COMPUTER VISION-BASED INSPECTIONS OF CIVIL INFRASTRUCTURE

Piotr TAUZOWSKI[1], Piotr JAROSIK[1], Mateusz ŻARSKI[2], Bartosz WÓJCIK[2], Mariusz OSTROWSKI[1], Bartlomiej BLACHOWSKI[1], Lukasz JANKOWSKI[1]

[1] Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego St. 5B, Warsaw 02-106, Poland
[2] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Baltycka St. 5, 44-100 Gliwice, Poland

Corresponding author: ptauzow@ippt.pan.pl

Abstract. *The uNET neural network architecture has shown very promising results when applied to semantic segmentation of biomedical images. The aim of this work is to check whether this architecture is equally applicable to semantic segmentation distinguishing the structural elements of railway viaducts. Artificial images generated by a computer graphics program rendering the 3D model of the viaduct in a photorealistic manner will be used as data sets. This approach produces a large number of images that provide a solid training set for machine learning model.*

Keywords

Computer vision, deep learning semantic segmentation.

1. Introduction

Maintaining the traffic infrastructure in good technical condition is a key element of rail and road traffic safety. The expansion of the traffic infrastructure increases the problems related to the maintenance of facilities. Therefore, there is a need for intensive monitoring of the technical condition of a large number of objects. The intensive development of vision techniques together with artificial intelligence allows for effective automation of the inspection process of traffic infrastructure facilities. Installed cameras, as well as remotely controlled or autonomous vehicles or drones can independently monitor by capturing the image. The next step is, image processing by artificial intelligence systems based on deep learning which recognizes whether a structure has some damages or not. To carry out the task, it is necessary to train deep neural networks capable of recognizing damage in the image. It is also necessary to distinguish parts of the structure in the image in order to properly locate the damage as well as to find out on how important the structural element they occur.

The aim of this study is to train a neural network model capable to distinguish structural components of a railway viaduct such as: non-bridge (*everything on the picture which not belong to the bridge*), slab, beam, column, nonstructural components (poles, cables, fences), rail, sleeper, other components. This task will be carried out using the method called semantic segmentation.

The first application of semantic segmentation with uNET was in the field of medical image analysis [1]. Ronneberger applied semantic segmentation to images of HeLa cells for distinguishing particular instances of the cells. In the structural monitoring application, other convolutional neural network architectures are most often used [4]. Wu et all. [5] applied uNET architecture for building damage detection. Hu et al. [6] performed segmentation of cracks in the road surface using the uNET convolutional network. In this paper we would like to examine if the uNET architecture is also suitable for application to structural monitoring and semantic segmentation of structural parts of the bridges.

An in-depth review of the state-of-the-art achievements in the field of semantic segmentation can be found in [3].

2. The Data Analysis

To train a neural network model, a substantial set of training and verification data is needed. This task uses a set of over 7,000 images artificially generated images of viaducts with application of computer graphics based on 3D models [1]. The data also includes a depth channel and eight channels with a segmentation mask. The number of masks corresponds to the number of types of structural objects being detected in the images and indicates where these objects are located in the images. This data forms the basis for neural network training.

The images contains a 3D model of a straight section of the railway viaduct in two versions: one with crossbeams between the columns and the other without it. The structural elements are also textured with images, most often concrete, but there are also stone textures. The ground is also textured with pictures from wood to stubble and sometimes there are even such strange textures difficult to determine what it is. There are many (probably about 100) versions of this scene with different textures for each version, several dozen shots of the viaduct are taken, most often along the axis. Interestingly, there is no intake perpendicular to the viaduct in the pledge. At a certain distance from the viaduct (approximately equal to the height of the viaduct), vertical and parallel to the viaduct panels are placed, on which various urban, landscape or even graffiti photos are placed, which are to represent various real locations of the viaduct. Below are some examples of the different texture versions (Fig. 1)

