Pavement Crack Segmentation Using an Attention-Based Deep Learning Model

Hieu Dao, Tung Khuc, Quan Truong, Cang Dinh, and Andy Nguyen

Abstract It has been observed that cracks, the most common sign of deterioration happing on the pavement, are difficult to detect at an early stage. Although the U-net-based model has detected well-established cracks, it shows some limitations when working with low-quality pavement images that are automatically captured by moving pavement-inspection vehicles. In this study, the attention technique is applied to the U-net model to enhance the results of pavement crack detection under some difficult pavement image conditions. Attention gates (AGs) are deployed at the skip connections of the U-net model to remove irrelevant regions by setting attention weights for each image part. This procedure helps the U-net model learn how to eliminate extraneous regions in the input image. Therefore, the technique minimizes the computational resources by ignoring wasted irrelevant operations and enhances crack segmentation results. The proposed model is verified using a real-life image packet of pavement. The performance of the attention U-net model illustrates better outcomes compared to the ones from the U-net model.

Keywords Deep learning · Attention gate · U-net · Pavement crack detection

1 Introduction

The development of digital image technologies and data processing algorithms makes it possible to detect and identify the deterioration of civil engineering projects and structures [1]. Due to the constant direct load of vehicles, the pavement surface requires regular monitoring and early detection of damage before it becomes severely damaged. Vision-based methods have been introduced in pavement surface maintenance, starting with images pictured by cameras mounted on a survey vehicle. Then,

H. Dao · T. Khuc (✉) · Q. Truong · C. Dinh
Faculty of Bridge and Road Engineering, Hanoi University of Civil Engineering, Hanoi, Vietnam
e-mail: tungkd@huce.edu.vn

A. Nguyen
School of Engineering, University of Southern Queensland, Springfield Central, QLD, Australia

T. Nguyen-Xuan et al. (eds.), *Proceedings of the 4th International Conference on Sustainability in Civil Engineering*, Lecture Notes in Civil Engineering 344,
https://doi.org/10.1007/978-981-99-2345-8_75

classical image processing algorithms were applied to detect cracks that yielded potential results [2]. Although these methods can detect cracks, their accuracy is still very limited. Since the algorithms are still developed with manually inputting parameters, their results will depend on the experience of monitoring and analysing engineers. Moreover, these methods are commonly less effective with automatically collected images, which always contain a lot of noise.

Recently, a new approach attracted the research community to apply deep learning techniques [3]. An initial study conducting the deep learning algorithm in crack detection has been developed by Kelvin Wang's group based on 3D laser images of pavement surfaces [4]. Majidifard uses Google's application programming interface (API) in street view to propose a method of pavement crack detection [5]. Early CNN-based models for pavement crack detection can only detect cracks at the image or block level. In order to detect the shapes and dimensions of pavement cracks, subsequent studies have proposed methods of pixel-level detection based on decode-encode networks of deep learning. One of the most well-known decode-encode algorithms named U-Net is used by Ju et al. [6] to propose a model called CrackU-net that can detect pavement cracks on images collected by vehicle-mounted smartphones. Chen et al. [7] used the SegNet model, a more minimalist model than the U-Net model, for real-time detection due to the outperformance of SegNet's calculating optimizations. Another study focuses on combining several deep learning models to improve pavement crack detection performance [8]. Image datasets used by those studies are CrackForest [9], CrackTree200 [10], or designed by their own groups [6, 11]. As it is seen that designing and preparing labelled images for pixel-level crack detection are laborious, augmentations are commonly applied [12].

