http://doi.org/10.35784/iapgos.6968

received: 18.12.2024 | revised: 20.03.2025 | accepted: 10.04.2025 | available online: 27.06.2025

APPLICATION OF YOLO AND U-NET MODELS FOR BUILDING MATERIAL IDENTIFICATION ON SEGMENTED IMAGES

Ruslan Voronkov, Mykhailo Bezuglyi

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Department of Computer-Integrated Technologies of Device Production, Kyiv, Ukraine

Abstract. *This paper is devoted to the analysis of existing convolutional neural networks and experimental verification of the YOLO and U-Net architectures for the identification and classification of building materials based on images of destroyed structures. The aim of the study is to determine the effectiveness of these models in the tasks of recognising materials suitable for reuse and recycling. This will help reduce construction waste and introduce a more environmentally friendly approach to resource management. The study examined several modern deep learning models for image processing, including Faster R-CNN, Mask R-CNN, FCN (Fully Convolutional Networks), and SegNet. However, the choice was made on the YOLO and U-Net architectures. YOLO is used for fast object identification in images, which allows for quick detection and classification of building materials, and U-Net is used for detailed image segmentation, providing accurate determination of the structure and composition of building materials. Each of these models has been adapted to the specific requirements of building materials analysis in the context of collapsed structures. Experimental results have shown that the use of these models allows achieving high accuracy of segmentation of images of destroyed buildings, which makes them promising for use in automated resource control systems.*

Keywords: image segmentation, neural networks, classification of building materials, YOLOv8, U-Net, deep learning

ZASTOSOWANIE MODELI YOLO I U-NET DO IDENTYFIKACJI MATERIAŁÓW BUDOWLANYCH NA SEGMENTOWANYCH OBRAZACH

Streszczenie*. Niniejszy artykuł poświęcony jest analizie istniejących konwolucyjnych sieci neuronowych i eksperymentalnej weryfikacji architektur YOLO i U-Net do identyfikacji i klasyfikacji materiałów budowlanych na podstawie obrazów zniszczonych konstrukcji. Celem badania jest określenie skuteczności tych modeli w zadaniach rozpoznawania materiałów nadających się do ponownego wykorzystania i recyklingu. Pomoże to zmniejszyć ilość odpadów budowlanych i wprowadzić bardziej przyjazne dla środowiska podejście do zarządzania zasobami. W badaniu przeanalizowano kilka nowoczesnych modeli głębokiego uczenia do przetwarzania obrazu, w tym Faster R-CNN, Mask R-CNN, FCN (Fully Convolutional Networks) i SegNet, jednak wybór padł na architektury YOLO i U-Net. YOLO służy do szybkiej identyfikacji obiektów na obrazach, co pozwala na szybkie wykrywanie i klasyfikację materiałów budowlanych, a U-Net służy do szczegółowej segmentacji obrazu, zapewniając dokładne określenie struktury i składu materiałów budowlanych. Każdy z tych modeli został dostosowany do specyficznych wymagań analizy materiałów budowlanych w kontekście zawalonych konstrukcji. Wyniki eksperymentów wykazały, że zastosowanie tych modeli pozwala osiągnąć wysoką dokładność segmentacji obrazów zniszczonych budynków, co czyni je obiecującymi do wykorzystania w zautomatyzowanych systemach kontroli zasobów.*

Słowa kluczowe: segmentacja obrazów, sieci neuronowe, klasyfikacja materiałów budowlanych, YOLOv8, U-Net, głębokie uczenie

Introduction

The development of unmanned aerial vehicles (UAVs) [2], combined with the capabilities of deep learning technologies [14], opens up new opportunities for automating complex processes, such as agricultural land analysis and crop monitoring [4], forest monitoring and forest fire detection [10], infrastructure inspection and damage detection [23], search and rescue operations [22], building damage assessment after natural disasters [5], etc. One of these processes is potentially the identification and classification of building materials in building-type structures [12]. Paper [19] proposes a method for detecting destroyed buildings in aerial photographs using deep learning. The authors use convolutional neural networks to analyse images and automatically identify damaged buildings. However, it is noted that the accuracy depends on weather conditions and the complexity of the objects. Paper [7] explores the use of UAVs to assess the condition of bridges using remote sensing. The authors analyse various methods of collecting and processing data from UAVs, such as high-precision photography and laser scanning, to detect structural defects and damage. Paper [8] investigates methods for detecting damage to buildings after military conflicts using remote sensing techniques. The authors use satellite and aerial image analysis to identify damage, considering various data processing algorithms. They show that a combination of different types of images and deep learning can significantly increase the accuracy of damage detection, providing recommendations for improving algorithms to improve results. Continuing this topic, paper [1] presents an assessment of building damage using Sentinel-1 and Sentinel-2 satellite data. The use of radar and optical images allows for detailed damage assessment, regardless of weather conditions, and the combination of data from both satellites improves the accuracy of the results. This allows for the creation of highly accurate damage maps, which contributes to effective planning for the recovery of post-conflict areas. Further, paper [6] discusses the possibilities of using drones in construction research, including monitoring and assessing the condition of building materials. The authors emphasise the possibility of automating assessment processes using artificial intelligence algorithms, which significantly improves the accuracy and speed of defect detection. However, challenges related to regulatory issues and technical limitations need to be addressed for the full implementation of such technologies. Paper [13] presents a system that uses Lale VTOL UAVs to detect objects in real time. This approach ensures efficiency in monitoring and surveillance tasks by combining flight stability with high detection accuracy. The authors emphasise the importance of such systems in rescue and military operations, where speed and accuracy are key. Finally, paper [15] discusses a methodology for rapidly assessing building damage caused by an earthquake using UAVs and GIS. This approach allows for rapid data collection and damage classification, which makes it extremely effective for rapid assessment of the consequences of natural disasters.