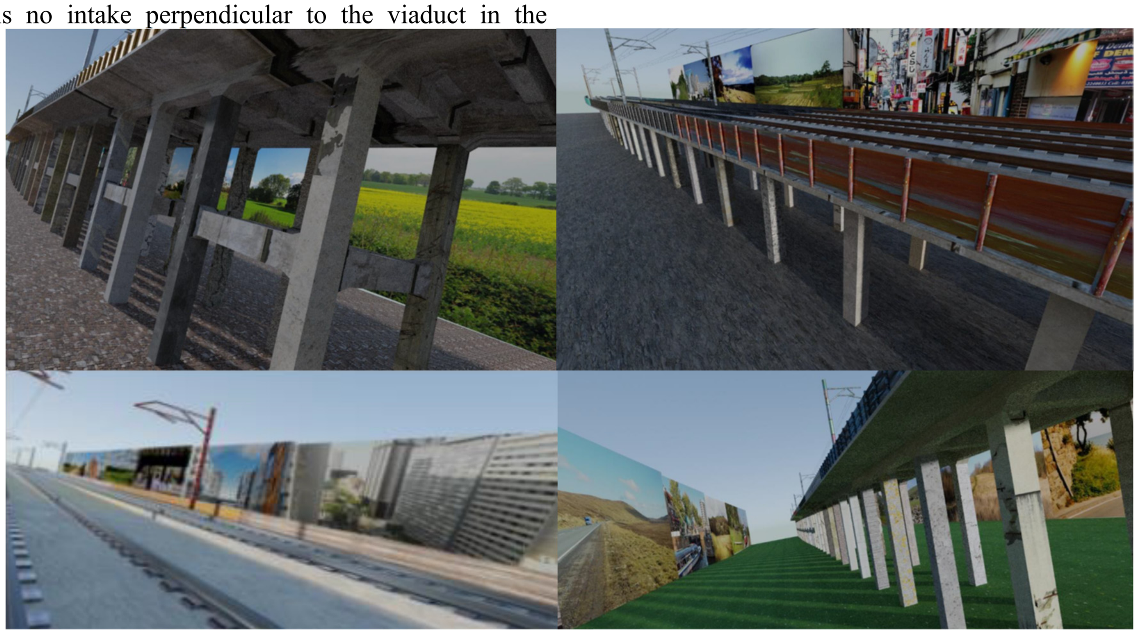

Fig. 1: Development of the variable z (unit) depending on the value of x (unit) and y (unit) in different graphic displays (a) Surf, (b) Plot3 and (c) Mesh.

In addition to the visual inspection of the data which has a qualitative nature, certain numerical characteristics should be established. For the multi-class semantic segmentation problem, one of the basic values is the empirical distribution of pixels belonging to individual classes. The histogram illustrating the pixel distribution for the presented images of viaducts images is depicted in the Fog. 2. The histogram shows that the most pixels belong to the nonbridge class. This is explained by the fact that most shots of the viaduct are taken from middle distance and include the background, sky or ground. Images of details of the structure are in the minority. As for structural elements such as beams, slabs or columns, with a slight predominance of columns. The reason for this quantitative advantage is the fact that the columns are a more exposed part of the structure than the slabs or beams and therefore take up more pixels in the photos. The least pixel-occupying objects are non-structural elements such as nonstructural elements (poles, cables, fences), tracks or railway sleepers. These elements are small in comparison with structural elements and, there are not many number of images of the track in the data set. Therefore the worst quality segmentation is observed on the images wit abovementioned nonstructural elements of railway infrastructure.

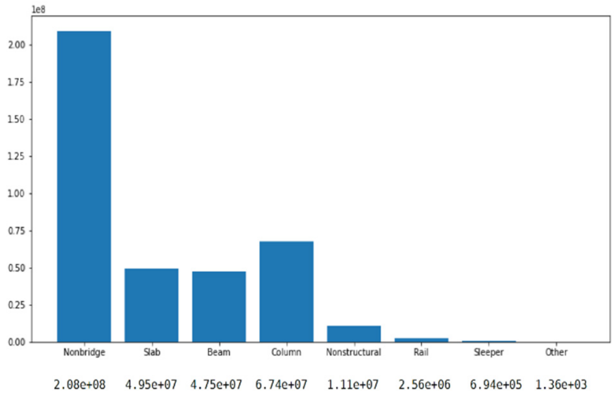

Fig. 2: Empirical pixel distribution of individual classes in particular segmentation masks

3. METHODOLOGY

The problem of multi-class semantic segmentation will be solved with application of neural network. The most effective in image processing are convolutional neural networks because they can learn to recognize certain patterns regardless of their location in the image. In this project, an architecture called uNET so called fully convolutional network is used. Diagram of the architecture is presented on the Fig 3. The issue was resolved in the TensorFlow environment using the Keras package. This package includes a ready-made implementation of the uNET network. This convolutional layer architecture was developed by Ronneberger et al. [2] for the purpose of biomedical image segmentation.

