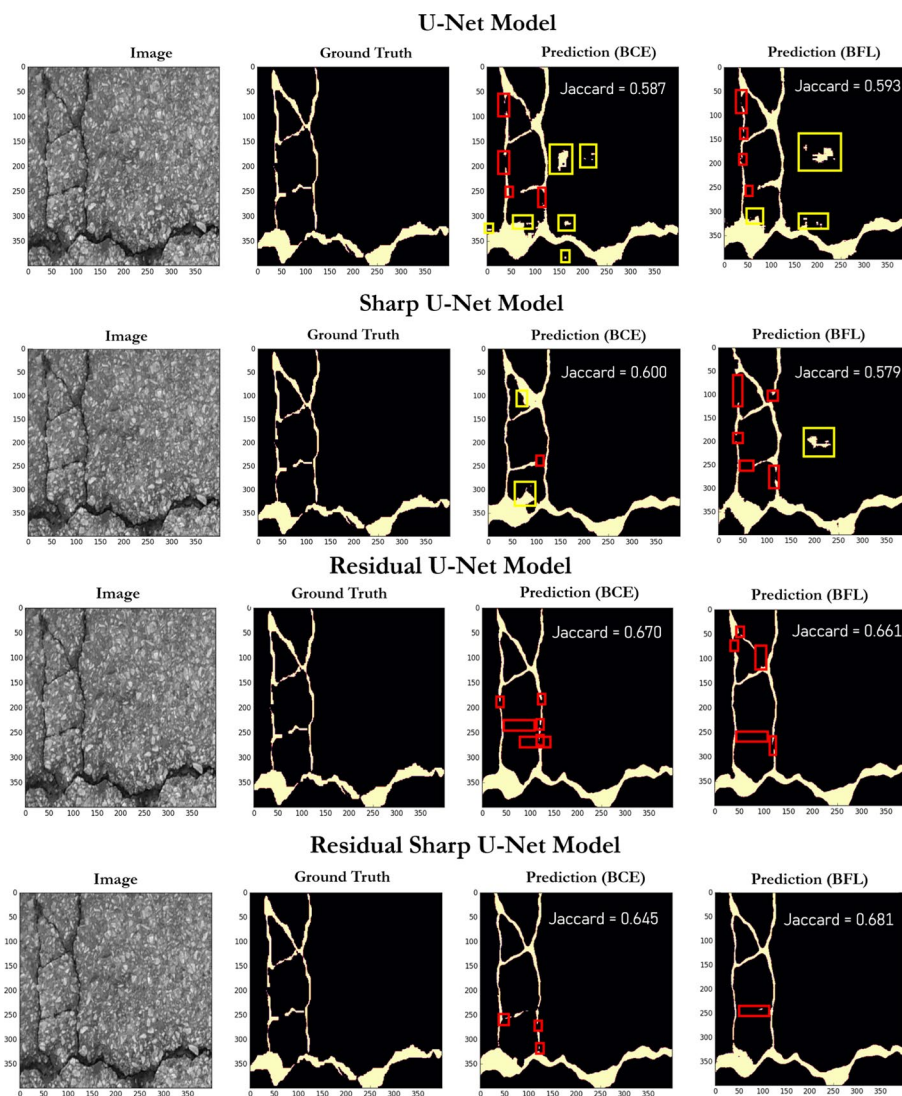


Fig. 9 Representation of the crack segmentation results of all models (11.2K dataset)

function. The Residual U-Net model addressed the problem of FPs, but still exhibited some FNs, as it failed to predict some of the crack regions. In contrast, the RS-Net model outperformed the other models in terms of both FP and FN regions, indicating its ability to capture fine crack features. However, further improvement is still needed to mitigate the small number of FNs that lead to under-segmentation of the predicted map. Fig. 11 shows the qualitative analysis of the results of the models on 9K datasets. The results show that the U-Net model continued to exhibit significant false positives and false negatives, resulting in over-segmentation and under-segmentation, respectively. Sharp U-Net, Residual U-Net, and RS-Net, among others, demonstrated varied degrees of success in reducing these concerns. The results from the 9K dataset corroborated the findings from the 11.2K dataset. Overall, the experimental results

