



## OPEN ACCESS

## EDITED BY

Chengyi Pu,  
Central University of Finance and  
Economics, China

## REVIEWED BY

Peng Lin,  
China University of Mining and  
Technology, China  
Enhedelihai Nilot,  
National University of Singapore,  
Singapore

## \*CORRESPONDENCE

Enhong Meng,  
✉ [enhm\\_1990@163.com](mailto:enhm_1990@163.com)

## SPECIALTY SECTION

This article was submitted to  
Environmental Informatics and Remote  
Sensing, a section of the journal  
Frontiers in Environmental Science

RECEIVED 22 November 2022

ACCEPTED 04 January 2023

PUBLISHED 22 February 2023

## CITATION

Lin X, Zhang J, Wu D, Meng E, Liu M, Li M  
and Gao F (2023), Intelligent identification  
of pavement cracks based on PSA-Net.  
*Front. Environ. Sci.* 11:1105467.  
doi: 10.3389/fenvs.2023.1105467

## COPYRIGHT

© 2023 Lin, Zhang, Wu, Meng, Liu, Li and  
Gao. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Intelligent identification of pavement cracks based on PSA-Net

Xuan Lin<sup>1</sup>, Jian Zhang<sup>2</sup>, Daifeng Wu<sup>3</sup>, Enhong Meng<sup>4\*</sup>, Maoyi Liu<sup>3</sup>,  
Meng Li<sup>2</sup> and Fuli Gao<sup>2</sup>

<sup>1</sup>School of Mechano-Electronic Engineering, Xidian University, Xi'an, China, <sup>2</sup>Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China, <sup>3</sup>Chongqing Urban Construction Investment (Group) Co., Ltd., Chongqing, China, <sup>4</sup>School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, China

The identification of pavement cracks is critical for ensuring road safety. Currently, manual crack detection is quite time-consuming. To address this issue, automated pavement crack-detecting technology is required. However, automatic pavement crack recognition remains challenging, owing to the intensity heterogeneity of cracks and the complexity of the backdrop, e.g., low contrast of damages and backdrop may have shadows with comparable intensity. Motivated by breakthroughs in deep learning, we present a new network architecture combining the feature pyramid with the attention mechanism (PSA-Net). In a feature pyramid, the network integrates spatial information and underlying features for crack detection. During the training process, it improves the accuracy of automatic road crack recognition by nested sample weighting to equalize the loss caused by simple and complex samples. To verify the effectiveness of the suggested technique, we used a dataset of real road cracks to test it with different crack detection methods.

## KEYWORDS

deep learning, pavement crack identification, pyramid pooling, feature fusion, risk assessment