1.1 Motivation and Research Objective

Although the U-net-based model has detected well-established cracks, the main challenge of the algorithm is the inaccurate crack detection outcomes when the pavement images are of low quality due to the movement of the pavement-inspection vehicle. In this study, the attention technique is proposed to overcome that problem. By implementing attention gates (AGs) at skip connections of the U-net model, the irrelevant regions are ignored by placing attention weights on each part of the image. This mechanism will improve the performance of the conventional U-net model and increase the ability to recognize pavement cracks in noisy and low-quality pavement images. Several studies applied attention U-Net to available datasets and showed its effectiveness in pavement crack segmentation [13, 14]. The common point of these studies is that they all use purpose-built datasets generated by still images, which significantly eliminates the noise that the real-life monitoring images may have. Therefore, in this research, image datasets captured by a pavement-inspection vehicle at a real-life highway are labelled and enhanced based on pavement image engineering properties to increase network training performance. Using a dataset that has been applied in real-life monitoring work, the research contribution is to

reduce the workload by improving the automatic pavement crack detection with the attention U-net model. The accuracy metrics are calculated that demonstrated the effectiveness of the proposed method.

2 Dataset

Pavement surface images are collected by two cameras mounted at the rear of a pavement-inspection vehicle. Due to being captured when the vehicle is moving, the image quality is significantly affected. Some popular imaging problems are blurring and having less contrast, which always influences the training effectiveness of deep learning models. To enhance the image quality, a contrast-based method is applied to increase the difference between light and dark parts of images [15]. Data labelling is also a challenge because the labelled image accuracy requires expert intervention, which leads to the number of labelled images is limited. Since a deep learning model always needs large enough datasets, several augmentation techniques are implemented to generate more images. Figure 1 presents the process of image enhancement and augmentation to develop datasets for the training model.

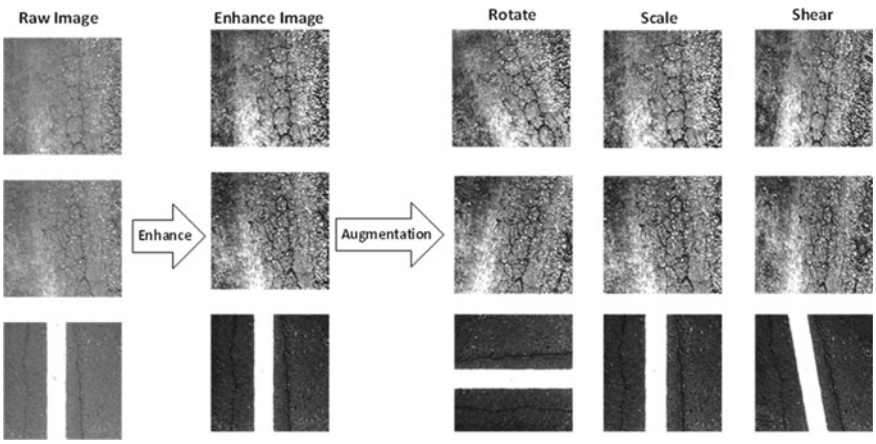


Fig. 1 Image enhancement and augmentation

3 The Attention U-Net Deep Learning Model

3.1 The U-Net Deep Learning Model

An advantage of the U-Net model is that the model can be trained with few images in the dataset. Moreover, the U-Net is a type of pixel classification algorithm, so it has potential for the application of pixel-level crack detection [16]. This model is a symmetric network consisting of two main parts called the encoder and decoder branches (Fig. 2). In this study, the input data are 480 × 480 grayscale images, and the output data are same-size binary segmentation maps. The main component is the convolutional block including a convolutional layer, a batch normalization and a ReLU activation function. The encoder branch is a conventional convolution network. After every two convolutional blocks, a 2 × 2 max-pooling layer with a stride of 2 is arranged to halve the feature map dimension. Since the feature map dimension is halved, the kernel's numbers in convolutional layers are twice as large as those in the previous layer.