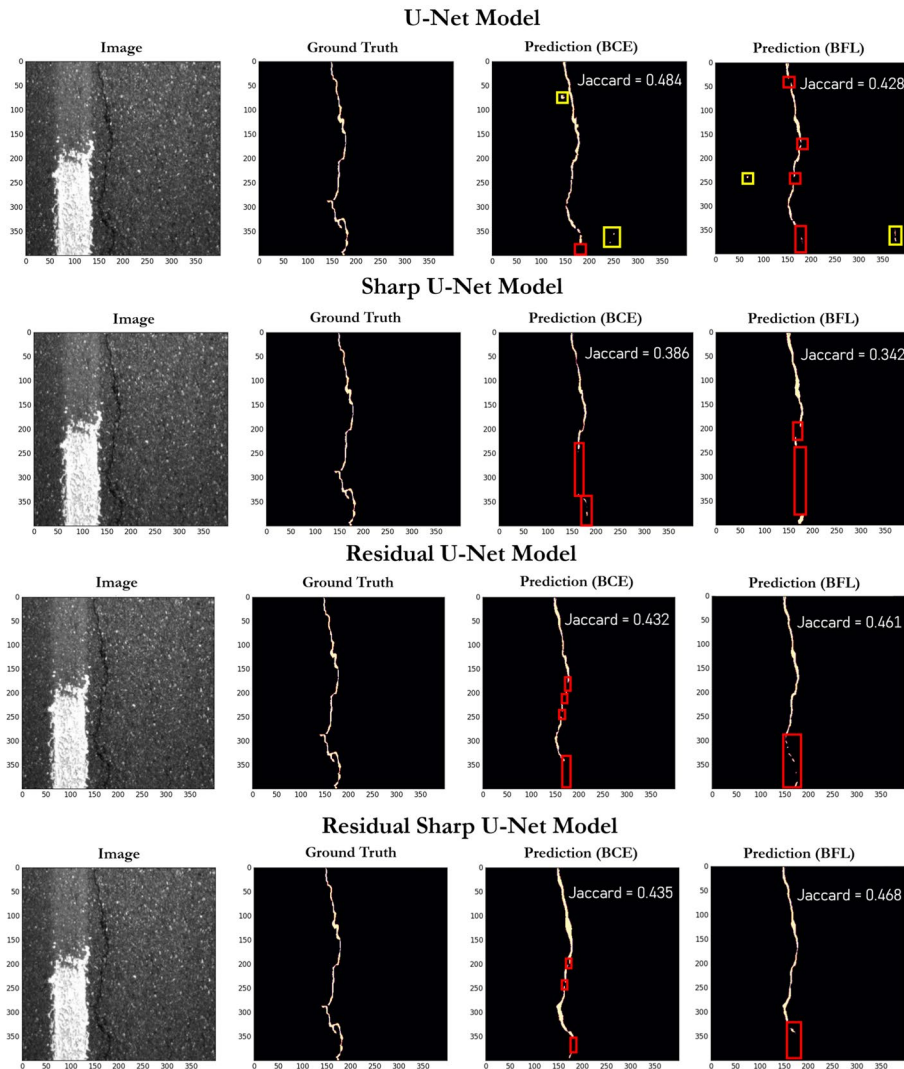


Fig. 10 Representation of the crack segmentation results of all models (11.2K dataset)

demonstrate that the RS-Net model trained on the 11.2K dataset has a strong capability for crack segmentation

Crack severity assessment of the RS-Net Model

The severity assessment of the cracks is performed on the segmentation maps obtained as a result from the RS-Net architecture using various morphological operations such as crack max-width, length, and mean width could be calculated as depicted in Fig. 12. These morphological features help the authorities in the assessment of the existing civil structures. The experimental result images are shown from top to bottom: original image, predicted image, crack segmentation image, crack skeleton image, and cracked surface image with severity labels. With lighter colors denoting bigger cracks, the medial-axis approach [77, 78] yields the crack skeleton, which is

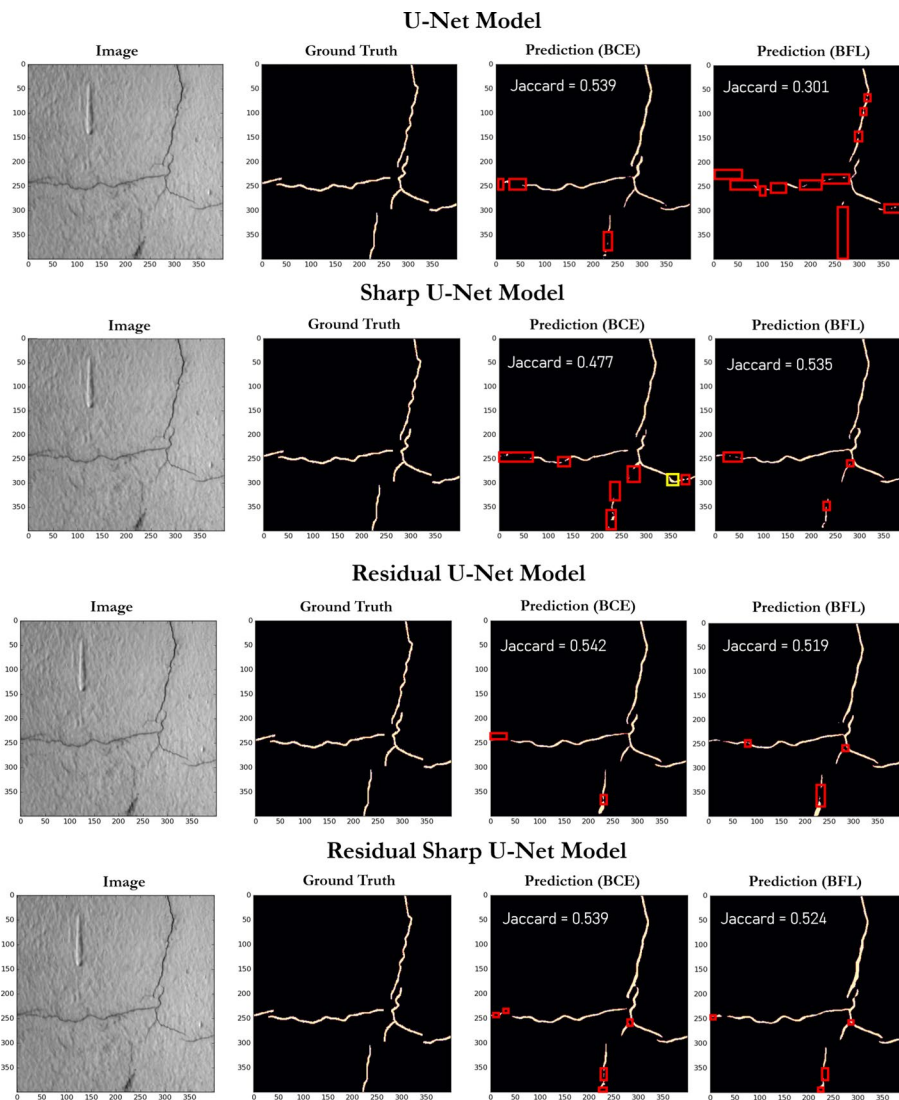


Fig. 11 Representation of the crack segmentation results of all models (9K dataset)