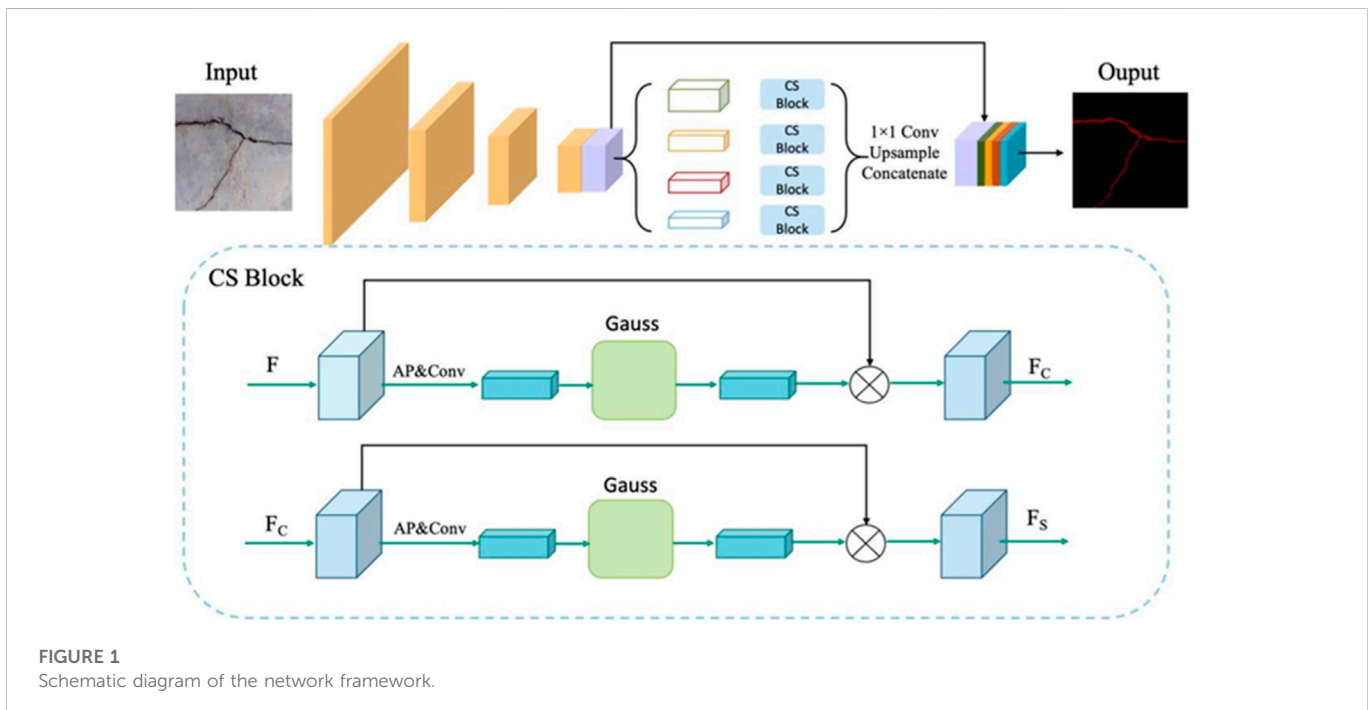
## 1 Introduction

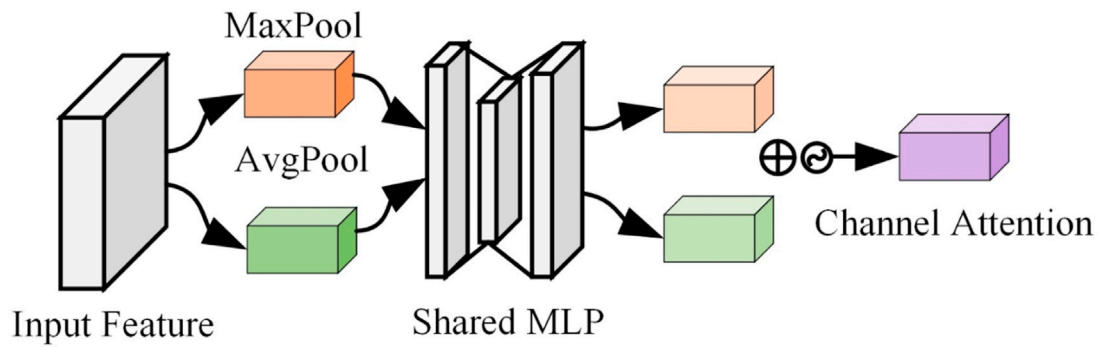
With the continuous development and improvement of China's roads as infrastructure construction, how to carry out scientific and intelligent maintenance management has become a fundamental research problem. The diagnosis of pavement diseases is an important issue in road maintenance, and cracking is the dominant type of pavement disease. The identification of pavement cracks is a critical duty for ensuring road safety. In this work on pavement crack detection, methods based on non-deep learning approaches are referred to as traditional crack-detecting methods. In the past few years, many researchers have been working on the automated detection of cracks. These works can be divided into five categories: 1) crack detection methods based on the wavelet transform: the wavelet transform decomposes the image into different frequency bands, and defect and noise are converted into distinct amplitude wavelet coefficients, which allows them to be applied to pavement crack detection work. [Peggy et al. \(2006\)](#) created a complicated coefficient map by applying a multi-scale 2D wavelet transform ([Peggy et al., 2006](#)); the crack region is then determined by scanning for the wavelet coefficients from the most enormous scale to the most miniature scale. However, this method cannot handle cracks with limited continuity or significant curvature characteristics. 2) Image thresholding crack detection method: Scholars use preprocessing algorithms to reduce lighting artifacts and then threshold the image to obtain candidate cracks. The processed crack images are further refined using morphological techniques ([Chambon and Jean-Marc, 2011](#); [Huang and Zhang, 2012](#); [Xu et al., 2013](#); [Li et al., 2015](#)). The aforementioned methods were further developed, and new graph-based methods can achieve crack candidate refinement ([Zou et al.,](#)

2012; Kelwin and Lucian, 2014; Marcos et al., 2016). 3) Handcrafted features and classification to achieve crack detection: Most conventional crack-detecting algorithms rely on handcrafted features and patch-based classifiers to achieve crack detection. Handcrafted features, such as HOG and LBP, are extracted from the image patches as descriptors. Then, a classifier, such as a support vector machine, is used to achieve crack recognition and classification (Hu and Zhao, 2010; Zakeri et al., 2013; Srivatsan et al., 2014; Rafal et al., 2015). 4) Crack detection is carried out based on edge detection: Yan et al. (2007) introduced a morphological filter in crack detection and used an improved median filter to remove noise to achieve crack detection (Yan et al., 2007). Albert and Nii (2008) applied the Sobel edge detector to detect cracks and used a two-dimensional empirical pattern decomposition algorithm to remove speckle noise (Albert and Nil, 2008). Stochastic structure forest was used by scholars and combined with structural information for crack detection (Piotr and Lawrence, 2013; Shi et al., 2016). 5) Complete crack detection based on the minimal path: The shortest path approach was suggested by Kass et al. (1988) to extract basic open curves from photographs for achieving crack detection (Kass et al., 1988). Vivek et al. (2012) proposed using an improved minimal path method to detect the same type of similar contours in the image structure to achieve crack detection (Vivek et al., 2012). This enhanced method requires less *a priori* knowledge about the topology and endpoints of the required curve. Amhaz et al. (2016) proposed a two-stage approach for crack detection: first, endpoints are selected in the local range, and second, the minimum path is selected in the global range to finally achieve crack detection (Amhaz et al., 2016). Nguyen et al. (2011) proposed a two-stage approach for crack detection by introducing freeform anisotropic features that offered a method that simultaneously considered strength and the shape of cracks to complete the identification and detection of cracks (Nguyen et al., 2011). However, traditional crack detection methods are extremely challenging to identify and detect due to the strong influence of human factors and the low efficiency of detection based on

