Proceedings of the Seventh International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 15, Paper 10.12
Civil-Comp Press, Edinburgh, United Kingdom, 2026
ISSN: 2753-3239, doi: 10.4203/ccc.15.10.12

Image-Based Crack Analysis of Railway Contact Wires Using a Two-Stage U-Net

Y. Yang[1], W. Liu[1], J. Liu[1] and T. Xin[2]

[1]Department of Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden
[2]Standard Metrology Institute, China Academy of Railway Sciences, Beijing, China

Abstract

The railway catenary system is essential for continuous electric power supply, and regular inspection is required to maintain safe and reliable long-term operation. As the component in direct contact with the pantograph, the contact wire is exposed to repeated mechanical and electrical loading, making surface cracking an important condition indicator. However, contact wire videos from operating railways are difficult to obtain. This paper therefore uses laboratory cyclic-loading videos as a controlled crack observation platform. We extract maximum-stroke keyframes to construct standardised contact wire keyframes for crack analysis. Standardised contact wire image regions are then segmented using a two-stage U-Net model, followed by mask cleaning, skeletonisation and pixel-level geometric measurement. The workflow provides measurable crack information for a safety-critical current-collection component, offering a basis for future real world contact wire inspection and maintenance decision support.

Keywords: railway contact wire, keyframe extraction, crack segmentation, skeletonization, geometric measurement, deep learning.

1 Introduction

The railway contact wire is a critical component of the pantograph–catenary system, and its condition directly affects current collection quality, maintenance planning, and operational safety. Among different surface damage forms, cracks and local defects are important visual indicators of contact-wire degradation [1]. Image-based analysis can segment visible crack regions and further convert them into measurable geometric indicators, such as crack length, width, and area.

Computer-vision and deep-learning methods have been increasingly used for catenary inspection. Existing studies have investigated automatic fastener defect detection on catenary support devices and surface defect detection of high-speed railway insulators [2,3]. These works demonstrate the potential of automated visual inspection for railway maintenance. However, they mainly focus on component localisation or defect-state recognition of catenary fittings, while pixel-level contact-wire crack segmentation and subsequent geometric quantification are still less explored.

Reliable crack data from real railway lines are difficult to collect because contact wires operate in open environments with complex conditions. Therefore, controlled laboratory data provide a practical intermediate step for developing and validating crack analysis methods before field deployment. Similar strategies have been adopted in infrastructure inspection, where fatigue-test photographs or laboratory specimen images were used for crack detection in steel bridges, concrete, and cementitious composites [4,5,6]. In the contact-wire domain, laboratory experiments have also been used to study fatigue crack initiation under combined mechanical and electrical loading [7]. These studies support the use of laboratory cyclic-loading videos as a controlled data source for contact-wire crack analysis.

Motivated by this, this study develops a visual analysis workflow for contact-wire surface cracks using laboratory cyclic-loading videos. First, a maximum-stroke keyframe extraction strategy is proposed to select representative frames and construct standardised 640 × 640 contact-wire image patches. Second, a two-stage ResNet34-U-Net model with attention-gated skip fusion is used to segment visible crack candidates in the extracted keyframes [8,9]. Third, a post-processing and skeleton-based quantification procedure is introduced to convert crack masks into pixel-level geometric indicators for estimating the relative damage degree of the specimen. With physical calibration, these pixel-level measurements can be further converted into actual crack length, width, and area. The proposed workflow provides a baseline model, annotation target, and descriptor format for future real-world contact-wire inspection datasets.

2 Laboratory Settings

In the laboratory validation step, a bending machine is used to apply repeated local bending to a contact wire and to generate visible surface cracking under controlled observation. Figure 1 shows the laboratory setup. The workpiece is fixed on the bending machine, and is repeatedly bent between two terminal stroke positions, with a motion period of 2s. During the test, the contact wire is repeatedly bent at the workpiece until fracture, and the full bending process is recorded by two cameras deployed on both sides of the machine. In our tests, two iPhone 16 Pro Max devices are used, each of which has a 48-megapixel sensor, and all videos are recorded at 4K resolution and 60 frames per second, providing high quality image sequences for crack candidate analysis. To improve the robustness of the subsequent keyframe extraction and crack candidate segmentation, videos were collected under different lighting conditions and at different shooting distances.