In the decoder branch, the max-pooling layer is replaced by the up-sample layer with 2 × 2 kernels to double the dimension of previously feature maps. Since the encoder branch aims to filter, highlight and extract image feature maps, the decoder branch tries to reconstruct those extracted features to map with labelled images for training. The skip connection links the feature maps from the encoder branch to its counterpart in the decoder branch. This mechanism intends to enhance the accuracy of the feature locations for reconstructing images. The sigmoid activation is located at the end of the network to calculate the probability of each of the classes, which are cracks or background pixels in pavement surface images.

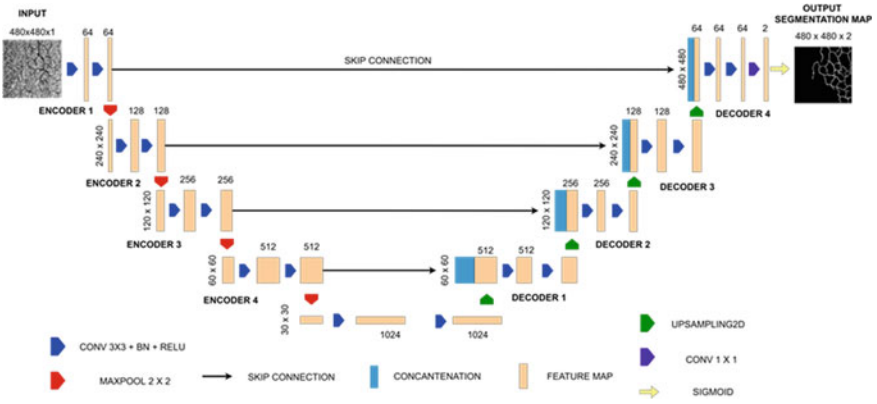


Fig. 2 Architecture of U-net model—modified from a figure in [16]

3.2 Attention Gates (AGs)

Two types of attention are studied: hard attention and soft attention. For hard attention, the relevant region is highlighted by cropping the image, so it only focuses on one region of an image at a time. It means that hard attention deep learning model has only two choices of either paying attention to a relevant region or not. Therefore, this attention type is non-differentiable and could not apply standard backpropagation for the training model as in deep learning methods. On the other hand, soft attention sets weights on each pixel in the image. These attention weights are learnable and are assigned high values to regions of salient feature. Due to this deterministic property, it can be differentiable and applied with standard backpropagation for training [17]. Figure 3a presents the attention U-Net model architecture used in this study. AGs are added to the skip connections and are calculated from the encoder and decoder up-sampling features.

Figure 3b explains the attention gate operation, where the input includes up-sampling feature s in the decoder and encoder feature f from the encoder. The up-sampling feature s is taken from the deeper layer and located before the output layer of the attention gate. Encoder feature f is the output of the layer in the encoder branch at the location of the corresponding attention gate. Because the dimensions of these features are not equal, the feature s has to go through a transposed convolution layer to double its dimensions. When the dimensions of feature s and f are the same, they are summed creating the resultant feature map. This map is then passed through two activation functions and resampled to get the attention coefficient map by using trilinear interpolation. Finally, the attention coefficient map is multiplied by the original feature f and results in the attention weight map [18].

By applying the attention gate, the U-net model takes an additional step to remove the irrelevant and focus on the regions that need attention. Figure 4 shows the original image, the enhanced image and their attention coefficient map. In Fig. 4c, the magnitude of the attention coefficient is represented in the brightness colour fashion. The coefficients at the crack region are clearly shown with the high brightness colour, while the coefficients in the background are assigned almost zero. However, the regions far away from the cracks also have high attention coefficients. This could be explained that the enhanced image still contains noise, and the attention gate cannot be clearly identified and eliminated. The identification and elimination of these regions could be done in the deeper layers of the U-Net model.