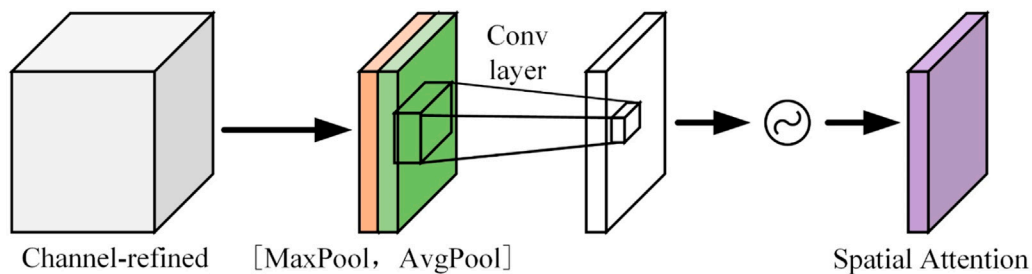
transformation methods and are not applicable to complex scenes. Their performance is still limited.

For the last several years, deep learning has seen extraordinary progress in the field of computer vision (Alex et al., 2017). Scholars have made several attempts to use deep learning techniques for crack identification. Zhang et al. (2016) developed a patch-based fracture detection neural network comprising four convolutional layers and two fully connected layers (Zhang et al., 2016). In addition, Zhang et al. (2016) compared their approach with handcrafted features to demonstrate the advantages of deep learning methods in feature representation. Pauly et al. (2017) used a deeper neural network to identify the road cracks (Pauly et al., 2017). Feng et al. (2017) proposed a deep active learning system to deal with the problem of limited labeling resources (Feng et al., 2017). Eisenbach et al. (2017) presented a road condition dataset for training a deep learning network and first evaluated and analyzed state-of-the-art methods for pavement distress detection. The approach mentioned previously considers crack detection as a patch-based classification challenge, dividing each picture into tiny patches and then training a deep neural network to determine whether or not each patch is a crack (Eisenbach et al., 2017). However, this approach has a complex operational process and is sensitive to the size of the patches. With the rapid development of semantic segmentation tasks (Jonathan et al., 2017; Vijay et al., 2017; Fan et al., 2018), many algorithms have been applied to different scenarios with good application results. Schmutge et al. (2017) proposed a SegNet-based crack segmentation method to detect cracks by aggregating crack probabilities. In conclusion, deep learning-based techniques show great potential for road crack detection applications (Stephen et al., 2017). Joshi et al. (2022) adopted a segmentation-based deep learning method for surface crack detection. Wang et al. (2022) presented fully convolutional network architecture for crack detection in fast-stitching images. To this end, we propose a new deep learning framework called PSA-Net for the particular task of road crack detection, which focuses on three aspects, namely, multi-scale feature information





**FIGURE 2**  
Schematic diagram of the channel attention module. MaxPool means maximum pooling, and AvgPool means average pooling.



**FIGURE 3**  
Schematic diagram of the spatial attention module.

extraction, spatio-temporal attention mechanism, and pyramid pooling, focusing on the contextual semantic information and edge information on crack images to achieve end-to-end pavement crack detection, which aims to improve the accuracy of crack intelligent recognition and detection.

## 2 Methods

For segmentation tasks, contextual information impacts the segmentation’s effectiveness. Generally speaking, when we judge the category of an object, besides directly observing its appearance, we sometimes also assist the environment in which it appears, ignoring these to make judgments that sometimes cause problems. The intelligent crack identification and detection of pavement in this paper are similar to the segmentation task, where we can improve the accuracy and precision of crack identification with the help of auxiliary information. First, this subsection describes the composition of the algorithm, and the network used consists of encoder and decoder architecture, as shown in Figure 1. We use ResNet-101 as the backbone in the feature extraction stage (He et al., 2016). The encoder uses the pre-trained model (ResNet-101) and the dilated convolution strategy to achieve the feature map extraction, and the extracted feature map is 1/8 the size of the input. The pyramid pooling module fuses the feature map to get the fused feature with general information, which is upsampled and connected with the feature map before pooling. Finally, the final output is obtained by a convolutional layer.