then displayed in various colors. The images in the fourth row of Fig. 12 appropriately represented the skeleton images of the original images in the first row.

The crack length is overestimated in the last row of the first 2 images, which might be owing to the use of morphological operations, notably opening and closure operations, to fill the holes and eliminate the single-pixel cracks. These techniques can occasionally fill in pixels that are not part of the crack or eliminate pixels that are part of the crack, increasing the number of predicted crack pixels. In the proposed work, the crack pixels were further classified as severe, moderate, or hairline based on their width, a hairline crack is defined as one with a width between 1 and 5 pixels, whereas a medium crack has a width larger than 5 but less than 10 pixels. Severe cracks are defined as those with a width of more than 10 pixels. The severity of the crack and the width of the pixels is used to categorize them. The 5th row of Fig. 12 depicts this classification.

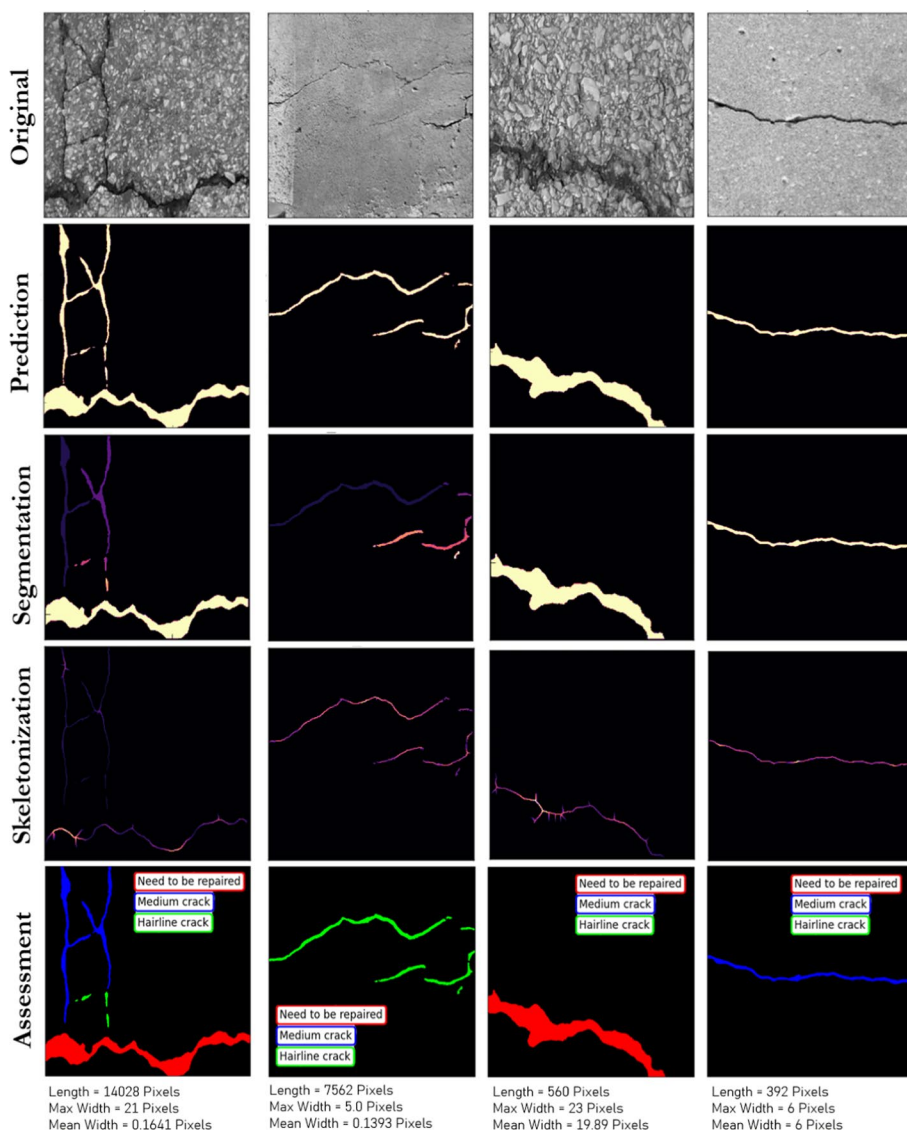


Fig. 12 Representation of the original image, predicted image, crack segmentation image, crack skeleton image, and cracked surface image with severity labels (11K)