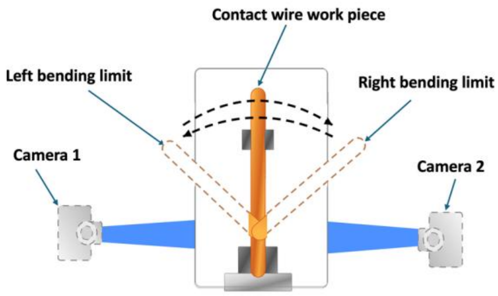

Figure 1: Overview of the laboratory setup.

3 Methods

Figure 2 illustrates the proposed workflow. Raw laboratory videos are first converted into maximum-stroke frames to form the dataset. A two-stage attention-gated ResNet34-U-Net is then used to segment the extracted frames. Finally, the predicted masks are cleaned, skeletonised and converted into pixel-level geometric indicators.

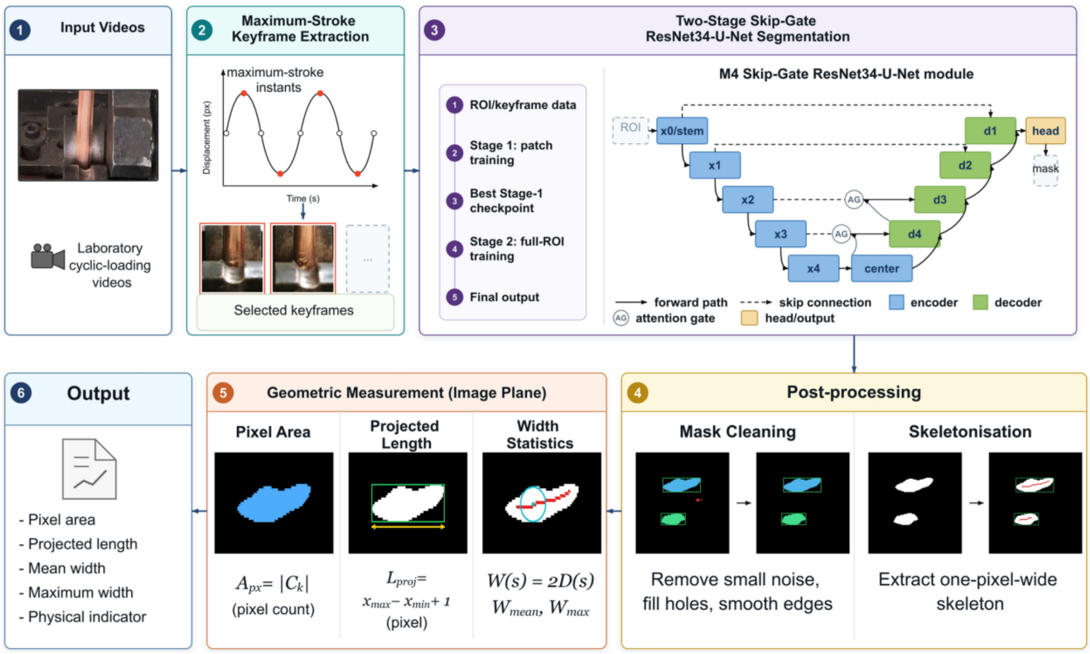

Figure 2: Overview of the proposed workflow from cyclic-loading videos to keyframe extraction, crack-candidate segmentation, post-processing and image-plane geometric measurement.

3.1 Maximum-stroke position keyframe extraction

In this paper, the maximum-stroke states refer to the two bending-limit positions of the contact wire in each loading cycle, as illustrated in Figure 1. This is to make the crack appearance more consistent between frames and reduce the effect of motion blur. Two reference frames, $t_{ref,1}$ and $t_{ref,2}$, are selected from these bending-limit states and used to identify repeatable maximum-stroke frames in the video.

For each video frame I_t, $t=1,...,T$, a contact wire region of interest (ROI) and a pose ROI are defined, as shown in Figure 3. In the contact wire ROI, hue, saturation and value colour thresholding is used to retain contact-wire pixels. The raw vertical position of the wire is then estimated as the mean vertical coordinate of the retained pixels:

$$d_{raw}(t) = \frac{1}{|M_t|} \sum_{(x,y)\in M_t} y, \tag{1}$$

where M_t is the retained contact-wire pixel set in frame t and $|M_t|$ is the number of retained pixels. Missing values are interpolated and the sequence is smoothed by a moving-average operator, giving the position signal d(t). A position similarity score $s_{pos,i}(t)$ is computed from the normalised distance between d(t) and the reference position $d(t_{ref,i})$.