### 2.1 Dilated convolution

Dilated convolution is a technique for solving picture semantic segmentation issues in which downsampling affects image resolution and results in information loss. By adding holes to expand the perceptual field, the original  $3 \times 3$  convolution kernel can have a perceptual area of  $5 \times 5$  (dilated rate = 2) or be more significant with the same number of parameters and computation, thus eliminating the need for downsampling. It has the advantage of increasing the field of perception and allowing each convolution output to contain a more extensive range of information. The information is without pooling data or creating loss ambiguity under the same computational conditions. Dilated convolution is often used in real-time image segmentation. Dilated convolution can be used when the network layer requires a bigger perceptual field, but the number or size of convolution kernels cannot be increased, owing to restricted computing resources. The feature extraction module of our network uses dilated convolution to increase the perceptual field and further improve the segmentation efficiency. The mathematical expression of dilated convolution is (Yu and Koltun, 2015)

$$M(i, j) = \sum_{x=1}^X \sum_{y=1}^Y h(i + atr \times x, j + atr \times y) \times g(x, y), \quad (1)$$

where  $i$  and  $j$  denote the positions of the image;  $X$  and  $Y$  represent the length and width of the input image, respectively;  $h(i, j)$  denotes the feature value of the input image at  $(i, j)$ ;  $atr$  denotes the void rate;  $g$  is the convolution kernel function; and  $M(i, j)$  denotes the result of the input image after convolution. Dilated convolution is to expand the

size of the convolution kernel by zero-filling in the standard convolution, so that it can better capture the context information on the feature map. The size of the dilated convolution is achieved by adjusting the atrous rate.

## 2.2 Pyramid pooling

The pyramid pooling module aggregates contextual data from several places and enhances the capacity to access global data. Experiments demonstrate that such an *a priori* representation (referring to PSP as a structure) is successful and produces outstanding results on a variety of datasets. The module incorporates four pyramid-scale features. The first row in red is the coarsest feature global pooling, generating a single bin output, and the next three rows are pooling features at different scales (as shown in Figure 1). To ensure the weight of the global features, if the pyramid has a total of  $N$  levels, a  $1 \times 1$  convolution is used after each class to reduce level channels to the original  $1/N$ . The size before unspooling is then obtained by bilinear interpolation and CONCAT function together. The pooling kernel size for the pyramid levels is settable and related to the input sent to the pyramid. We used four ranks with kernel sizes of  $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ , and  $6 \times 6$ .

## 2.3 Feature fusion

Feature fusion is a popular component of current network topologies that merges features from distinct layers or branches. It is often performed using basic operations (such as summation or splicing); however, this is not always the best option. This is a unified general scheme for attentional feature fusion that applies to the most common scenarios, including short-hop and long-hop connections and feature fusion induced in the inception layer. As shown in Figure 1 and Figure 2, we present the multi-scale channel attention module, which solves the challenges of fusing information supplied at distinct scales to better fuse features with inconsistent meanings and scales. We also show that the early integration of feature maps might be a bottleneck, which can be solved by adding another attention level. With fewer parameters or network layers, our model outperforms the latest network on both road crack segmentation datasets, suggesting the more sophisticated attention mechanism used for feature fusion has excellent potential to consistently produce better results than direct feature fusion.

The input feature maps in the channel attention module are pooled and averaged into the shared MLP layer. Then, the output features of the shared MLP layer are summed by sending elementwise and activated by the sigmoid function to obtain the feature map of the channel attention module. The channel attention module (CAM) compresses the feature map in spatial dimensions to get a one-dimensional vector and then operates on it. The channel attention module focuses on what is essential in this graph. Mean pooling has feedback for every pixel point on the feature map. In contrast, maximum pooling has feedback for gradients only where the response is most evident in the feature map when performing gradient backpropagation calculations.

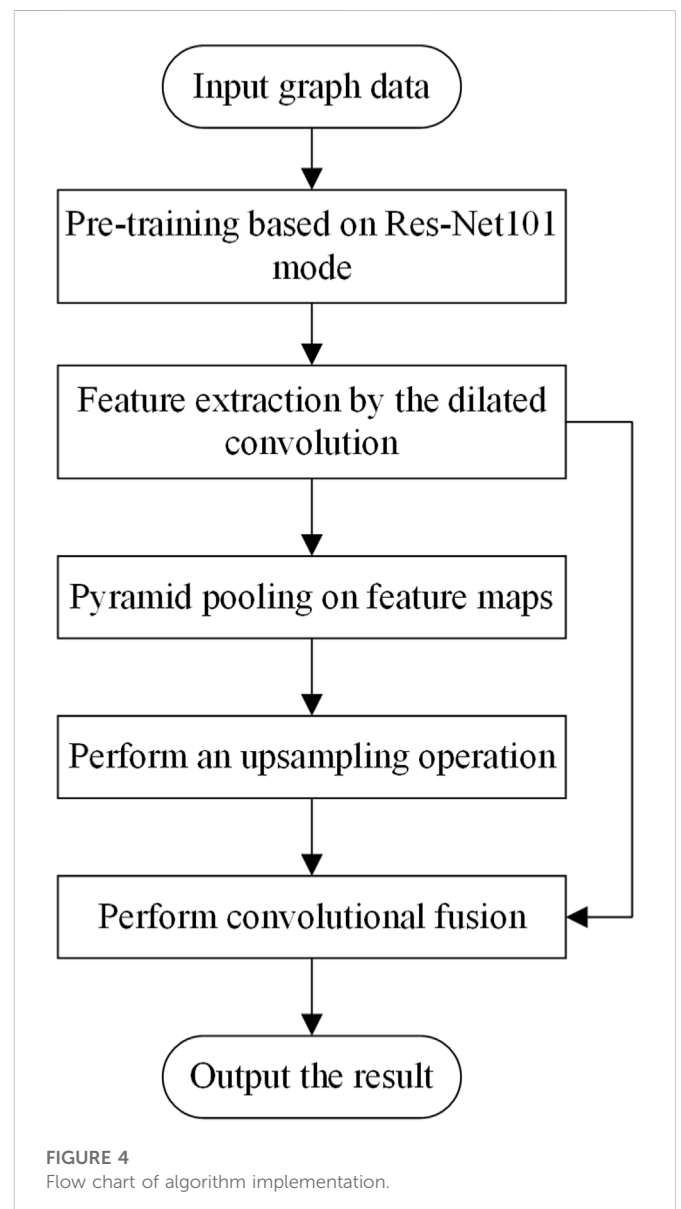
The feature map generated from the channel attention module is utilized as an input in the spatial attention module (as shown in Figure 3). First, execute maximum and average pooling depending