4 Experiment and Evaluation

As described in the previous section, a dataset of 20 pavement images was augmented for training two deep learning models: a conventional U-net and an attention U-net. The augmented image dataset consists of 320 images including 288 images for training and 32 images for testing. The deep learning networks were written by Python

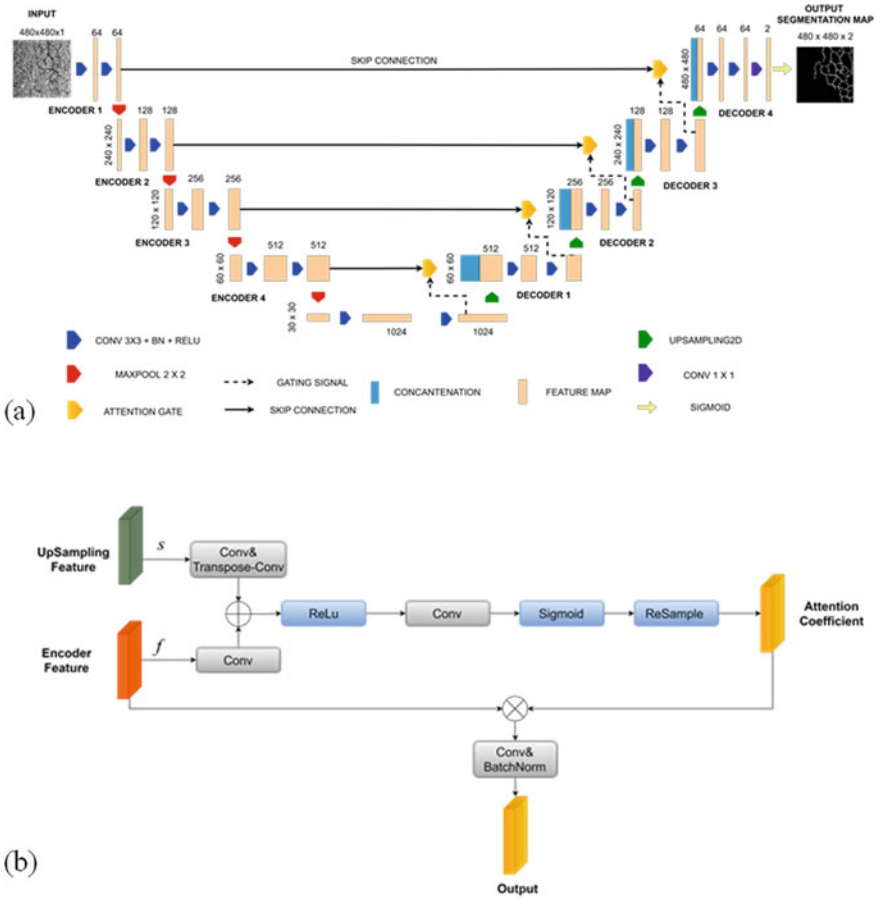


Fig. 3 **a** Architecture of the attention U-net model. **b** Schematic of the attention gate (AG)—modified from figures in [18]

(v3.7) in the Google Colaboratory (Google Colab) environment. This environment allows users to program Python through a web browser and performs tasks in a cloud environment provided by Google. The configuration provided by Google Colab in this research is GPU NVIDIA Tesla T4 and 25 Gb RAM. The image data in this study were processed by using the Albumentation library, a Python-based computer vision toolbox [19]. Figure 5 shows the training accuracy and loss of the attention U-net model after 50 epochs. The last values of accuracy and loss are 0.9923 and 0.1341, respectively.

After training, the last model (at the final epoch) was evaluated with Recall, Precision and F1-score:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