The position cue is complemented by a pose cue because reflections, colour noise and partial occlusion can disturb the position estimate. The pose ROI is cropped, converted to greyscale, resized and normalised, and then compared with the two reference pose crops using normalised template correlation. The resulting pose similarity $s_{pose,i}(t)$ is smoothed and rescaled to [0,1]. The final score for reference state i is a linear fusion of the position and pose similarities:

$$F_i(t) = \alpha_s\, s_{pos,i}(t) + \alpha_p\, s_{pose,i}(t), \quad i\in\{1,2\}. \tag{2}$$

For each score curve, frames above the quantile threshold $\theta_i=Q_{0.85}(F_i)$ are merged into candidate intervals. For a retained interval $[t_s,t_e]$, the extracted keyframe is

$$t_i^{\star} = \arg\max_{t\in[t_s,t_e]} F_i(t). \tag{3}$$

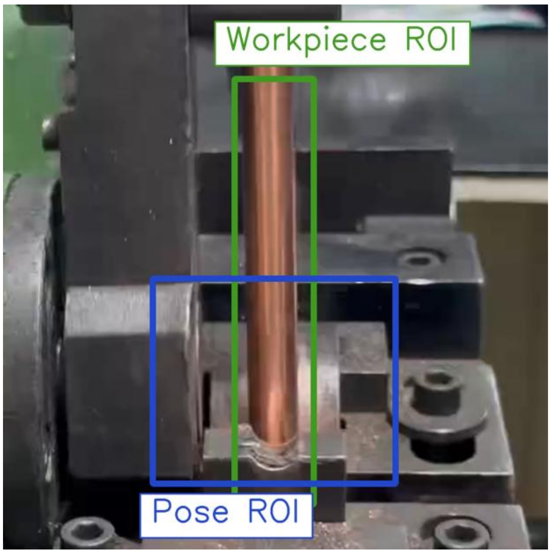

Figure 3: Definition of the workpiece ROI and pose ROI for maximum-stroke keyframe extraction.

3.2 Attention-gated crack candidate segmentation

After keyframe extraction, pixel-level segmentation is used to identify visible crack candidate regions. A ResNet34-U-Net is adopted as the segmentation backbone, and the training stage is illustrated in Figure 4.

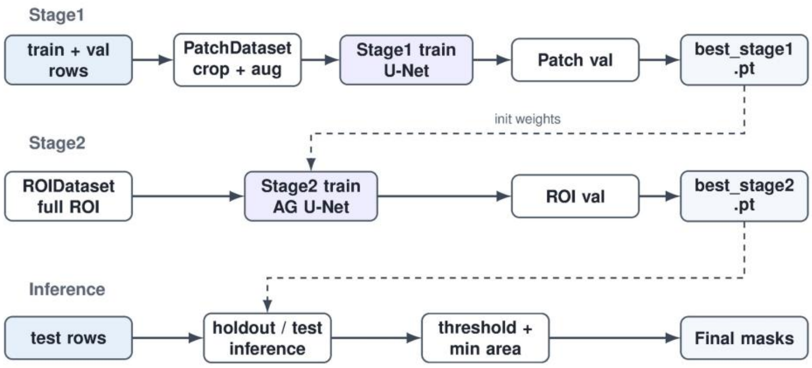

Figure 4: Flowchart of the two-stage training pipeline.

The two-stage strategy addresses the small pixel proportion of crack candidates in the contact wire images. In Stage 1, the network is trained on positive and negative patches so that the appearance of local cracks, and boundary signals are learned with a higher occurrence of crack-pixel [10]. Hard normal replay is used to repeatedly expose the model to metallic texture, surface marks and fixture-related backgrounds that resemble cracks. In Stage 2, the best Stage 1 checkpoint is transferred to full-image training. Attention gates are added to the decoder skip-fusion paths at levels d4 and d3, using decoder context to modulate encoder features before fusion [11]. Hard normal replay is also used during full-image training as a false-positive control mechanism.