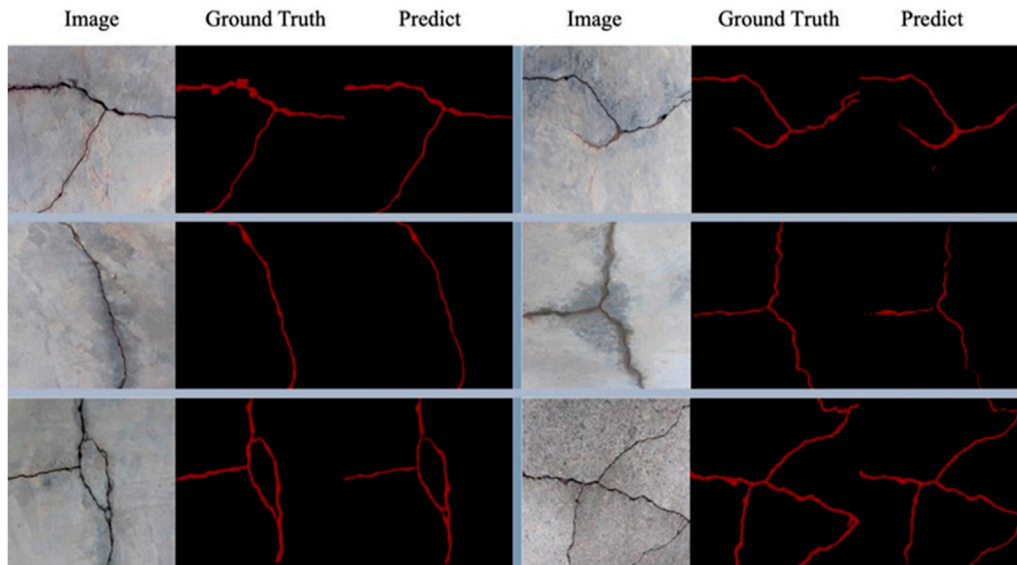
on the channel, followed by the CONCAT operation on both layers. The feature map produced from the spatial attention module is then obtained by the sigmoid function after convolution is conducted and reduced to one channel.

The spatial attention module compresses channels and performs mean and maximum pooling in the channel dimension. The final pooling operation is to extract the channel's greatest value, and the number of extractions is  $H \times W$ . The average pooling procedure is used to acquire the channel's average value, and the number of extractions is also  $H \times W$ . As a result, a two-channel feature map is generated.

## 2.4 Loss function

The loss function is the gap between the predicted and actual values of the model. That is to find a standard to help the training mechanism optimize parameters at any time, so as to find the parameters with the highest precision of the network role at all times toameterork proposes to





**FIGURE 5**  
Road surface crack detection results.

**TABLE 1 Comparison of the prediction performance of different methods.**

| Method         | Precision | Recall | Accuracy | F1    |
|----------------|-----------|--------|----------|-------|
| HED            | .9390     | .9025  | .9846    | .9452 |
| RCF            | .9451     | .9155  | .9219    | .9392 |
| PCN            | .9550     | .9249  | .9412    | .9433 |
| PSPNet         | .9652     | .9323  | .9721    | .9534 |
| DeepLabv3      | .9877     | .9432  | .9844    | .9567 |
| PAS-Net (ours) | .9955     | .9450  | .9870    | .9644 |

**TABLE 2 Comparison of the prediction performance of different methods.**

| Method                   | Dice |
|--------------------------|------|
| CNN + CS                 | .87  |
| CNN + CS + ASPP          | .92  |
| ResNet + CS              | .90  |
| ResNet + CS + ASPP       | .94  |
| ResNet (pre) + CS + ASPP | .98  |

describe a criterion to help the training mechanism in optimizing the par facilitate finding the network at the greatest accuracy. We want the predicted value to be infinitely close to the actual value, so the difference needs to be minimized (in this process, the loss function needs to be introduced). The choice of the loss function in this process is very critical. In specific projects, some loss functions calculate the gradient of the difference falling fast, while others fall slowly, so choosing the appropriate loss function is also very critical.