Table 5 Comparison of the proposed RS-Net with SOTA algorithms

Model	Accuracy	Loss	Jaccard	Precision	Recall	mIoU	Dice coefficient
U-Net [89]	0.943-0.970	0.05-0.15	NA	NA	NA	NA	NA
VGG-UNet [90]	0.9706	0.4364	NA	NA	NA	0.7035	NA
DeepLabV3+ [91]	0.5950	NA	NA	0.6870	0.8150	NA	NA
DeepLabV3 [91]	0.5950	NA	NA	0.6870	0.8150	NA	NA
DA-Net [91]	0.7570	NA	NA	0.8740	0.8490	NA	NA
RS-Net	0.9892	0.0644	0.4681	0.7883	0.6486	0.5409	0.6331

Comparison with State-of-the-art algorithms

In the proposed work, a comparative analysis with state-of-the-art algorithms (SOTA) trained on D2 (11.2K) dataset to assess the performance of the RS-Net algorithm. The dataset D1 has been relatively underutilized in the literature, limiting the ability to conduct a comprehensive comparison with state-of-the-art (SOTA) methods. The comparison table of the algorithms using D2 is depicted in Table 5.

It is evident from the table that a majority of studies utilizing the same dataset have not consistently reported all evaluation metrics. This inconsistency in reporting metrics across studies poses challenges in conducting a comprehensive and fair comparison of algorithms. Nonetheless, the comparison results show that the proposed RS-Net algorithm has higher accuracy (0.989) and lower loss (0.064) than the other algorithms, making it a promising approach for crack detection. However, it has lower precision, recall, and mIoU compared to some of the other algorithms. The high Dice coefficient indicates that the RS-Net algorithm can produce more accurate segmentation masks. Overall, the RS-Net algorithm shows great potential for crack detection in civil infrastructure, but further improvements are needed in some areas.

Pixels to physical length conversion

Previous research has investigated how to convert crack dimensions from pixel units to engineering units across different camera orientations. A significant challenge in-camera image-processing systems for measuring crack width is accurately determining the conversion factor [92, 93]. In the proposed work, pixel-to-physical length conversion is performed by capturing images at two specific distances, 16 cm and 40 cm from the ground surface. Notably, the high-resolution images at 16 cm (8192 × 6144 pixels) are specified for accurate analysis. The conversion factors presented in Table 6 define the relationships between pixels and millimeters. For the 16 cm distance, 1 pixel is approximately equivalent to 0.0268 mm in width and 0.0292 mm in height.

At the 40 cm distance, 1 pixel corresponds to about 0.0732 mm in width and 0.0716 mm in height. These values are essential for translating digital measurements into real-world values, ensuring that assessments of civil engineering structures are conducted with precision and real-world relevance, ultimately facilitating more accurate infrastructure maintenance. A comparison of the calculated crack width, obtained through pixel-to-physical length conversion, with the actual crack width in a test image as depicted in Fig. 13. The linear crack’s maximum width was calculated, and after applying the conversion factor, the physical width of the crack was determined to be 12.21 mm. When manually measuring the crack on a scale, the actual crack width was approximately

Table 6 Pixel-to-physical length conversion factors for different distances and dimensions

Distance (mm)	Image resolution	Dimension	Conversion factor (mm)
16	8192 × 6144	Width	0.0268
42	8192 × 6144	Width	0.0732
16	8192 × 6144	Height	0.0292
42	8192 × 6144	Height	0.0716

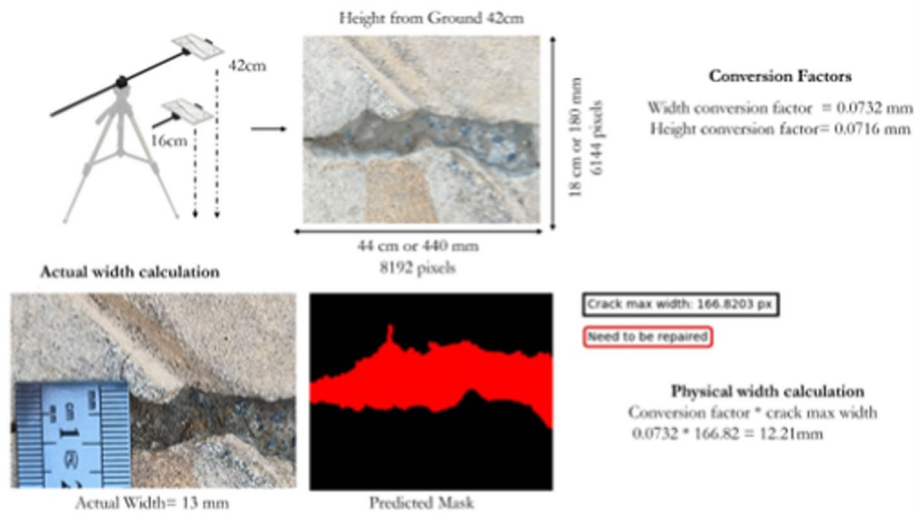


Fig. 13 Crack width and severity estimation for linear cracking

12 mm. The absolute error between the calculated and actual widths is approximately 0.21 mm. To ensure a fair comparison, a conservative approach considers this error as 1 mm, reflecting the practical nature of civil engineering assessments.

