

Semantic Segmentation Using Lightweight DeepLabv3+ for Desiccation Crack Detection in Soil [†]

Hui Yean Ling ¹, See Hung Lau ^{1,*}, Siaw Yah Chong ¹, Min Lee Lee ² and Yasuo Tanaka ¹

¹ Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kajang 43000, Selangor, Malaysia; hyling.99@utar.my (H.Y.L.); chongsiauw@utar.edu.my (S.Y.C.); yasuo@utar.edu.my (Y.T.)

² Department of Civil Engineering, University of Nottingham Malaysia, Semenyih 43500, Selangor, Malaysia; minlee.lee@nottingham.edu.my

* Correspondence: laush@utar.edu.my

[†] Presented at the 2024 IEEE 6th International Conference on Architecture, Construction, Environment and Hydraulics, Taichung, Taiwan, 6–8 December 2024.

Abstract: Soil desiccation cracks in natural clayey soil pose significant risks to the stability of civil and geotechnical structures. Traditional methods for detecting these cracks are often inefficient and prone to inaccuracies. Therefore, we applied a deep learning approach of semantic segmentation based on DeepLabv3+ to detect desiccation cracks. To enhance computational efficiency, a pretrained lightweight network, MobileNetV2, was employed as the backbone for the DeepLabv3+ model. The model was trained and tested on a dataset of natural clayey soil crack images obtained through laboratory tests. Evaluation metrics including precision, recall, F1 score, and intersection over union (IoU) were used to assess the segmentation performance. The model took 17.13 min to train and achieved an inference speed of 0.43 s per image. DeepLabv3+ achieved better performance than the traditional segmentation method, with a precision of 95.76%, a recall of 84.12%, an F1 score of 89.56%, and an IoU of 81.10%. The model also demonstrated the capability to handle images with shading conditions and the presence of spots. DeepLabv3+ with MobileNetV2 as a backbone network was proven to be effective and efficient as a backbone in soil desiccation crack detection and segmentation.

Keywords: semantic segmentation; soil desiccation crack; crack detection; deep learning; DeepLabv3+



Academic Editors: Teen-Hang Meen, Hsi-Hsien Yang and Cheng-Fu Yang

Published: 8 April 2025

Citation: Ling, H.Y.; Lau, S.H.; Chong, S.Y.; Lee, M.L.; Tanaka, Y. Semantic Segmentation Using Lightweight DeepLabv3+ for Desiccation Crack Detection in Soil. *Eng. Proc.* **2025**, *91*, 2. <https://doi.org/10.3390/engproc2025091002>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Desiccation cracks form when a soil mass is subjected to an arid climate due to evaporation and volumetric shrinkage [1,2]. The phenomenon is prevalent particularly in clayey soils because of its high tendency to shrink and swell in response to the change of moisture in the surroundings [3]. Desiccation cracking of natural clayey soil impacts the mechanical and hydraulic properties of the soil [4–7]. These detrimental effects are a significant concern in geotechnical and environmental engineering, for they can lead to slope instability [8], settlement and bearing capacity issues in foundations [2], and landfill leachate leakage [9].

With weakened stability and integrity of clayey soil and escalating climate change, accurate detection and analysis of these cracks in various applications are crucial for timely maintenance and prevention of potential failures. The analysis of soil cracks usually involves the quantitative measurement of geometrical characteristics such as crack intensity factor and crack width [5,10–12].

Early methods for analyzing soil crack networks relied on manual measurements, which are inefficient and inaccurate due to human errors [13]. The problem was greatly alleviated by implementing digital image processing technology. Crack network acquisition in this method is followed by grey scaling, binarization, and optional post-processing for denoising to segment the crack networks from the crack images [14–17]. However, the method relies heavily on human intervention, especially in fine-tuning binarization and post-processing processes and on the quality of crack images. The presence of disturbance, such as uneven illumination and rough soil surface, significantly affects the accuracy of crack network acquisition. This prompted the need for an advanced system that performs soil crack detection efficiently and accurately in all circumstances. Previous study results demonstrated the potential of a machine learning-based image processing model in the segmentation and quantification of soil cracks [18]. The model proved to be effective in handling uneven illumination while quantifying crack parameters with minimal processing time and satisfying accuracy, highlighting the promising use of machine learning.