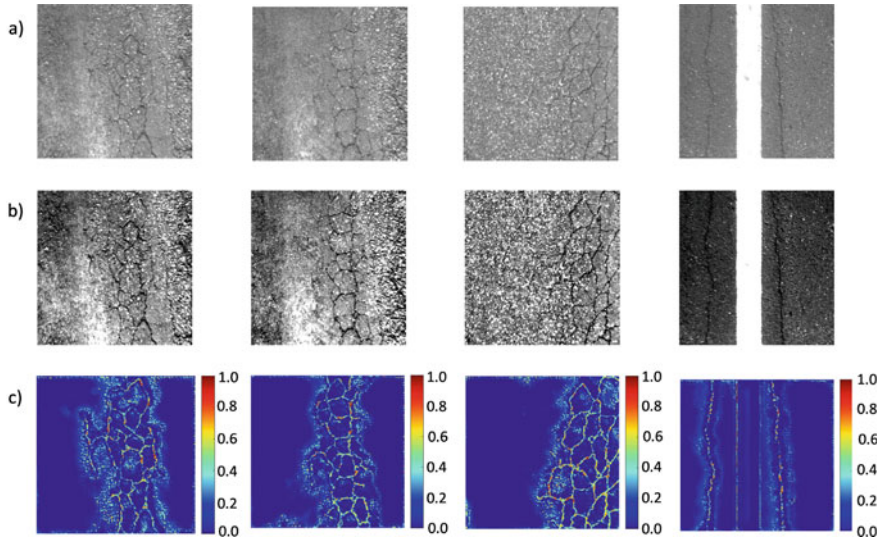
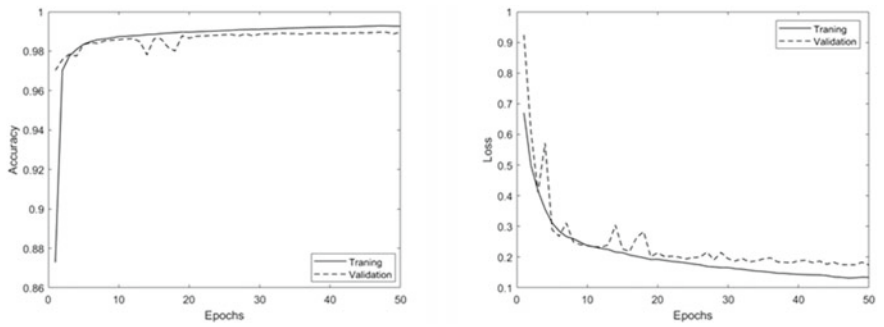


Fig. 4 **a** Original image; **b** enhanced image; **c** the attention coefficient map of the corresponding enhanced image



a) Training b) Loss

Fig. 5 Model accuracy and loss during the training process

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$F1 - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{3}$$

where TP is the true positive (pixels that correctly match the labelled crack region); TN is the true negative (non-defected pavement pixels are correctly detected by the trained model); FP is the false positive (defective pixels are detected incorrectly);

Table 1 Comparison of metrics and running time between the U-net model and attention U-net model

	Precision	Recall	F1-score	Time (seconds)
U-net	0.829	0.817	0.823	3244
Attention U-net	0.893	0.892	0.892	4162

FN is the false negative (defective pixels are not detected by the trained model); Precision is the proportion of correct crack pixels in the detection results; Recall is the proportion of correct crack pixels that are not ignored by the deep learning model, and $F1$-score is the harmonic mean of both Precision and Recall.

Table 1 presents the metric results of each model determined based on absolute pixel accuracy without using tolerance margin. It is seen that the resulting metrics of the attention U-net are better than the U-net model outcomes. Results of the attention U-net model achieve Precision = 0.893 and Recall = 0.892 that outperform the U-net model's values with Precision = 0.829 and Recall = 0.817. It means that the attention U-Net model is capable of sufficiently and accurately detecting the crack shape better than the U-Net model.

Figure 6 includes the original image taken from the pavement surface dataset, the enhanced image, the manually labelled image as ground truth and the detected result image from U-Net and the attention U-Net model. The cracks predicted by both models are almost completely compared to the labelled data. However, the identification results from the attention U-Net are more accurate because the detected crack lines are not as broken as those of the U-Net model. Figure 7 shows more clearly the difference between the detection pixel results of the two deep learning models. In these images, the white pixel represents the crack correctly detected, and black pixels are detected as undamaged pavement background. Green pixels are those that the model did not identify as cracks, but were labelled as damaged. In addition, purple pixels were recognized by the model as cracks, but in the labelled image those pixels were not labelled as cracks. Although the results from both models still have some false detections in the crack edge regions, the purple pixels in the attention U-Net model results are significantly reduced. In the U-Net model, green pixels represent some parts of the crack that are completely ignored, which can be detected precisely with the attention U-Net model.