As the most common loss function, cross-entropy is not optimal for MIS tasks, like road crack image segmentation, in which objects often occupy only a small area, or some medical image processing tasks such as, for example, retinal vascular and eye segmentation. We use the Dice coefficient loss function instead of the common cross loss due to its performance in the presence of ground truth, which is widely used to evaluate segmentation performance. Suppose  $k$  is the class label, where  $k = \{1, 2, \dots, K\}$ , and  $K \in \mathbb{N}$ . The ground truth label vector and the predicted probability vector can be expressed as  $Y = \{y_1(k), y_2(k), \dots, y_i(k), \dots, y_N(k)\}$ , where  $\hat{y}_i(k) \in [0, 1, 2, \dots, n]$  and  $\hat{Y}_i(k) \in [0, 1, 2, \dots, n]$ . The Dice loss function can be expressed as follows:

$$L_{Dice} = 1 - \frac{2\omega_k \sum_i^N \hat{y}_i(k) y_i(k)}{\sum_i^N \hat{y}_i^2(k) + \sum_i^N y_i^2(k)} \tag{2}$$

where  $N$  denotes the number of pixels and  $k$  and  $\omega_k$  are the number of classes and the category weights, respectively.

Figure 4 shows the flow chart of the algorithm of the proposed method in the paper.

### 3 Numerical experiments

#### 3.1 Data set

We collected and labeled a road crack dataset. The dataset contains a total of 600 images. We split the 600 images into a training set, a test set, and a validation set in the ratio of 8:1:1. We also conducted extensive experiments on several public benchmark datasets. The experimental results demonstrate that our method achieves SOTA performance compared to the most common methods.

## 3.2 Evaluation indicators

We used the PyTorch deep learning framework to build our network. We considered several performance metrics for experimental comparison to better evaluate the experimental results, including accuracy, F1-score, recall, precision, and other evaluation metrics.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{F1 - score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

where TP denotes the true-positive case, FP denotes the false-positive case, FN denotes the false-negative case, and TN denotes the true-negative case.

## 3.3 Experimental details and experimental results

Our proposed network is an end-to-end architecture system that uses a ResNet network that has been pre-trained on ImageNet for four downsampling operations in the encoder phase. We used the PyTorch deep learning framework to build our network. The training and testing platforms were both Ubuntu 18.04, and in terms of hardware configuration, two 3090 graphics cards were used, each with 24G of video memory. In the training process, we used the small batch stochastic gradient descent (SGD) method with a batch size of 8 and a learning rate of .0001. We also used Adam optimization and SGD methods for comparison experiments, and chTor found that SGD usually performs better but that Adam converges in a shorter time. Although Adam converges faster, we carry out a better and more biased performance in terms of performance and time.

The results of our test on the dataset are shown in Figure 5. It is worth noting that these datasets are not exposed to the network in advance to better validate the effectiveness of the proposed method, where image is the original input of an RGB image, ground truth represents our label, and predict is the resulting map predicted by our algorithm. The experimental results show that the results achieved by our algorithm on the crack images have been less different from ground truth, which is enough to prove the effectiveness and accuracy of our algorithm.

The specific performance comparison of different methods is shown in Table 1. The values of the four metrics of the HED (holistically nested edge detection) method such as precision, recall, accuracy, and F1 are .9390, .9025, .9846, and .9452, respectively. The values of the four metrics of the RCF (richer convolutional feature) method such as precision, recall, accuracy, and F1 are .9451, .9155, .9219, and .9392, respectively. The values of the four metrics of the FCN (fully convolutional network) method such as precision, recall, accuracy, and F1 are .9451, .9155, .9219, and .9392, respectively. Our method PAS-Net has values of .9955, .9450, .9870, and .9644 for the

four metrics such as precision, recall, accuracy, and F1, respectively. It can be seen that our method achieves the most advanced performance in all four evaluation metrics compared to these state-of-the-art methods.

To further validate the adequate performance of our method, we disassembled it into multiple parts. We conducted a large number of combinatorial experiments to validate the efficiency of each module fully. As shown in Table 2, the most basic CNN + CS (context spatial) yields a dice metric of .87 on our dataset. CNN + CS + ASPP (atrous spatial pyramid pooling) yields a dice metric of .92 on our dataset. We use ResNet as the backbone; the combination of ResNet + CS gives a dice metric of .90 on our dataset, and ResNet + CS + ASPP provides a dice metric of .94 on our dataset. In addition, the best performance of .98 is achieved when using ResNet that has been pre-trained on ImageNet. Step-by-step tests further validate the effectiveness of the proposed method in this study.