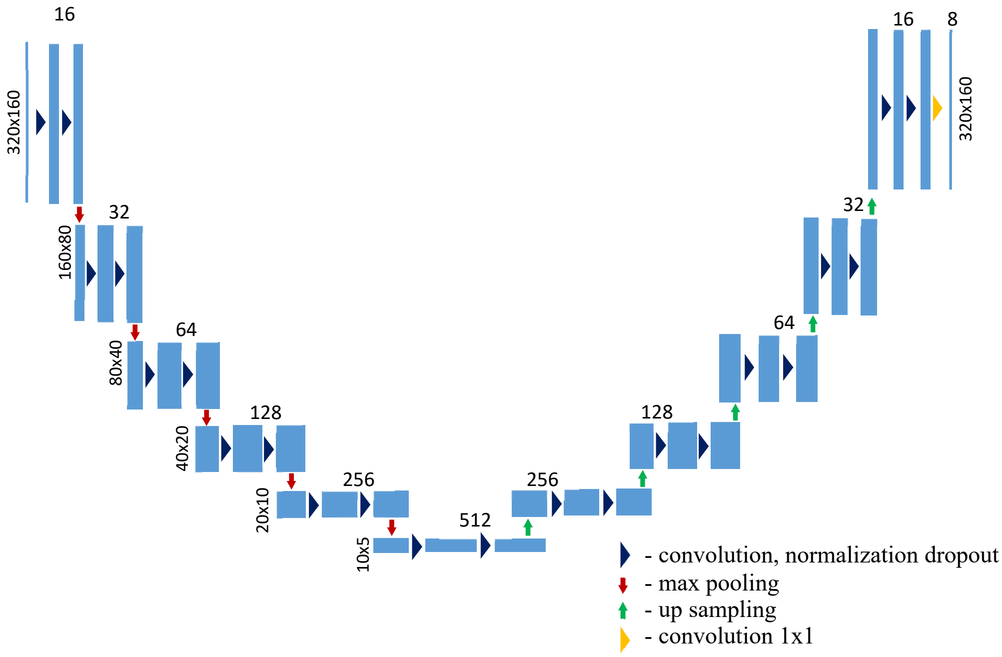

Fig. 3: Diagram of uNET version neural network implemented in *keras_unet* package.

4. The Results

Two uNET models were trained. In the first one, only RGB images were used, i.e. three channels were used as input. The second model additionally contained the depth channel which is part of the data set. The network learning process consisted of 70 epochs. The learning process of neural networks is shown in the plots below (Fig. 4).