Discussion

The study presents a novel crack architecture i.e., RS-Net for crack segmentation in civil infrastructure. The proposed architecture is trained on two publicly available dataset D1 and D2 consisting of 11.2 and 9K images. The proposed architecture is compared with U-Net, Residual U-Net and Sharp U-Net based on various loss functions and several metrics such as accuracy, loss, Jaccard, precision, recall, mIoU, and Dice coefficient. The experimental results showed that all four models achieved promising accuracy (> 0.97), which is a good indication that the models can effectively perform the crack segmentation task however, RS-Net consistently outperformed the other models in both loss functions, BCE and BFL. The results also indicate that increasing the number of samples in the training set without sufficient variation among the data samples does not enhance the performance of the models, as observed in the comparison between datasets D1 and D2. However, despite the differences in dataset sizes, both datasets showed the Residual U-Net and RS-Net models to be effective for crack segmentation. Moreover, the results in Table 3 and 4 indicates that the Sharp U-Net and RS-Net models achieved better performance when trained using the BCE loss function. The utilization of the BCE loss function demonstrated the RS-Net model's ability to optimize its parameters effectively for achieving accurate segmentation outcomes. This is evident from the mIoU values of 0.5404 and 0.5459 achieved by the model on Dataset D2 and D1, respectively.

On dataset D1, the U-Net model shows promising accuracy of (0.978), precision (0.7206), and recall (0.6141). However, the values for Jaccard (0.415), mIoU (0.493), and Dice coefficient (0.585) were lower, suggesting that the model's capability to accurately capture the intricate shapes and details of the cracks in the segmentation task was limited. On the other hand, the Sharp U-Net model performed better in capturing the shape

and features of the cracks, as it had higher values for Jaccard (0.449), mIoU (0.519), and Dice coefficient (0.614). The comparison between U-Net and Sharp U-Net showed that the latter performed slightly better in terms of Jaccard and Dice coefficient metrics which is due to the integration of the sharp blocks in the skip connection of the U-Net model. Residual U-Net also showed promising performance, especially in terms of mIoU (0.5129) and Dice coefficient (0.6137) metrics, indicating that the use of residual connections can also improve help the model to capture more contextual information and produce better segmentation results. The RS U-Net combines the benefits of both sharp and residual U-Net and outperforms the other model and achieved the highest values for most of the evaluation metrics i.e., Jaccard (0.468), mIoU(0.540) and Dice coefficient (0.633).

Similarly, on dataset D2, the RS-Net architecture surpass the other models in term of accuracy (0.958), loss (0.0471), Jaccard (0.499), precision (0.798), recall (0.669), mIoU (0.545) and Dice Coefficient (0.652). Additionally, the Residual U-Net and RS-Net models had fewer parameters than the other models and took longer to train, but their performance was comparable to the U-Net and Sharp U-Net models. This suggests that using residual blocks in the models could be an effective way to reduce the number of parameters while maintaining good performance. Overall, the study highlights that incorporating advanced techniques such as residual connections and sharp blocks to the U-Net can lead to better segmentation results, and further research can explore other techniques for enhancing the U-Net architecture by integrating various loss functions and modules. Moreover, it is also clear from the qualitative results that the RS-Net architecture can be integrated with various morphological operations for assessing the severity of cracks using segmentation maps. The crack pixels based on their width can be used to categorize the cracks in the structure into severe, moderate, or hairline crack labels which are then used to assess the existing civil structures. However, more improvement is needed to eliminate the overestimation of crack length in the morphological operations.

Conclusion

From the above discussion, it can be concluded that RS-Net architecture proves effective for crack segmentation in civil structures by incorporating advanced techniques such as residual connections and sharp blocks within the U-Net framework. RS-Net with BCE shows superior accuracy on both datasets, highlighting its robust performance. It can also be concluded that the integration of RS-Net with morphological operations enhances the assessment of crack severity in civil infrastructures. This study provides valuable insights into developing precise and efficient crack segmentation models for civil infrastructure. However, limitations include the need for dataset expansion with greater sample variation to further enhance model accuracy, and the elimination of overestimations in crack length during morphological operations.

Future work

In future work, we aim to enhance the performance of the RS-Net architecture by incorporating a larger and more diverse dataset while also focusing on improving its generalizability to address longer crack lengths caused by discontinuities in the crack

pixels. Additionally we will explore additional loss functions to further refine our model's performance across various segmentation challenges.

Author contributions

The authors L.A., H.A-J, M.S., W.K., and F.A-N. collectively contributed to the conceptualization, methodology, software development, validation, and data curation of this study. L.A., H.A-J and F.A-N. performed the formal analysis. The original draft of the manuscript was prepared by L.A. and later reviewed and edited by all authors. M.S. and W.K. were responsible for visualization. The project was supervised by H.A-J. and F.A-N. with project administration handled by H.A-J. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 August 2023 Accepted: 10 August 2024