## 4 Conclusion

This study proposes a new deep learning network architecture called PSA-Net for road crack detection. Our algorithm is developed from three aspects, namely, multi-scale feature information extraction, spatio-temporal attention mechanism, and pyramidal pooling, focusing on the contextual semantic information and edge information on crack images, and it is an end-to-end segmentation algorithm. We designed PSA-Net for pavement crack detection without increasing the number of network parameters. We conducted our experiments on our road crack detection datasets. The experimental results show that our PSA-Net offers a significant advantage on various datasets, sufficient to prove the algorithm's effectiveness. In the future, we will continue to contribute to road crack detection and consider combining the detection model with the segmentation model to improve the performance of our algorithm further (H et al., 2013).

## Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## Author contributions

XL proposed the architecture of the algorithm and wrote the paper. JZ helped write the paper and implement some of the algorithms. DW and MaL provided experimental data. EM gave insight into the implementation process. MeL and FG labeled the images and helped further refine the paper.

## Funding

This work was supported by the China National Key R&D Program under grant 2022YFF0802600.

## Conflict of interest

DW and ML were employed by Chongqing Urban Construction Investment (Group) Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Albert, A., and Nii, A. (2008). Evaluating pavement cracks with bidimensional empirical mode decomposition. *EURASIP. J. Adv. Sig. Pr.* 2008, 861701. doi:10.1155/2008/861701
- Alex, K., Ilya, S., and Geoffrey E. H. (2017). Imagenet classification with deep convolutional neural networks. *Commun. Acm.* 60 (6), 84–90. doi:10.1145/3065386
- Amhaz, R., Chambon, S., Idier, J., and Baltazart, V. (2016). Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection. *Ieee. Trans. Intell. Transp.* 17 (10), 2718–2729. doi:10.1109/TITS.2015.2477675
- Chambon, S., and Jean-Marc, M. (2011). Automatic road pavement assessment with image processing: Review and comparison. *Int. J. Geophys.* 2011, 1–20. doi:10.1155/2011/989354
- Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., et al. (May 2017). “How to get pavement distress detection ready for deep learning? A systematic approach,” in Proceedings of the 2017 International Joint Conference on Neural Networks (IEEE), Anchorage, AK, USA 2161–4407.
- Fan, H., Mei, X., Danil, P., and Ling, H. (2018). Multi-level contextual rnns with attention model for scene labeling. *Ieee. Trans. Intell. Transp.* 19 (11), 3475–3485. doi:10.1109/TITS.2017.2775628
- Feng, C., Liu, M., Kao, C., and Lee, T. (June 2017). “Deep active learning for civil infrastructure defect detection and classification,” in Proceedings of the ASCE International Workshop on Computing in Civil Engineering (Seattle, Washington: UnivIWCC), 298–306.
- He, K., Zhang, X., Ren, S., and Sun, J. (June 2016). Deep residual learning for image recognition. Proceedings of the IEEE Conf. Comput. Vis. pattern Recognit., Las Vegas, NV, USA 770–778.
- Hu, Y., and Zhao, C. (2010). A novel lbp based methods for pavement crack detection. *J. pattern Recognit. Res.* 5 (1), 140–147. doi:10.13176/11.167
- Huang, W., and Zhang, N. (December 2012). “A novel road crack detection and identification method using digital image processing techniques,” in Proceedings of the 7th International Conference on Computing and Convergence Technology (Seoul: ICCCT), 397–400.
- Jonathan, L., Evan, S., and Trevor, D. (2017). Fully convolutional networks for semantic segmentation. *Ieee. Trans. Pattern Anal.* 39 (4), 640–651. doi:10.1109/TPAMI.2016.2572683
- Joshi, D., Singh, T., and Sharma, G. (2022). Automatic surface crack detection using segmentation-based deep-learning approach. *Eng. Fract. Mech.* 268, 108467. doi:10.1016/j.engfracmech.2022.108467
- Kass, A., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *Int. J. Comput. Vis.* 1, 321–331. doi:10.1007/bf00133570
- Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models, *International journal of computer vision* 1.4 321–331. doi:10.1007/s11263-022-01685-7
- Kelwin, S., and Lucian, C. (October 2014). “Pavement pathologies classification using graph-based features,” in Proceedings of the IEEE International Conference on Image Processing (Paris, France: IEEE), 793–797.
- Li, P., Wang, C., Li, S., and Feng, B. (September 2015). “Research on crack detection method of airport runway based on twice-threshold segmentation,” in Proceedings of the Fifth International Conference on Instrumentation & Measurement, Computer, Communication, and Control (Qinhuangdao, China: IMCCC), 1716–1720.
- Marcos, Q., Juan, T., and Jose, M. (2016). A simplified computer vision system for road surface inspection and maintenance. *Ieee. Trans. Intell. Transp.* 17, 608–619. doi:10.1109/TITS.2015.2482222
- Nguyen, T. S., Stephane, B., Florent, D., and Manuel, A. (September 2011). “Free-form anisotropy: A new method for crack detection on pavement surface images,” in Proceedings of the 18th IEEE International Conference on Image Processing (Brussels, Belgium: ICIP), 1069–1072.
- Pauly, L., Peel, H., Luo, S., Hogg, D., and Fuentes, R. (July 2017). “Deeper networks for pavement crack detection,” in Proceedings of the 34th International Symposium in Automation and Robotics in Construction (Taipei, Taiwan: IAARC), 479–485.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Peggy, S., Jean, D., Vincent, L., and Dominique, B., (October 2006). “Automation of pavement surface crack detection using the continuous wavelet transform,” in Proceedings of the IEEE International Conference on Image Processing. Atlanta, GA, USA: IEEE, 30–37.