In light of the successful applications of deep learning and computer vision technologies, researchers have applied deep neural networks to soil desiccation crack detection [19–21]. Han et al. [19] proposed a system combining crack localization and segmentation using a mask regional convolutional neural network (Mask R-CNN). The model detects cracks with instance segmentation to identify, localize, and generate individual labels for each crack in an image. Xu et al. [20] used an attention residual network-UNet for the semantic segmentation of the entire soil crack networks. Both studies showed the promising ability of deep learning methods with their high segmentation accuracy in crack detection. In addition, a comparative study was conducted on ground crack detection, and the results demonstrated the superiority of deep learning methods over traditional segmentation methods in segmentation accuracy and consistency [21].

Despite the plethora of image segmentation deep learning frameworks available, existing studies have only examined a limited subset of these tools. Models such as Mask R-CNN and attention Res-UNet, which were trained from scratch, demand extensive training times and significant computational resources. An overlooked framework in this domain is DeepLabv3+. This model employs an encoder–decoder architecture and incorporates the atrous spatial pyramid pooling (ASPP) module, delivering exceptional semantic segmentation performance [22–24].

In this study, a deep learning approach based on DeepLabv3+ was applied to soil desiccation crack semantic segmentation and detection. The pretrained lightweight MobileNetV2 network was employed as the backbone for the DeepLabv3+ model to minimize training and computational efforts. A soil crack dataset was prepared from a set of natural clayey soil crack images for the model's training, validation, and testing. Evaluation metrics including precision, recall, F1 score, and IoU were determined to evaluate the detection performance.

2. Materials and Methods

2.1. Data Preparation

Drying and wetting processes of natural clayey soil with high plasticity were conducted. In the experiments, soil crack images were taken under different photographic conditions, and the images were sliced into smaller patches for deep-learning model training. A total of 220 images at a resolution of 480×480 were generated and were randomly split at a ratio of 9:1 into training and validation datasets. Four unsliced images at the resolution of 960×960 were reserved for the testing of the model. All the images were manually annotated to produce binary ground truth masks with the foreground as white (1) and the background as black (0) pixels as shown in Figure 1. The training dataset was

processed with random augmentation methods for scaling, rotating, reflecting, translating, and hue/saturation/value (HSV) color jittering to prevent overfitting.



Figure 1. Image dataset. (a) Original image patch; (b) Binary mask.

2.2. DeepLabv3+ Architecture

The architecture of the DeepLabv3+ model for automatic semantic segmentation of soil desiccation crack networks is shown in Figure 2. An encoder–decoder structure is observed, where the encoder extracts and compresses features from the input image, and the decoder reconstructs these features to generate the segmented output. The encoder starts with a backbone network, in this case, a pretrained MobileNetV2, to extract both high- and low-level features to be fed to the ASPP module and decoder path, respectively.