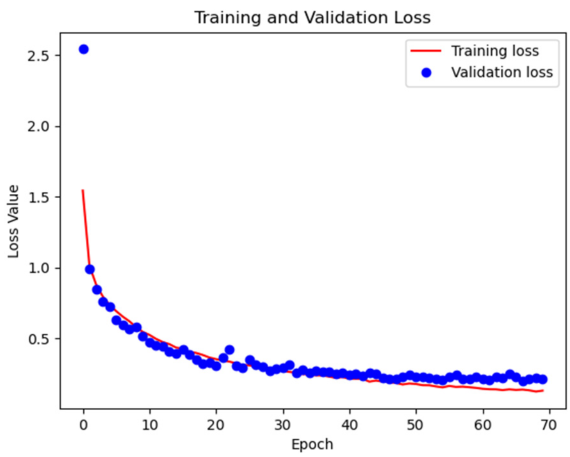

Fig. 4: Training process of the uNET model

The figure shows the beginning of the overfitting process. Therefore, we believe that the use of more than 70 epochs has no rational justification. In the first, the 2D model, categorical accuracy is equal to 0. 9628 while the loss is equal 0. 0.0971. In the three-dimensional model, significantly better results was achieved. Categorical accuracy was 0.9760 while the loss was equal to 0. 0631. Of course, in practice, when inspection images are captured, they are not made in a stereoscopic version, hence they will not have a depth channel. In the examples below, several chosen examples of semantic segmentation has been presented. Especially in the first example (Fig 5) we will check the model's ability to distinguish one of the most difficult structural elements of beams and slabs. The elements are most often located in the shaded part of the bridge, they do not differ in texture and their distinguishing in the photo may even be difficult for a person who is not a civil engineer. The results below show that the most challenging part is to distinguish beams and slabs. A similar examples in which we deal with the difference between beams and boards is shown in Fig. 6,7. The test example additionally includes a railing and a bitmap with a city photo in the background. Our model distinguishes the elements on it correctly. As mentioned before. Semantic segmentation is the most difficult in images containing elements of railway infrastructure, such as tracks, overhead contact lines.

These elements are recognized by the model but some inaccuracies can be noticed (Fig 8).

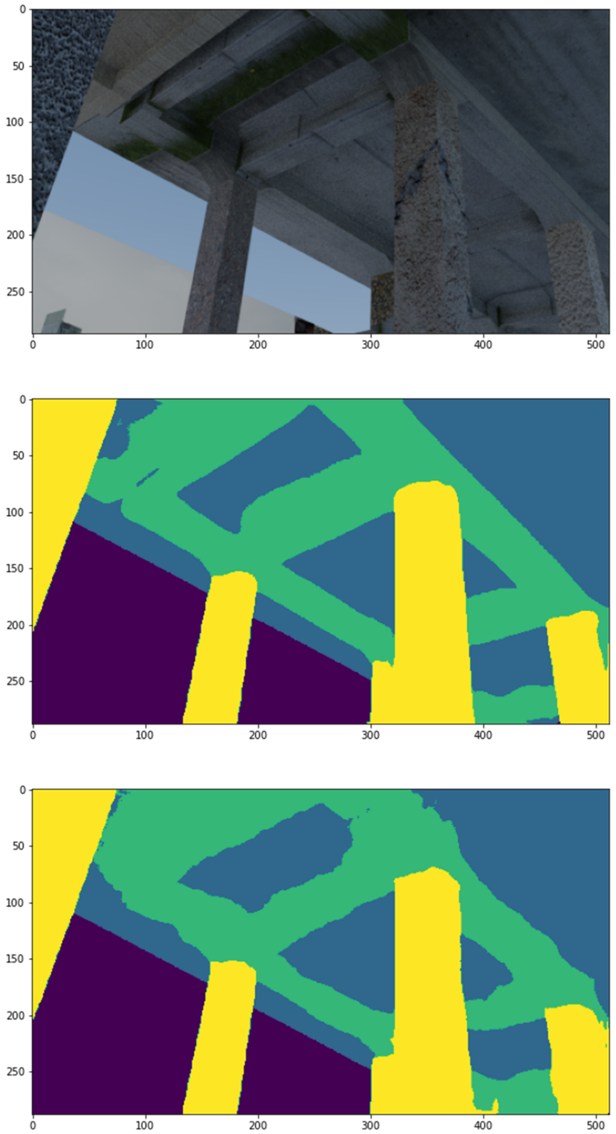

Fig. 5: Test to distinguish between beams and slabs upper image: original image, middle image reference model with depth field, bottom image: target model results