3.3 Conversion from segmentation masks to geometric indicators

The final module converts each binary crack-candidate mask into component-level and image-level geometric descriptors, which is designed to move from crack localisation to damage assessment. Small isolated regions are removed, and the remaining 8-neighbour connected components are denoted as $C_1,C_2,...,C_n$. For each component, the pixel area is $A(C_k)=|C_k|$. Since the contact wire is approximately aligned with the horizontal image axis in the laboratory images, the projected length and its relative ratio are computed as

$$L_{proj}(C_k)=x_{max,k}-x_{min,k}+1. \tag{4}$$

$$L_{wire}=x_{wire,right}-x_{wire,left}+1. \tag{5}$$

$$R_{proj}(C_k)=\frac{L_{proj}(C_k)}{L_{wire}}. \tag{6}$$

The width statistics are estimated from the medial-axis skeleton. For each component skeleton S_k, the Euclidean distance transform gives the local half-width, and the local width is estimated as [12]

$$D(p)=\min_{q\notin C}\|p-q\|_2, \quad W(s)=2D(s), \quad s\in S_k, \tag{7}$$

where $C=\cup_k C_k$. The mean and maximum widths are calculated from W(s) along the skeleton. Component-level measurements are finally aggregated into image-level descriptors, including the number of crack candidate components, total candidate area, maximum projected length, maximum projected length ratio and maximum width. All length and width measurements in this paper are reported in pixel units, while the projected length ratio is dimensionless.

4 Results and discussion

This section presents the experimental results of the proposed workflow. It first reports the construction of standardised crack-candidate keyframes, then evaluates the selected crack segmentation model, and finally analyses the skeleton-based geometric indicators extracted from the predicted crack masks.

4.1 Construction of standardised crack-candidate keyframes

Figure 5 shows an example of the maximum-stroke keyframe extraction result. The fused scores of the two reference maximum-stroke states follow the cyclic loading pattern and identify repeatable bending-limit frames. These selected frames are used as standardised crack-candidate image samples for subsequent segmentation and geometric measurement.

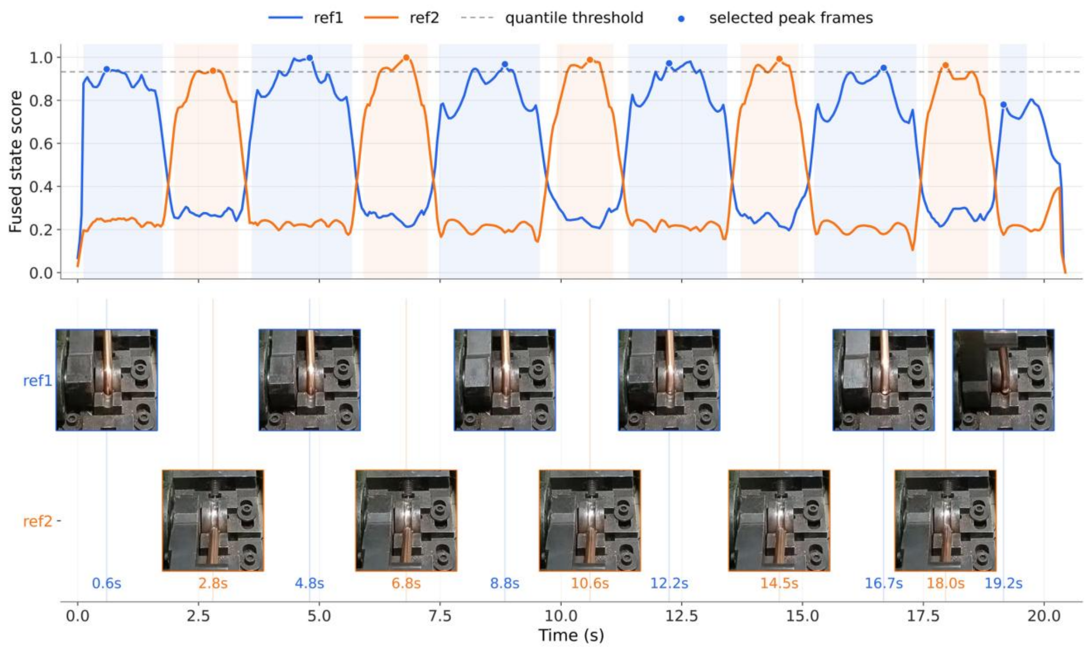

Figure 5: Fusion scores and corresponding keyframes for the ref1 and ref2 states.