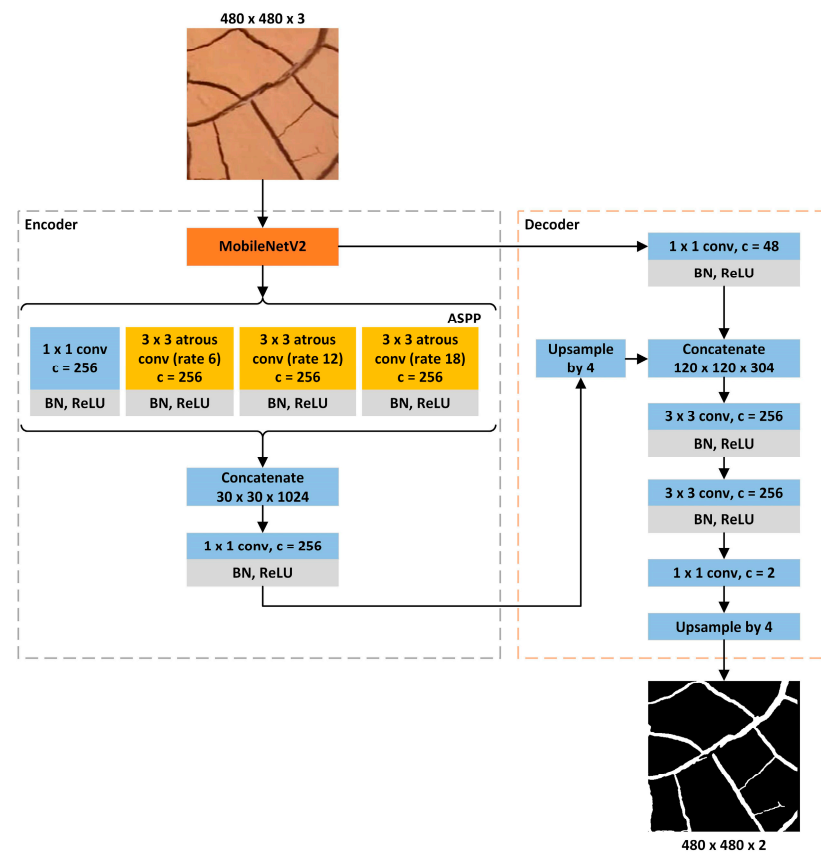


Figure 2. DeepLabv3+ architecture, c stands for number of output channels.

MobileNetV2 is a neural network that is well known for its lightweight and computational efficiency due to its depthwise separable convolution [25]. Its basic building block is an inverted residual block with a linear bottleneck structure (Table 1 and Figure 3). The network architecture and its sequences of MobileNetV2 used for DeepLabv3+ are shown in Table 2. In

the table, n represents the number of repetitions of a similar operator. At every level, the first block has a stride of s while the rest have a stride of 1. Despite having a stride of 1, the first blocks in sequences of $30^2 \times 64$ and $30^2 \times 96$ do not contain a residual connection.

Table 1. MobileNetV2 building block [25].

Input	Operator	Output
$h \times w \times k$	1×1 conv2d, BN, ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3×3 Depthwise $s = s$, BN, ReLU6	$h/s \times w/s \times (tk)$
$h/s \times w/s \times tk$	1×1 conv2d, BN	$h/s \times w/s \times k'$

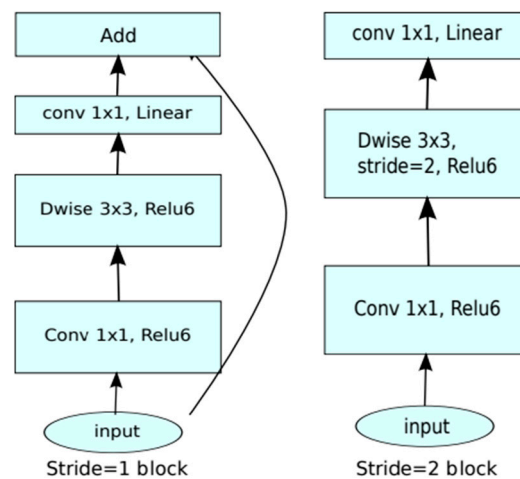


Figure 3. Building block configuration with different stride [25].

Table 2. Architecture of MobileNetV2 backbone network.

Input	Operator	t	c	n	s
$480^2 \times 3$	Conv2d	-	32	1	2
$240^2 \times 32$	Bottleneck	1	16	1	1
$240^2 \times 16$	Bottleneck	6	24	2	2
$120^2 \times 24$	Bottleneck	6	32	3	2
$60^2 \times 32$	Bottleneck	6	64	4	2
$30^2 \times 64$	Bottleneck	6	96	3	1
$30^2 \times 96$	Bottleneck	6	160	3	1
$30^2 \times 160$	Bottleneck	6	320	1	1

t = expansion factor; c = number of output channels; n = module repetitions; s = stride.

ASPP uses atrous separable convolutions (Figure 4) and spatial pyramid pooling to capture multi-scale features [22]. The ASPP module consists of 4 parallel branches with the same number of output channels (1×1 convolution; 3×3 convolutions with dilation rates of 6, 12, and 18); these outputs are concatenated and passed through another 1×1 convolution to generate encoded feature representations.