Published online: 17 August 2024

References

1. Graybeal BA, Phares BM, Rolander DD, Moore M, Washer G. Visual inspection of highway bridges. *J Nondestruct Eval.* 2002;21(3):67–83.
2. Phares BM, Washer GA, Rolander DD, Graybeal BA, Moore M. Routine highway bridge inspection condition documentation accuracy and reliability. *J Bridge Eng.* 2004;9(4):403–13.
3. Alampalli S, Rehm KC. Impact of i-35w bridge failure on state transportation agency bridge inspection and evaluation programs. In: *Structures Congress 2011*, 2011;1019–1026
4. Asakura T, Kojima Y. Tunnel maintenance in Japan. *Tunnell Underground Space Technol.* 2003;18(2–3):161–9.
5. Hasan U, Whyte A, Al Jassmi H. Life-cycle asset management in residential developments building on transport system critical attributes via a data-mining algorithm. *Buildings.* 2018;9(1):1.
6. Baig F, Ali L, Faiz MA, Chen H, Sherif M. How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment? *J Hydrol.* 2024;633:131040.
7. Philip B, Xu Z, AlJassmi H, Zhang Q, Ali L. Asenn: attention-based selective embedding neural networks for road distress prediction. *J Big Data.* 2023;10(1):164.
8. Sinha SK, Fieguth PW. Morphological segmentation and classification of underground pipe images. *Mach Vision Appl.* 2006;17:21–31.
9. Sinha SK, Fieguth PW. Automated detection of cracks in buried concrete pipe images. *Autom Constr.* 2006;15(1):58–72.
10. Chambon S, Subirats P, Dumoulin J. Introduction of a wavelet transform based on 2d matched filter in a markov random field for fine structure extraction: application on road crack detection. In: *Image Processing: Machine Vision Applications II*, 2009;7251:87–98. SPIE.
11. Fujita Y, Hamamoto Y. A robust automatic crack detection method from noisy concrete surfaces. *Mach Vision Appl.* 2011;22:245–54.
12. Hoang N-D. Detection of surface crack in building structures using image processing technique with an improved otsu method for image thresholding. *Adv Civil Eng.* 2018;2018:3924120.
13. Kamaliardakani M, Sun L, Ardakani MK. Sealed-crack detection algorithm using heuristic thresholding approach. *J Comput Civil Eng.* 2016;30(1):04014110.
14. Abdel-Qader I, Abudayyeh O, Kelly ME. Analysis of edge-detection techniques for crack identification in bridges. *J Comput Civil Eng.* 2003;17(4):255–63.
15. Chisholm T, Lins R, Givigi S. Fpga-based design for real-time crack detection based on particle filter. *IEEE Trans Ind Inform.* 2019;16(9):5703–11.
16. Ali L, Harous S, Zaki N, Khan W, Alnajjar F, Al Jassmi H. Performance evaluation of different algorithms for crack detection in concrete structures. In: *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2021:53–58. IEEE.

17. Ali L, Sallabi F, Khan W, Alnajjar F, Aljassmi H. A deep learning-based multi-model ensemble method for crack detection in concrete structures. In: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 2021;38:410–418. IAARC Publications.
18. Silva WRld, Lucena DSd. Concrete cracks detection based on deep learning image classification. In: Proceedings, 2018;2:489. MDPI.
19. Flah M, Suleiman AR, Nehdi ML. Classification and quantification of cracks in concrete structures using deep learning image-based techniques. *Cement Concrete Compos.* 2020;114:103781.
20. Ali L, Alnajjar F, Jassmi HA, Gocho M, Khan W, Serhani MA. Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures. *Sensors.* 2021;21(5):1688.
21. Ali L, Valappil NK, Kareem DNA, John MJ, Al Jassmi H. Pavement crack detection and localization using convolutional neural networks (cnns). In: 2019 International Conference on Digitization (ICD), 2019;217–221. IEEE.
22. Ali R, Chuah JH, Talip MSA, Mokhtar N, Shoaib MA. Structural crack detection using deep convolutional neural networks. *Autom Constr.* 2022;133:103989.
23. Ali L, Alnajjar F, Zaki N, Aljassmi H. Pavement crack detection by convolutional adaboost architecture. In: 8th Zero Energy Mass Custom Home International Conference, ZEMCH 2021, 2021;383–394. ZEMCH Network.
24. Liu F, Liu J, Wang L. Asphalt pavement crack detection based on convolutional neural network and infrared thermography. *IEEE Trans Intell Transp Syst.* 2022;23(11):22145–55.
25. Chaiyasarn K, Khan W, et al. Damage detection and localization in masonry structure using faster region convolutional networks. *GEOMATE J.* 2019;17(59):98–105.
26. Ma D, Fang H, Wang N, Zhang C, Dong J, Hu H. Automatic detection and counting system for pavement cracks based on pccgan and yolo-mf. *IEEE Trans Intell Transp Syst.* 2022;23(11):22166–78.
27. Deng J, Lu Y, Lee VC-S. Imaging-based crack detection on concrete surfaces using you only look once network. *Struct Health Monit.* 2021;20(2):484–99.
28. Pan Y, Zhang G, Zhang L. A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. *Autom Constr.* 2020;119:103357.
29. Dong Z, Wang J, Cui B, Wang D, Wang X. Patch-based weakly supervised semantic segmentation network for crack detection. *Constr Build Mater.* 2020;258:120291.
30. Zhang L, Shen J, Zhu B. A research on an improved unet-based concrete crack detection algorithm. *Struct Health Monit.* 2021;20(4):1864–79.
31. Ye X-W, Jin T, Chen P-Y. Structural crack detection using deep learning-based fully convolutional networks. *Adv Struct Eng.* 2019;22(16):3412–19.
32. Xue Y, Li Y. A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. *Comput Aided Civil Infrastruct Eng.* 2018;33(8):638–54.
33. Zheng X, Zhang S, Li X, Li G, Li X. Lightweight bridge crack detection method based on segnet and bottleneck depth-separable convolution with residuals. *IEEE Access.* 2021;9:161649–68.
34. Nasiruddin Khiliji T, Lopes Amaral Loures L, Rezazadeh Azar E. Distress recognition in unpaved roads using unmanned aerial systems and deep learning segmentation. *J Comput Civil Eng.* 2021;35(2):04020061.
35. Ronneberger O, Fischer P, Brox T. 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;234–241. Springer.
36. Cheng J, Xiong W, Chen W, Gu Y, Li Y. Pixel-level crack detection using u-net. In: TENCON 2018–2018 IEEE Region 10 Conference, 2018;0462–0466. IEEE.
37. Jenkins MD, Carr TA, Iglesias MI, Buggy T, Morison G. A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks. In: 2018 26th European Signal Processing Conference (EUSIPCO), 2018;2120–2124. IEEE.
38. König J, Jenkins MD, Barrie P, Mannion M, Morison G. A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating. In: 2019 IEEE International Conference on Image Processing (ICIP), 2019;1460–1464. IEEE.
39. König J, Jenkins MD, Barrie P, Mannion M, Morison G. Segmentation of surface cracks based on a fully convolutional neural network and gated scale pooling. In: 2019 27th European Signal Processing Conference (EUSIPCO), 2019;1–5. IEEE.
40. Lau SL, Chong EK, Yang X, Wang X. Automated pavement crack segmentation using u-net-based convolutional neural network. *IEEE Access.* 2020;8:114892–9.
41. Wang W, Su C. Convolutional neural network-based pavement crack segmentation using pyramid attention network. *IEEE Access.* 2020;8:206548–58.
42. Liu F, Wang L. Unet-based model for crack detection integrating visual explanations. *Constr Build Mater.* 2022;322:126265.
43. König J, Jenkins MD, Mannion M, Barrie P, Morison G. Optimized deep encoder-decoder methods for crack segmentation. *Digit Signal Process.* 2021;108:102907.
44. Ghosh S, Singh S, Maity A, Maity HK. Crackweb: a modified u-net based segmentation architecture for crack detection. In: IOP Conference Series: Materials Science and Engineering, 2021;1080:012002. IOP Publishing.
45. Huyen J, Ma T, Li W, Yang H, Xu Z. Pixelwise asphalt concrete pavement crack detection via deep learning-based semantic segmentation method. *Struct Control Health Monit.* 2022;29(8):2974.
46. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X. Residual attention network for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2017;3156–3164.
47. Yang Y, Zhao Z, Su L, Zhou Y, Li H. Research on pavement crack detection algorithm based on deep residual unet neural network. In: Journal of Physics: Conference Series, 2022;2278:012020. IOP Publishing.
48. Yu G, Dong J, Wang Y, Zhou X. Ruc-net: a residual-unet-based convolutional neural network for pixel-level pavement crack segmentation. *Sensors.* 2022;23(1):53.
49. Liu D, Xu M, Li Z, He Y, Zheng L, Xue P, Wu X. A multi-scale residual encoding network for concrete crack segmentation. *J Intell Fuzzy Syst.* 2024;46(1):1379–92.