4.2 Crack-candidate segmentation using the selected U-Net module

Segmentation performance is evaluated using Dice, intersection over union (IoU) and recall, which are commonly used pixel-level segmentation metrics [13]. Table 1 shows that the selected attention-gated two-stage U-Net gives the best segmentation performance among the tested variants. Compared with the single-stage U-Net, the two-stage transfer strategy improves Dice, IoU and recall on test sets, indicating that patch-level pre-training helps the model learn the local appearance of small crack regions before full-ROI training. Adding attention gates at the d4 and d3 skip-fusion levels further improves the scores. The attention-gated two-stage U-Net is therefore used as the selected segmentation module for the subsequent geometric measurement step.

Model	Setting	Dice	IoU	Recall
Single-stage U-Net	Full images	0.412	0.305	0.400
Two-stage U-Net	Patches and Full images	0.544	0.439	0.529
Attention-gated two-stage U-Net	Patches and Full images	**0.617**	**0.504**	**0.633**

Table 1: Performance of crack candidate segmentation models on the test set.

4.3 Skeleton-based crack geometry measurement

Figure 6 shows a representative post-processing example for a predicted mask. Mask cleaning removes one small isolated component and retains two main crack-candidate components for measurement.

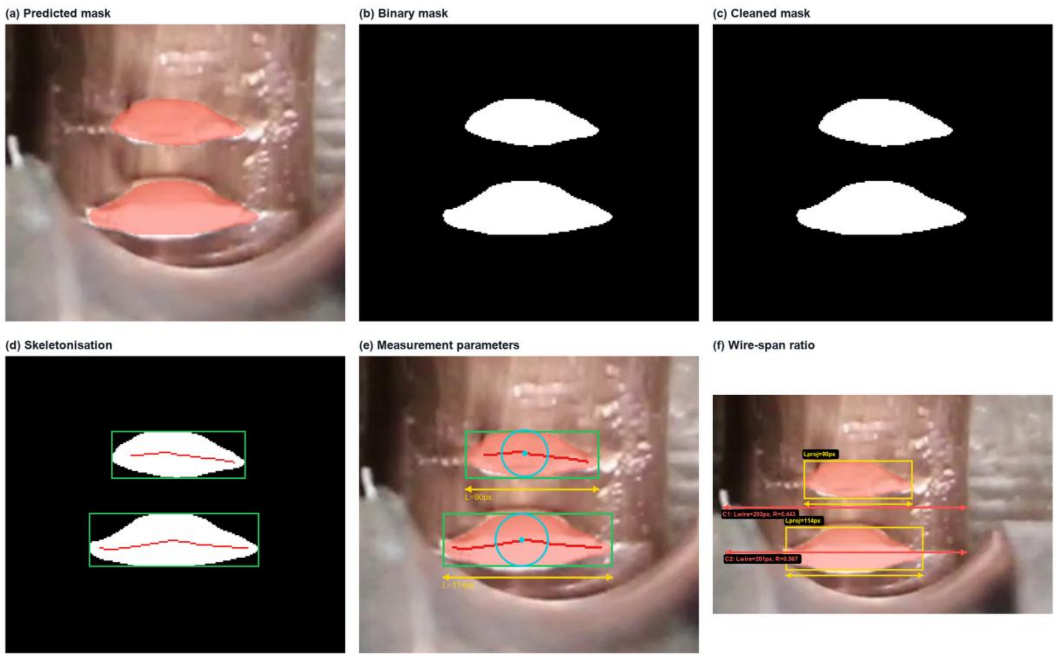

Figure 6: Representative post-processing result for a predicted crack candidate mask.

The resulting component-level and image-level geometric indicators are listed in Table 2. In this example, the retained components have projected lengths of 90 pixels(px) and 114 px, corresponding to length ratios of 0.443 and 0.567, respectively, which indicates the extent of damage to the contact-wire workpiece.

Component	Area (px)	Projected length (px)	Wire span (px)	Length ratio	Mean width (px)	Max width (px)
C1	1979	90	203	0.443	23.57	32.00
C2	2805	114	201	0.567	24.57	36.00
Summary	4784	114	–	0.567	–	36.00

Table 2: Example geometric indicators for the representative post-processing result.

5 Conclusions

This paper presented a video-to-indicator workflow for crack candidate analysis of railway contact wires under laboratory cyclic loading. The workflow converts cyclic-loading videos into standardised maximum-stroke keyframes. It then segments visible crack candidate regions with a two-stage attention-gated ResNet34-U-Net and extracts image-plane geometric indicators using skeleton-based measurement.