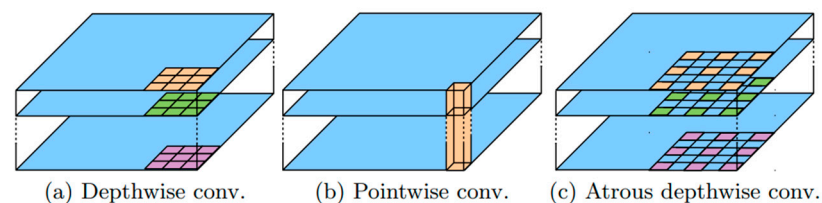


Figure 4. Example of atrous depthwise separable convolution at a dilation rate of 2 [22].

2.3. Model Training and Validation

Using the MATLAB (R2023b) Deep Learning Toolbox, the model was trained on an Nvidia RTX 3060 GPU with 12 GB of memory at a batch size of 4. The adaptive moment estimation (Adam) was adopted as an optimizer, and the MobileNetV2 backbone was initialized with weights obtained by pre-training on ImageNet to reduce the training effort. The initial learning rate during training was 10^{-4} and decayed at a factor of 0.1 every 15 epochs for 50 epochs. The validation dataset was utilized to validate the performance of the model at the end of every epoch.

The loss function was used to measure the difference between ground truth and model prediction and then update the network weights based on it. The generalized dice loss function was used and is expressed in (1).

$$L = 1 - \frac{2 \sum_{k=1}^K w_k \sum_{m=1}^M Y_{km} T_{km}}{\sum_{k=1}^K w_k \sum_{m=1}^M (Y_{km}^2 + T_{km}^2)} \quad (1)$$

where L is the measured loss for a prediction, K is the number of classes, M is the number of pixels in the image, Y is the predicted segmentation, T is the ground truth, and w_k is the class-specific weighting factor that is used to balance the contributions from each class. The factor is defined as follows:

$$w_k = \frac{1}{\left(\sum_{m=1}^M T_{km} \right)^2} \quad (2)$$

To prevent overfitting, L2 regularization is utilized with weight decay set to 0.0005. L2 regularization is expressed as follows:

$$\text{L2 Regularization} = \lambda \sum (w_i^2) \quad (3)$$

$$\text{Loss}_{\text{L2 regularized}} = \text{Loss}_{\text{original}} + \text{L2 Regularization} \quad (4)$$

where λ is the weight decay and w_i is the weights. The regularization term is added to the loss to form a regularized loss.

2.4. Evaluation Standards

Pixels generated by the model for prediction are categorized into the following: (i) true positives (TP), which indicate correct crack pixel identification; (ii) false positives (FP), which predict soil pixels as cracks; (iii) true negatives (TN), which correctly predict background pixels; and (iv) false negatives (FN), which indicate miss out of actual crack pixels. Then, the model segmentation performance is assessed by calculating precision, recall, F1 score, and IoU (5)–(8).

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

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (8)$$

In addition, Otsu's global thresholding, a widely used traditional method, was adopted for performance comparison. This method was employed to generate binary masks by optimizing a global threshold value based on the maximum of the between class-variance [26].

3. Results and Discussion

The computational efficiency of the model was evaluated in terms of training time and inference speed. The model required 17.13 min to train, demonstrating minimal training effort needed for DeepLabv3+ with pre-trained MobileNetV2 backbone fine-tuning. The model produced inferences on the test dataset at a speed of 0.43 s per image and 2.33 frames per second. Overall, the model achieved satisfactory computational efficiency with low training effort required and fast inference speed.

The overall segmentation performance of the traditional and DeepLabv3+ methods is shown in Table 3, and the segmentation masks are illustrated in Figure 5. DeepLabv3+ shows significantly better performance than the traditional method. The overall precision is increased substantially by 61.1% and overall IoU by 47.03%. The high recall value of the traditional method is not related to segmentation accuracy. The broad categorization of pixels to the foreground class is observed from its masks (Figure 5), which greatly reduces *FNs*. The high precision, F1 score, and IoU achieved by DeepLabv3+ indicate the model's ability to produce accurate segmentation with minimal mistakes and precise localization of cracks. The metrics show that DeepLabv3+ with MobileNetV2 is effective for soil desiccation crack image segmentation.