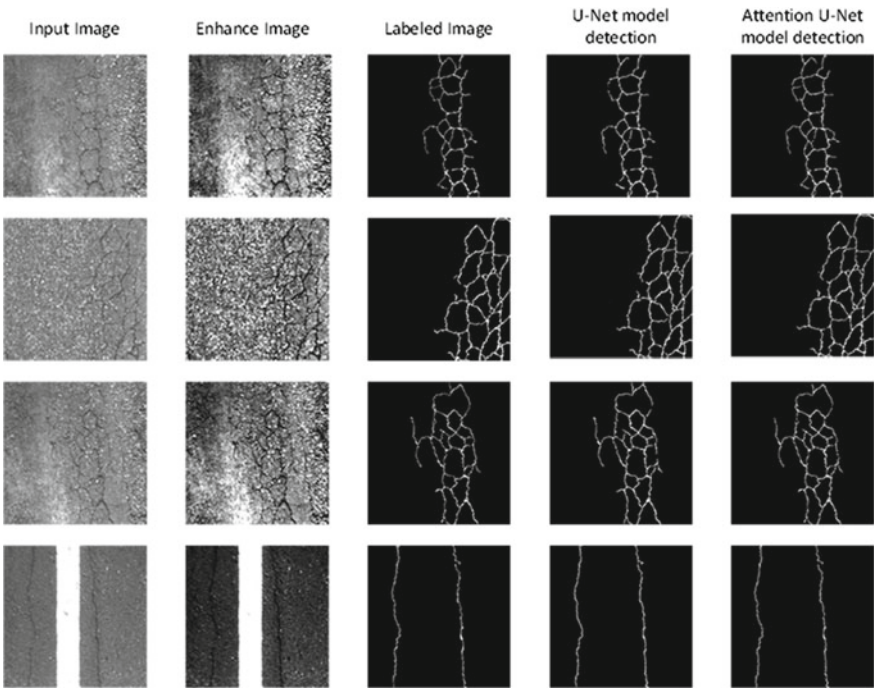


Fig. 6 Comparison of the models' performances

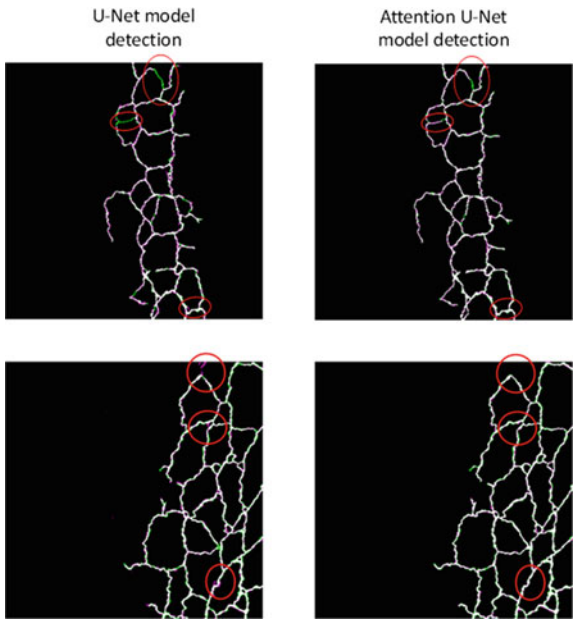


Fig. 7 The crack detection pixel results of the U-Net model and the Attention U-Net model

5 Conclusion

In this study, an attention U-Net-based deep learning model is studied for automatically detecting pavement cracks. The proposed model has the ability to accurately detect the crack shape and dimension better than a conventional U-net performance. By adding the attention gates to the U-net model, the cracks in pavement surfaces can be detected more accurately. The proposed approach has been evaluated using the pavement surface dataset at a real-life expressway and achieving some potential results. The model's detection results are $Precision = 0.893$ and $Recall = 0.892$, which is a higher rate of detection compared to previous studies. The results show that the attention U-net model is suitable for real-life crack detection tasks, which can replace human labour.