50. Wu Z, Lu T, Zhang Y, Wang B, Zhao X. Crack detecting by recursive attention u-net. In: 2020 3rd International Conference on Robotics, Control and Automation Engineering (RCAE), 2020;103–107. IEEE.
51. Augustauskas R, Lipnickas A. Improved pixel-level pavement-defect segmentation using a deep autoencoder. *Sensors*. 2020;20(9):2557.
52. Mohan S, Bhattacharya S, Ghosh S. Attention w-net: Improved skip connections for better representations. In: 2022 26th International Conference on Pattern Recognition (ICPR), 2022;217–222. IEEE.
53. Cui X, Wang Q, Dai J, Xue Y, Duan Y. Intelligent crack detection based on attention mechanism in convolution neural network. *Adv Struct Eng*. 2021;24(9):1859–68.
54. Yu C, Du J, Li M, Li Y, Li W. An improved u-net model for concrete crack detection. *Mach Learn Appl*. 2022;10:100436.
55. Wang J, Liu F, Yang W, Xu G, Tao Z. Pavement crack detection using attention u-net with multiple sources. In: Pattern Recognition and Computer Vision: Third Chinese Conference, PRCV 2020, Nanjing, China, October 16–18, 2020, Proceedings, Part II 3, 2020;664–672. Springer.
56. Xiang X, Zhang Y, El Saddik A. Pavement crack detection network based on pyramid structure and attention mechanism. *IET Image Process*. 2020;14(8):1580–6.
57. Zou Q, Zhang Z, Li Q, Qi X, Wang Q, Wang S. Deepcrack: learning hierarchical convolutional features for crack detection. *IEEE Trans Image Process*. 2018;28(3):1498–512.
58. Liu Y, Yao J, Lu X, Xie R, Li L. Deepcrack: a deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*. 2019;338:139–53.
59. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: a nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, 2018;3–11. Springer.
60. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen Y-W, Wu J. 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;1055–1059. IEEE.
61. Zhang J, Jin Y, Xu J, Xu X, Zhang Y. Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. *arXiv preprint arXiv:1812.00352* 2018.
62. Khaledyan D, Marini TJ, Baran MT, O'Connell A, Parker K. Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: sharp attention unet. *PLoS ONE*. 2023;18(12):0289195.
63. Basu A, Deb M, Das A, Dhal KG. Information added u-net with sharp block for nucleus segmentation of histopathology images. *Opt Memory Neural Netw*. 2023;32(4):318–30.
64. Senapati P, Basu A, Deb M, Dhal KG. Sharp dense u-net: an enhanced dense u-net architecture for nucleus segmentation. *Int J Mach Learn Cybern*. 2024;15(6):2079–94.
65. Xie K, Yang Y, Pagnucco M, Song Y. Electron microscope image registration using laplacian sharpening transformer u-net. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022;310–319. Springer.
66. Qian L, Zhou X, Li Y, Hu Z. Unet#: a unet-like redesigning skip connections for medical image segmentation. *arXiv preprint arXiv:2205.11759* 2022.
67. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging*. 2019;39(6):1856–67.
68. Zioulis N, Albanis G, Drakoulis P, Alvarez F, Zarpalas D, Daras P. Hybrid skip: a biologically inspired skip connection for the unet architecture. *IEEE Access*. 2022;10:53928–39.
69. Liu K, Han X, Chen BM. Deep learning based automatic crack detection and segmentation for unmanned aerial vehicle inspections. In: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2019;381–387. IEEE.
70. Kulkarni S, Singh S, Balakrishnan D, Sharma S, Devunuri S, Korlapati SCR. Crackseg9k: a collection and benchmark for crack segmentation datasets and frameworks. In: European Conference on Computer Vision, 2022;179–195. Springer.
71. Zou Q, Cao Y, Li Q, Mao Q, Wang S. Cracktree: automatic crack detection from pavement images. *Pattern Recognit Lett*. 2012;33(3):227–38.
72. Yang F, Zhang L, Yu S, Prokhorov D, Mei X, Ling H. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Trans Intell Transp Syst*. 2019;21(4):1525–35.
73. Shi Y, Cui L, Qi Z, Meng F, Chen Z. Automatic road crack detection using random structured forests. *IEEE Trans Intell Transp Syst*. 2016;17(12):3434–45.
74. Hadinata PN, Simanta D, Eddy L, Nagai K. Multiclass segmentation of concrete surface damages using u-net and deeplabv3+. *Appl Sci*. 2023;13(4):2398.
75. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. *IEEE Geosci Remote Sens Lett*. 2018;15(5):749–53.
76. Zunair H, Hamza AB. Sharp u-net: depthwise convolutional network for biomedical image segmentation. *Comput Biol Med*. 2021;136:104699.
77. Hilditch C. Comparison of thinning algorithms on a parallel processor. *Image Vision Comput*. 1983;1(3):115–32.
78. Montero AS, Lang J. Skeleton pruning by contour approximation and the integer medial axis transform. *Comput Graph*. 2012;36(5):477–87.
79. Tran TS, Nguyen SD, Lee HJ, Tran VP. Advanced crack detection and segmentation on bridge decks using deep learning. *Constr Build Mater*. 2023;400:132839.
80. Tran VP, Nguyen SD, Lee HJ, Tran TS, Elipse C. Gan-xgb-cavity: automated estimation of underground cavities' properties using gpr data. *Neural Comput Appl*. 2023;35(25):18357–76.
81. Qu Z, Mei J, Liu L, Zhou D-Y. Crack detection of concrete pavement with cross-entropy loss function and improved vgg16 network model. *IEEE Access*. 2020;8:54564–73.
82. Dais D, Bal IE, Smyrou E, Sarhosis V. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. *Autom Constr*. 2021;125:103606.