Table 3. Comparison of traditional and DeepLabv3+ segmentation performance.

Metrics	Traditional Method	DeepLabv3+
Precision (%)	34.66	95.76
Recall (%)	95.27	84.12
F1 score (%)	50.82	89.56
IoU (%)	34.07	81.10

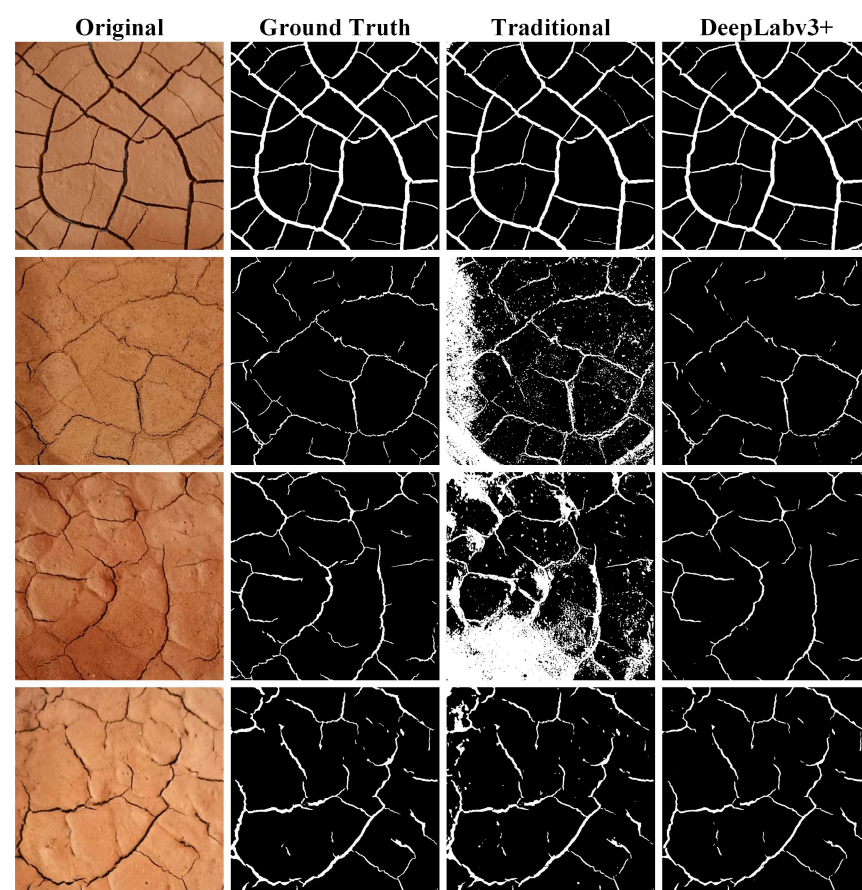


Figure 5. Segmentation mask on test images.

Figure 5 shows that the traditional method faced challenges in handling images with varied shading conditions, as evidenced by images in rows 2 and 3. The traditional method classifies shadowy regions and spots as cracks due to its inability to generalize spatial relationships at a higher level, resulting in a lack of flexibility in generating accurate segmentation results. Unlike the traditional method, the deep learning network captures feature representations from low to high levels, better distinguishing actual cracks from artifacts such as shadows and spots. This allows DeepLabv3+ to precisely segment images even when they contain shadows and spots.

Despite the high-pixel resolution, the images show blurry crack lines and edges. Several finer cracks were difficult to discern without closer inspection and human judgment. This hinders the accurate segmentation of soil cracks. A closer inspection shows that DeepLabv3+ effectively handles images with distinct crack networks, such as row 1, where crack edges are clearer (Figure 5). However, when dealing with blurry crack edges, the model presents challenges in differentiating fine cracks from the backgrounds, which negatively impacts the IoU performance. This suggested that while DeepLabv3+ is robust for most images, it still struggles with images containing indistinct crack boundaries, necessitating improvement using various techniques.