Piotr, D., and Lawrence, Z., (December 2013). “Structured forests for fast edge detection,” in Proceedings of the IEEE International Conference on Computer Vision. Sydney, NSW, Australia: ICCV, 1841–1848.

Rafal, K., Pawel, S., Adam, T., Andrzej, R., Andrzej, P., Pawel, R., et al. (June 2015). “Asphalt surfaced pavement cracks detection based on histograms of oriented gradients,” in Proceedings of the 22nd International Conference on Mixed Design of Integrated Circuits & Systems (Torun, Poland: IEEE), 579–584.

Schmugge, S. J., Rice, L., Lindberg, J., et al. (2017). “Crack segmentation by leveraging multiple frames of varying illumination[C],” in IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE), 1045–1053.

Shi, Y., Cui, L., Qi, Z., Meng, F., and Chen, Z. (2016). Automatic road crack detection using random structured forests. *Ieee. Trans. Intell. Transp.* 17 (12), 3434–3445. doi:10.1109/TITS.2016.2552248

Srivatsan, V., Sobhagya, J., Karan, S., Lars, W., and Christoph, M. (March 2014). “Vision for road inspection,” in Proceedings of the IEEE Winter Conference on Applications of Computer Vision (Steamboat Springs, CO, USA: IEEE), 115–122.

Stephen J, S., Lance, R., John, L., Robert, G., Chris, J., and Min C, S. (March 2017). “Crack segmentation by leveraging multiple frames of varying illumination,” in Proceedings of the 17th IEEE Winter Conference on Applications of Computer Vision (Santa Rosa, CA, USA: IEEE), 1045–1053.

Vijay, B., Alex, K., and Roberto, C. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *Ieee. Trans. Pattern Anal.* 39 (12), 2481–2495. doi:10.1109/TPAMI.2016.2644615

Vivek, K., Anthony, K., and James, T. (2012). Detecting curves with unknown endpoints and arbitrary topology using minimal paths. *Ieee. Trans. Pattern. Anal.* 34 (10), 1952–1965. doi:10.1109/TPAMI.2011.267

Wang, S., Liu, C., and Zhang, Y. (2022). Fully convolution network architecture for steel-beam crack detection in fast-stitching images. *Mech. Syst. Signal Process.* 165, 108377. doi:10.1016/j.ymssp.2021.108377

Xu, W., Tang, Z., Zhou, J., and Ding, J. (September 2013). “Pavement crack detection based on saliency and statistical features,” in Proceedings of the 20th IEEE International Conference on Image Processing (Melbourne, VIC, Australia: ICIP), 4093–4097.

Yan, M., Bo, S., Xu, K., and He, Y. (June 2007). “Pavement crack detection and analysis for high-grade highway,” in Proceedings of the 8th International Conference on Electronic Measurement and Instruments (Xian, China, 548–552. ICEMI

Yu, F., and Koltun, V., (2015). Multi-scale context aggregation by dilated convolutions. <https://arxiv.org/abs/1511.07122>. doi:10.48550/arXiv.1511.07122

Zakeri, H., Nejad, F. M., Fahimifar, A., and Torshizi, A. D., (June 2013). “A multi-stage expert system for classification of pavement cracking,” in Proceedings of the Joint World Congress of the International-Fuzzy-Systems-Association and Annual Meeting of the North-American-Fuzzy-Information-Processing-Society (Edmonton, AB, Canada: IFSA/NAFIPS), 1125–1130.

Zakeri, E., Moezi, S. A., Bazargan-Lari, Y., and Mao, Q. (2013). Control of a ball on sphere system with adaptive feedback linearization method for regulation purpose[J]. *Majlesi J. of Mechatronic Engineering* 2 (3), 23–27.

Zhang, L., Yang, F., Zhang, Y., and Zhu, Y. (September 2016). “Road crack detection using deep convolutional neural network,” in Proceedings of the 23rd IEEE International Conference on Image Processing (Phoenix, AZ, USA: ICIP), 3708–3712.

Zou, Q., Cao, Y., Li, Q., Mao, Q., and Wang, S. (2012). Cracktree: Automatic crack detection from pavement images. *Pattern Recogn. Lett.* 33, 227–238. doi:10.1016/j.patrec.2011.11.004