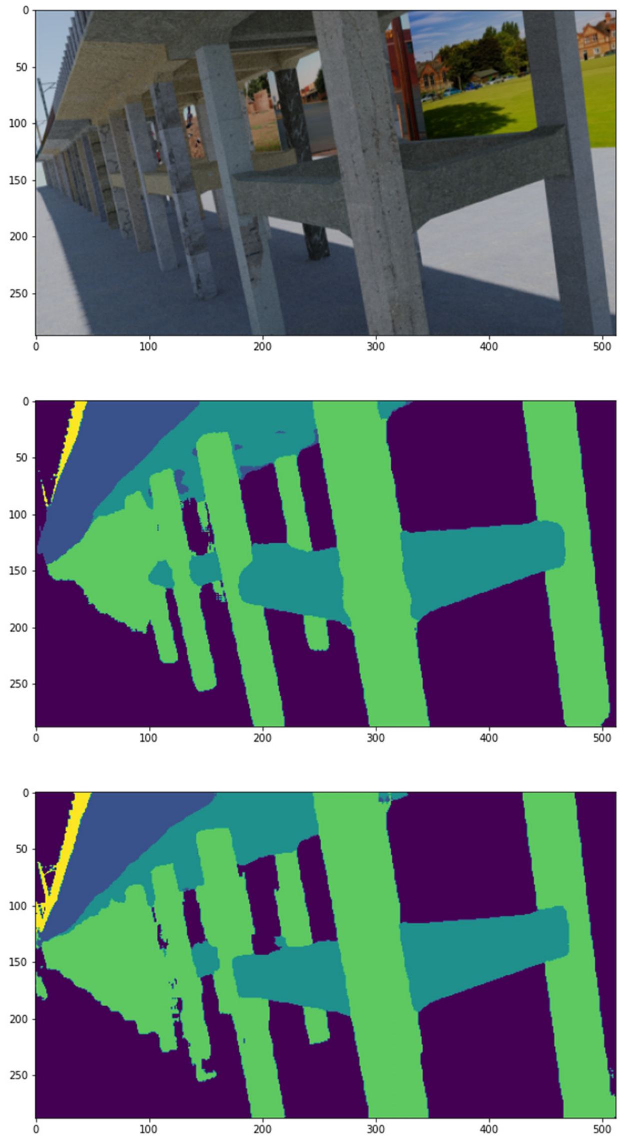

Fig. 6: Test to distinguish structural and nonstructural elements from the background. Upper original image, middle image reference model with depth field, bottom image: target model results

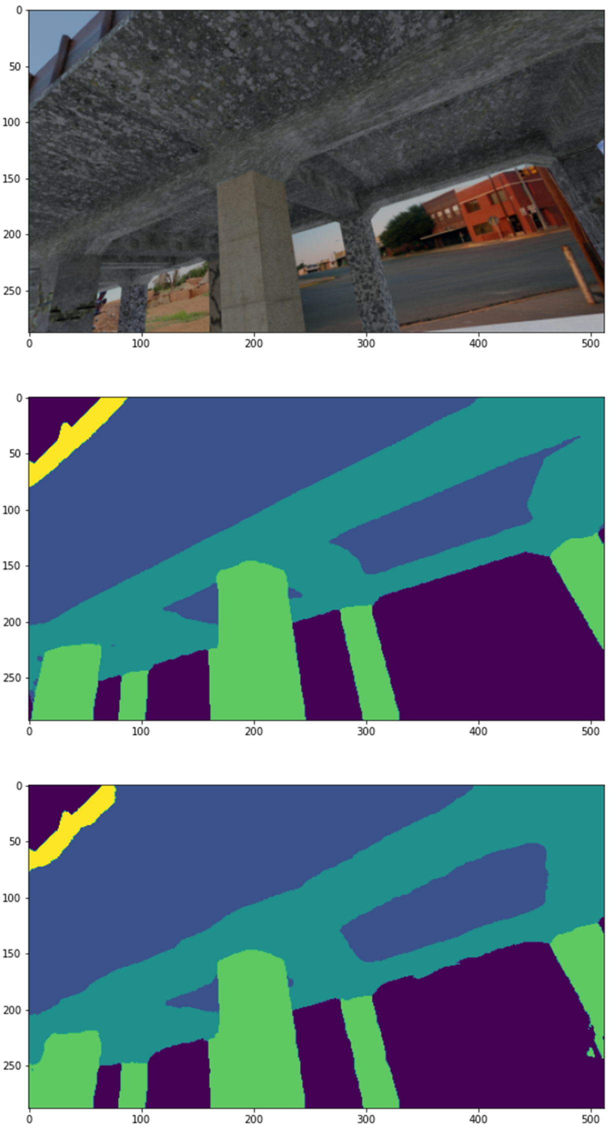

Fig. 7: Test to distinguish structural and nonstructural elements railing (yellow). Upper original image, middle image reference model with depth field, bottom image: target model results