4. Conclusions

A novel approach was developed for the semantic segmentation and detection of soil desiccation cracks using a lightweight DeepLabv3+ model with a pretrained MobileNetV2 backbone. The model achieved high computational efficiency with a training time of 17.13 min and an inference speed of 0.43 s per image. A total of 220 images were used to train the model, and four were used for model testing. The model significantly improved segmentation accuracy compared with the traditional method, showing higher precision, F1 score, and IoU. The DeepLabv3+ model is capable of operating in different conditions and providing precise segmentation results even in the presence of shadows and spots. However, the model still needs to deal with blurry crack edges, especially blurry fine cracks. However, advanced deep learning models like DeepLabv3+, enhanced by pretrained lightweight networks such as MobileNetV2, efficiently and effectively detect soil cracks for the monitoring and management of soil desiccation cracks.

Author Contributions: Conceptualization, H.Y.L. and S.H.L.; methodology, H.Y.L. and S.H.L.; software, H.Y.L.; validation, H.Y.L., S.H.L., and S.Y.C.; formal analysis, H.Y.L., S.H.L., and Y.T.; investigation, H.Y.L., S.Y.C., and M.L.L.; resources, H.Y.L., S.H.L., and S.Y.C.; data curation, H.Y.L.; writing—original draft preparation, H.Y.L.; writing—review and editing, H.Y.L., S.H.L., and S.Y.C.; visualization, H.Y.L. and S.H.L.; supervision, S.H.L., S.Y.C., M.L.L., and Y.T.; project administration, H.Y.L., S.H.L., and S.Y.C.; funding acquisition, S.H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Higher Education (MOHE) Malaysia through the Fundamental Research Grant Scheme project (FRGS/1/2022/TK06/UTAR/02/36).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: The authors wish to thank Universiti Tunku Abdul Rahman for the facilities provided for the study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Kodikara, J.; Costa, S. Desiccation Cracking of soils: A review of investigation approaches, underlying mechanisms, and influencing factors. In *Multiphysical Testing of Soils and Shales*; Laloui, L., Ferrari, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 21–32. [\[CrossRef\]](#)
- Tang, C.S.; Zhu, C.; Cheng, Q.; Zeng, H.; Xu, J.J.; Tian, B.G.; Shi, B. Desiccation cracking of soils: A review of investigation approaches, underlying mechanism, and influencing factors. *Earth Sci. Rev.* **2021**, *216*, 103586. [\[CrossRef\]](#)
- Jones, L.D.; Jefferson, I. Expansive soils. In *ICE Manual of Geotechnical Engineering*; ICE Publishing: London, UK, 2012; Volume 1, pp. 413–441.
- Albrecht, B.A.; Benson, C.H. Effect of desiccation on compacted natural clays. *J. Geotech. Geoenviron.* **2001**, *127*, 67–75. [\[CrossRef\]](#)
- Cheng, Q.; Tang, C.S.; Xu, D.; Zeng, H.; Shi, B. Water infiltration in a cracked soil considering effect of drying-wetting cycles. *J. Hydrol.* **2021**, *593*, 125640. [\[CrossRef\]](#)
- Li, H.D.; Tang, C.S.; Cheng, Q.; Li, S.J.; Gong, X.P.; Shi, B. Tensile strength of clayey soil and the strain analysis based on image processing techniques. *Eng. Geol.* **2019**, *253*, 137–148. [\[CrossRef\]](#)
- Morris, H.; Graham, J.; Williams, D.J. Cracking in drying soils. *Can. Geotech. J.* **1992**, *253*, 263–277. [\[CrossRef\]](#)
- Wang, L.L.; Tang, C.S.; Shi, B.; Cui, Y.J.; Zhang, G.Q.; Hilary, I. Nucleation and propagation mechanisms of soil desiccation cracks. *Eng. Geol.* **2018**, *238*, 27–35. [\[CrossRef\]](#)