The results support the feasibility of this workflow. The keyframe selection step provides standardised crack-candidate images for subsequent analysis. Among the tested segmentation variants, the attention-gated two-stage U-Net achieves the best overall performance, with test Dice, IoU and recall values of 0.617, 0.504 and 0.633, respectively. The post-processing and skeleton-based measurement steps further convert predicted masks into descriptors including area, projected length, projected length ratio and width. In particular, the projected length ratio provides a relative damage-degree indicator under image-plane analysis, while the retained pixel-level measurements provide the basis for future conversion to physical damage dimensions after calibration.

The practical significance of the work is that it provides a bridge between controlled laboratory crack observation and future real-world contact wire inspection. These outputs can support future field data annotation, model pre-training, algorithm benchmarking and maintenance oriented interpretation once real inspection images are collected. Future work will extend to physical-unit calibration, and validate the segmentation and measurement modules under realistic field inspection conditions.

Acknowledgement

The authors would like to thank Yingjie Dong, Qun Yu and Yan Xu for the support in on-site measurement and data processing. The project is supported by the Research Project of China Academy of Railway Sciences Corporation Limited (2024YJ281) and the Research Project of China National Railway Administration (SJ2026-079).

References

[1] B. Olivier, F. Guo, Y. Qian, D.P. Connolly, "A Review of Computer Vision for Railways", *IEEE Transactions on Intelligent Transportation Systems*, 26(7), 11034–11065, 2025, DOI:10.1109/TITS.2025.3552011.

[2] J. Chen, Z. Liu, H. Wang, A. Núñez, Z. Han, "Automatic Defect Detection of Fasteners on the Catenary Support Device Using Deep Convolutional Neural Network", *IEEE Transactions on Instrumentation and Measurement*, 67(2), 257–269, 2018, DOI:10.1109/TIM.2017.2775345.

[3] G. Kang, S. Gao, L. Yu, D. Zhang, "Deep Architecture for High-Speed Railway Insulator Surface Defect Detection: Denoising Autoencoder With Multitask Learning", *IEEE Transactions on Instrumentation and Measurement*, 68(8), 2679–2690, 2019, DOI:10.1109/TIM.2018.2868490.

[4] C.V. Dung, H. Sekiya, S. Hirano, T. Okatani, C. Miki, "A Vision-Based Method for Crack Detection in Gusset Plate Welded Joints of Steel Bridges Using Deep Convolutional Neural Networks", *Automation in Construction*, 102, 217–229, 2019, DOI:10.1016/j.autcon.2019.02.013.

[5] A.K. Das, C.K.Y. Leung, K.T. Wan, "Application of Deep Convolutional Neural Networks for Automated and Rapid Identification and Computation of Crack Statistics of Thin Cracks in Strain Hardening Cementitious Composites (SHCCs)", *Cement and Concrete Composites*, 122, 104159, 2021, DOI:10.1016/j.cemconcomp.2021.104159.

[6] J.A. Guzmán-Torres, M.Z. Naser, F.J. Domínguez-Mota, "Effective Medium Crack Classification on Laboratory Concrete Specimens via Competitive Machine Learning", *Structures*, 37, 858–870, 2022, DOI:10.1016/j.istruc.2022.01.061.

[7] Ö. Sunar, D. Fletcher, "Experimental Investigation on the Arc Damage and Fatigue Crack Initiation Risk of Copper-Silver Contact Wires", *IEEE Transactions on Power Delivery*, 38(2), 777–784, 2023, DOI:10.1109/TPWRD.2022.3198734.

[8] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 234–241, 2015, DOI:10.1007/978-3-319-24574-4_28.

[9] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, 2016, DOI:10.1109/CVPR.2016.90.

[10] Y. Liu, J. Yao, X. Lu, R. Xie, L. Li, "DeepCrack: A Deep Hierarchical Feature Learning Architecture for Crack Segmentation", *Neurocomputing*, 338, 139–153, 2019, DOI:10.1016/j.neucom.2019.01.036.

[11] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker, D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas", arXiv:1804.03999, 2018, DOI:10.48550/arXiv.1804.03999.

[12] S. Jin, S.E. Lee, J.-W. Hong, "A Vision-Based Approach for Autonomous Crack Width Measurement with Flexible Kernel", *Automation in Construction*, 110, 103019, 2020, DOI:10.1016/j.autcon.2019.103019.

[13] A.A. Taha, A. Hanbury, "Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool", *BMC Medical Imaging*, 15(1), 29, 2015, DOI:10.1186/s12880-015-0068-x.