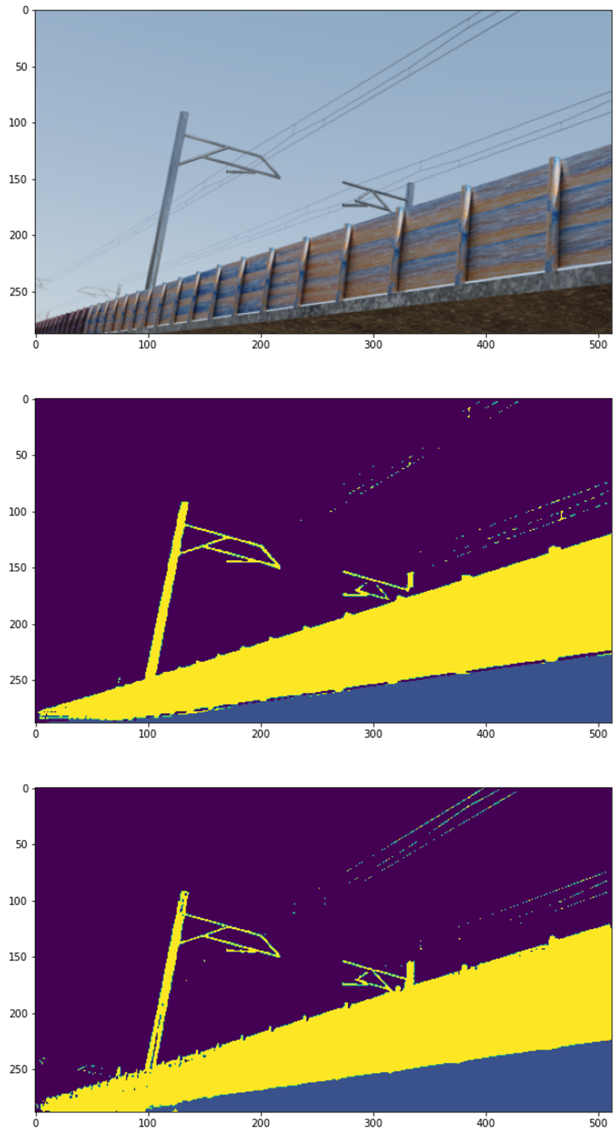

Fig. 8: Test to distinguish nonstructural railway lines of the overhead contact line. Upper original image, middle image reference model with depth field, bottom image: target model results

5. Conclusions

This article presents the uNET model for the semantic segmentation of structural elements of railway viaducts. The following semantic segmentation classes are distinguished: 1 – Non-bridge, 2 - Slab, 3 - Beam, 4 - Column, 5 - Nonstructural components (Poles, Cables, Fences), 6 - Rail, 7 - Sleeper, 8 - Other components. The trained network model contained over 7 million parameters. Very good training measures for the model were obtained: loss (categorical cross entropy): 0.0971, categorical accuracy: 0.9628. These good results allow us to state that the uNET network is a good solution for semantic segmentation not only in biomedical applications but also in engineering problems, in particular semantic segmentation of railway viaducts.

Acknowledgements

The authors are grateful for the financial support received from the project "*Development of regional network on autonomous systems for structural health monitoring*" financed by the Visegrad Fund, granted under the grant agreement 22110360.

References

[1] Spencer B.F. Jr., Hoskere V., Narazaki Y., *Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring*, Engineering 5 (2) (2019) 199-222,https://doi.org/10.1016/j.eng.2018.11.030.

[2] Ronneberger O., Fischer P., Brox T. (2015) *U-Net:*

Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[3] Mo Y., Wu Y., Yang X., Liu F., Liao Y., *Review the state-of-the-art technologies of semantic* segmentation *based on deep learning*, Neurocomputing, 2022, https://doi.org/10.1016/j.neucom.2022.01.005.

[4] Lingxin, Z., Junkai, S. & Baijie, Z. *A review of the research and application of deep learning-based computer vision in structural damage* detection. Earthq. Eng. Eng. Vib. 21, 1–21 (2022). https://doi.org/10.1007/s11803-022-2074-7

[5] Wu C., Zhang, F., Xia J., Xu Y., Li Guoqing., Xie J. Du Z. & Liu R. (2021). *Building Damage Detection Using U-Net with Attention Mechanism from Pre- and Post-Disaster Remote Sensing Datasets*. Remote Sensing. 13. 905. 10.3390/rs13050905.