- Cheng, Q.; Tang, C.S.; Zeng, H.; Zhu, C.; An, N.; Shi, B. Effects of microstructure on desiccation cracking of a compacted soil. *Eng. Geol.* **2020**, *265*, 105418. [\[CrossRef\]](#)
- Miller, C.J.; Mi, H.; Yesiller, N. Experimental analysis of desiccation crack propagation in clay liners. *J. Am. Water Resour. Assoc.* **1998**, *34*, 677–686. [\[CrossRef\]](#)
- Tang, C.S.; Cui, Y.J.; Tang, A.M.; Shi, B. Experiment evidence on the temperature dependence of desiccation cracking behavior of clayey soils. *Eng. Geol.* **2010**, *114*, 261–266. [\[CrossRef\]](#)
- Zeng, H.; Tang, C.S.; Cheng, Q.; Lin, L.; Xu, J.J. Desiccation cracking behavior of soils. *Jpn. Geotech. Soc. Spec. Publ.* **2019**, *7*, 90–95. [\[CrossRef\]](#)
- Dasog, G.S.; Shashidhara, G.B. Dimension and volume of cracks in a vertisol under different crop covers. *Soil. Sci.* **1993**, *156*, 424–428. [\[CrossRef\]](#)
- Tang, C.S.; Shi, B.; Liu, C.; Zhao, L.Z.; Wang, B.J. Influencing factors of geometrical structure shrinkage cracks in clayey soils. *Eng. Geol.* **2008**, *101*, 204–217. [\[CrossRef\]](#)
- Liu, C.; Tang, C.S.; Shi, B.; Suo, W.B. Automatic quantification of crack patterns by image processing. *Comput. Geosci.* **2013**, *57*, 77–80. [\[CrossRef\]](#)
- Lu, Y.; Liu, S.; Weng, L.; Wang, L.; Li, Z.; Xu, L. Fractal analysis of cracking in a clayey soil under freeze-thaw cycles. *Eng. Geol.* **2016**, *208*, 93–99. [\[CrossRef\]](#)
- Singh, S.P.; Rout, S.; Tiwari, A. Quantification of desiccation cracks using image analysis technique. *Int. J. Geotech. Eng.* **2018**, *12*, 383–388. [\[CrossRef\]](#)
- Ling, H.Y.; Lau, S.H.; Chong, S.Y.; Lee, M.L.; Tanaka, Y. Quantifying desiccation cracks for expansive soil using machine learning technique in image processing. *Int. J. Integr. Eng.* **2024**, *16*, 8–15. [\[CrossRef\]](#)
- Han, X.L.; Jiang, N.J.; Yang, Y.F.; Choi, J.; Singh, D.N.; Beta, P.; Du, Y.J.; Wang, Y.J. Deep learning based approach for the instance segmentation of clayey soil desiccation cracks. *Comput. Geotech.* **2022**, *146*, 104733. [\[CrossRef\]](#)
- Xu, J.J.; Zhang, H.; Tang, C.S.; Cheng, Q.; Tian, B.G.; Liu, B.; Shi, B. Automatic soil crack recognition under uneven illumination condition with the application of artificial intelligence. *Eng. Geol.* **2022**, *296*, 106495. [\[CrossRef\]](#)
- Pham, M.V.; Ha, Y.S.; Kim, Y.T. Automatic detection and measurement of ground crack propagation using deep learning networks and an image processing technique. *Measurement* **2023**, *215*, 112832. [\[CrossRef\]](#)
- Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic segmentation. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018. [\[CrossRef\]](#)
- Fu, H.; Meng, D.; Li, W.; Wang, Y. Bridge crack semantic segmentation based on improved Deeplabv3+. *J. Mar. Sci. Eng.* **2021**, *9*, 671. [\[CrossRef\]](#)
- Nguyen, T.G.; Do, T.L.; Nguyen, T.N.; Nguyen, N.N. Semantic segmentation of cracks using DeepLabv3+. In Proceedings of the Third International Conference on Sustainable Civil Engineering and Architecture, Da Nang City, Vietnam, 19–21 July 2023; Reddy, J.N., Wang, C.M., Luong, V.H., Le, A.T., Eds.; Springer: Singapore, 2024; pp. 1539–1546. [\[CrossRef\]](#)

25. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [[CrossRef](#)]
26. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2008.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.