83. Junior GS, Ferreira J, Millán-Arias C, Daniel R, Junior AC, Fernandes BJ. Ceramic cracks segmentation with deep learning. *Appl Sci*. 2021;11(13):6017.
84. Dorafshan S, Thomas RJ, Maguire M. Sdnet 2018: an annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. *Data Brief*. 2018;21:1664–8.
85. Amhaz R, Chambon S, Idier J, Baltazart V. Automatic crack detection on 2d pavement images: an algorithm based on minimal path selection, accepted to *IEEE trans. Syst Intell Transp*. 2015;17(10):2718–29.
86. Yu Y, Guan H, Li D, Zhang Y, Jin S, Yu C. Ccapfn: a context-augmented capsule feature pyramid network for pavement crack detection. *IEEE Trans Intell Transp Syst*. 2020;23(4):3324–35.
87. Cui L, Qi Z, Chen Z, Meng F, Shi Y. Pavement distress detection using random decision forests. In: *Data Science: Second International Conference, ICDS 2015, Sydney, Australia, August 8-9, 2015, Proceedings 2, 2015*;95–102. Springer.
88. Benz C, Debus P, Ha HK, Rodehorst V. Crack segmentation on uas-based imagery using transfer learning. In: *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), 2019*;1–6. IEEE.
89. Coca L-G, Cusmuluc CG, Iftene A. Automatic tarmac crack identification application. *Proc Compu Sci*. 2021;192:478–86.
90. Aprilyanto J, Yohannes Y. Implementasi arsitektur vgg-unet dalam melakukan segmentasi keretakan pada citra bangunan. In: *MDP Student Conference, 2023*;2:257–264.
91. Lee T, Kim J-H, Lee S-J, Ryu S-K, Joo B-C. Improvement of concrete crack segmentation performance using stacking ensemble learning. *Appl Sci*. 2023;13(4):2367.
92. Valença J, Puente I, Júlio E, González-Jorge H, Arias-Sánchez P. Assessment of cracks on concrete bridges using image processing supported by laser scanning survey. *Construction and Building Materials*. 2017;146:668–78.
93. Kim H, Lee J, Ahn E, Cho S, Shin M, Sim S-H. Concrete crack identification using a uav incorporating hybrid image processing. *Sensors*. 2017;17(9):2052.

Publisher's Note

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