[6] Hu X., Yao M., Zhang D., *School Road crack segmentation using an attention residual U-Net with generative adversarial learning*, Mathematical Biosciences and Engineering, 2021, Volume 18, Issue 6: 9669-9684. doi: 10.3934/mbe.2021473

About Authors

Piotr TAUZOWSKI Master's studies at the University of Technology at the Faculty of Civil Engineering. Post graduate studies in Big Data Analysis at the Faculty of Faculty of Electronics and Information Technology. PhD study in Institute of Fundamental Technological Research Polish Academy Of Sciences. Position: research specialist at Department of Information and Computational Science. Author or co-author of many papers as well as computer programs in the field of finite element analysis, reliability and optimization, incl. Numpress Explore system (http://numpress.ippt.pan.pl/) used for optimization and reliability analysis. Coordinator of Polish-Hungarian cooperation between the Polish and Hungarian Academy of Sciences. Research Interests: Robust, Design and Topology Optimization, Sensitivity Analysis, Software engineering, Programming, Big Data analysis, Machine Learning. Programming skill: C++, MATLAB, Python.

Mateusz ZARSKI is employed at IITiS PAN as an assistant in the Computer Vision Systems Group, where he conducts research on the development of machine learning methods in civil engineering. Studying at the Silesian University of Technology in 2015, he defended the title of engineer in the specialization of Building and Engineering Structures (bridges), and then in 2017 Master's Degree in Statistical Modeling of Building Life Cycle. Currently, he is a PhD student in the Simulations in Engineering project at the Silesian University of Technology. In the profession of a construction engineer, he worked in 2014-2015 as a construction engineer at Łęgprzem and in 2017-2018 as an assistant designer of bridge structures at InMost Projekt. He was also involved in inspections of bridges. Working scientifically, he dealt with the application of computer vision and deep machine learning methods in construction and combining machine learning with the Internet of Things (IoT) technology in the context of security protection.

Bartosz WOJCIK is employed at IITiS PAN as an assistant in the Computer Vision Systems Group, where he conducts research on the application of computer vision in construction. In 2014 and 2016, he received the titles of an engineer and a master's degree in Civil Engineering in the specialization of Structures and Engineering (bridges). Currently, he is a participant in interdisciplinary doctoral studies at the Silesian University of Technology. As part of his professional career, he had contact with infrastructure construction at every stage of the construction life, starting from design, through construction, and ending with maintenance. His scientific work combines modern technologies such as BIM (Building Information Modeling), 3D reconstruction and artificial intelligence in the context of inspection of bridges.

Mariusz OSTROWSKI is PhD Student at the Institute of Fundamental Technological Research of the Polish Academy (IPPT PAN) of Sciences in Warsaw (Poland) since 2017. He received the Master degree at Lublin University of Technology in 2016. He worked for space industry (UV2VIS project, which is related with Sentinel-5 mission) from 2017 to 2018. He also dealt with modelling of electromechanical systems and multiphysics problems for industry with aid of Matlab® software and other similar tools. His current research covers the fields of semi-active modal control of vibration and optimal sensors placement. His main interests are: modal and optimal control, mathematical modelling, model updating, structural health monitoring, optimisation and energy harvesting.

Bartlomiej BLACHOWSKI is a member of the scientific staff of the Institute of Fundamental Technological Research of the Polish Academy of Sciences (IPPT PAN). He received an MSc degree at the Silesian University of Technology in 2000, a PhD degree at IPPT PAN in 2005, and a habilitation degree at IPPT PAN in 2020. His research areas include wide-sense structural dynamics, in particular strategies of vibration control, algorithms for optimal structural design, and the use of sensor networks for structural health monitoring.

He is equally interested in modern computational methods of structural. analysis, experimental techniques such as operational modal analysis, and hardware solutions for wireless acceleration measurements of engineering structures.

Łukasz JANKOWSKI is an associate professor at the Institute of Fundamental Technological Research of the Polish Academy of Sciences (IPPT PAN). He received the MSc degrees in mathematics (2001) and in computer science (2000), the PhD degree in mathematical physics (2004), and the DSc in mechanics (IPPT PAN, 2014). The research interests involve the area of inverse problems in structural mechanics and include smart, adaptive structures, semi-active structural control and structural health monitoring**.**