Acknowledgements This research is funded by Hanoi University of Civil Engineering (HUCE) under the grant number of 10-2022/KHXD-TĐ.

References

1. Khuc T, Catbas FN (2018) Structural identification using computer vision–based bridge health monitoring. J Struct Eng 144:04017202
2. Wang KC, Li Q, Gong W (2007) Wavelet-based pavement distress image edge detection with a trous algorithm. Transp Res Rec 2024:73–81
3. Brisolin J, Nguyen A, Ullah F, Bourke A, Khuc T, Nsanabo V (2022) Smart automated fault detection for improved road maintenance planning in Australia
4. Wang KC, Zhang A, Li JQ, Fei Y, Chen C, Li B (2017) Deep learning for asphalt pavement cracking recognition using convolutional neural network. Airfield Highway Pavements 2017:166–177
5. Majidifard H, Jin P, Adu-Gyamfi Y, Buttlar WG (2020) Pavement image datasets: a new benchmark dataset to classify and densify pavement distresses. Transp Res Rec 2674:328–339
6. Huyan J, Li W, Tighe S, Xu Z, Zhai J (2020) CrackU-net: a novel deep convolutional neural network for pixelwise pavement crack detection. Struct Control Health Monit 27:e2551
7. Chen T, Cai Z, Zhao X, Chen C, Liang X, Zou T, Wang P (2020) Pavement crack detection and recognition using the architecture of segNet. J Ind Inform Integ 18:100144
8. Lau SL, Chong EK, Yang X, Wang X (2020) Automated pavement crack segmentation using u-net-based convolutional neural network. IEEE Access 8:114892–114899
9. Shi Y, Cui L, Qi Z, Meng F, Chen Z (2016) Automatic road crack detection using random structured forests. IEEE Trans Intell Transp Syst 17:3434–3445
10. Qu Z, Mei J, Liu L, Zhou D-Y (2020) Crack detection of concrete pavement with cross-entropy loss function and improved VGG16 network model. IEEE Access 8:54564–54573
11. Zhang K, Zhang Y, Cheng H-D (2020) CrackGAN: pavement crack detection using partially accurate ground truths based on generative adversarial learning. IEEE Trans Intell Transp Syst 22:1306–1319
12. Huang Z, Chen W, Al-Tabbaa A, Brilakis I (2022) NHA12D: a new pavement crack dataset and a comparison study of crack detection algorithms. arXiv preprint arXiv:2205.01198
13. König J, Jenkins MD, Barrie P, Mannion M, Morison G (2019) A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating. In: Proceedings of the 2019 IEEE international conference on image processing (ICIP). IEEE, pp 1460–1464

14. Wu Z, Lu T, Zhang Y, Wang B, Zhao X (2020) Crack detecting by recursive attention U-Net. In: Proceedings of the 2020 3rd international conference on robotics, control and automation engineering (RCAE). IEEE, pp 103–107
15. Negi SS, Bhandari YS (2014) A hybrid approach to image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram. In: International conference on recent advances and innovations in engineering (ICRAIE-2014). IEEE, pp 1–6
16. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, New York, pp 234-241
17. Jetley S, Lord NA, Lee N, Torr PH (2018) Learn to pay attention. arXiv preprint arXiv:1804.02391
18. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B (2018) Attention u-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999
19. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020) Albumentations: fast and flexible image augmentations. Information 11:125