In today's context, when it is important to quickly and accurately assess the condition of buildings and infrastructure, especially after natural disasters or emergencies, the issue of identifying building materials to analyse their condition and reuse potential is gaining new relevance. Traditional assessment methods, such as manual analysis, have significant drawbacks, including significant time and human resources that are not always available in emergency situations, as well as an increased risk for professionals who are forced to work in hazardous conditions [11]. Accordingly, there is a need to create automated systems for remote detection and analysis of the composition of building structures based on digital image processing.

Digital image processing methods have the potential to automate building condition assessment processes, reducing the workload of specialists and significantly speeding up the analysis process. However, there are certain difficulties associated with ensuring high accuracy of material recognition, especially when using UAVs, when the image quality may be insufficient due to noise or weather conditions [18]. In addition, colour image processing is complicated due to the nonlinear relationships between colour components, which requires the development of new algorithms to improve the accuracy of material identification on complex backgrounds [3].

In view of the above, the relevance of the study is to develop methods for the automatic identification of building materials using deep learning. The aim of the study is to determine the effectiveness of the proposed models in the tasks of recognising materials suitable for reuse and recycling. This will not only speed up the process of assessing the condition of buildings, but also increase its accuracy, especially when time and resources are limited. The results obtained can be used to create automated building analysis tools that will be widely used in various industries, such as construction, engineering, and humanitarian missions.

1. Methods and tools

Given the need to process large volumes of visual data acquired by UAVs, the task of developing efficient image segmentation algorithms arises. One of the key stages of this process is the selection of an appropriate neural network capable of accurately identifying various building materials in high-resolution images. Given the peculiarities of UAV data, namely the presence of noise, lighting changes, and a variety of material textures, the choice of neural network architecture is critical to achieving high segmentation accuracy.

By analysing modern architectures of convolutional neural networks, it is determined that ResNet (Residual Networks) is one of the most effective models for image classification and object detection. The ResNet architecture is based on the concept of residual blocks, which ensures efficient transfer of gradients through a large number of layers. In studies where ResNet was used to detect building materials, the model achieved an accuracy of about 89%, which indicates its high performance. However, ResNet is characterised by high computational requirements due to the large number of layers and parameters [16]. Another important architecture is VGG-16, which has a simple structure with 16 convolutional layers. This model is often used for classification and segmentation, but has significant limitations related to computational complexity. In recent studies analysing building materials, VGG-16 achieved an accuracy rate of 83%. However, its high memory consumption and computational resources may limit the use of this model in real-world applications [21]. When analysing approaches to real-time object detection, it should be noted that YOLO (You Only Look Once) is one of the most effective models that allows for the simultaneous segmentation of multiple objects. In a recent study on segmentation of building materials, YOLOv8 achieved a mAP of 96.7% at a high processing speed. However, one of the drawbacks of YOLO is the difficulty of detecting small objects, which can be critical in some scenarios [17]. U-Net is noted as a model specifically designed for semantic segmentation tasks. Its architecture has a symmetrical encoder-decoder structure, which allows for high pixel-level accuracy. In building materials segmentation, U-Net achieved an accuracy of up to 96% IoU, which is an impressive performance for complex scenes. Although this model is somewhat slower than YOLO due to its computational requirements, its effectiveness in segmentation tasks is undeniable [9]. Finally, Mask R-CNN is an improvement of Faster R- CNN that adds instance segmentation to the object detection task. This model achieves up to 95% mAP accuracy in complex scenes. The main disadvantage of Mask R-CNN is its high computational complexity, which makes it difficult to use for large amounts of data or in real time [20].

In the study, the choice of YOLOv8 and U-Net models is based on their unique advantages that meet the specifics of the task. YOLOv8, with an accuracy of up to 96.7% mAP and low computing requirements, provides high real-time processing speed, which is critical for analysing data obtained from unmanned aerial vehicles. This allows for the rapid identification of various building materials, which is important for effective planning of further actions in the context of infrastructure restoration.

On the other hand, U-Net, with its encoder-decoder architecture and 96% IoU accuracy, demonstrates high performance in semantic segmentation tasks, particularly at the pixel level. This allows for detailed analysis and classification of materials, which is important in the context of the accuracy of complex segmentation of images from destroyed buildings. Although U-Net may have some limitations in processing speed, its ability to accurately segment makes it an ideal choice for tasks requiring detailed analysis.

Table 1. Comparative table of CNN models

model	architecture	depth	accuracy	advantages	disadvantages
ResNet	Residual networks with feedback	Up to 152 layers	~89%	Deep architecture, high accuracy	High computational cost
VGG-16	13 convolutional layers and 3 bound	16 layers	~83%	Simple architectur, good feature extraction	High memory consumption
YOLOv8	Simultaneous grid segmentation	24 layers	~97%	Fast segmentation in real time	Lower accuracy on small objects
U-Net	Encoder-decoder	23 layers	~96%	High pixel- level segmentation accuracy	Slower processing time
R-CNN mask	Masks branch with regional offers	Depending on the backbone	~95%	High accuracy for multiple objects	High computation at cost

Table 1 presents the characteristics of convolutional neural networks such as ResNet, VGG-16, YOLOv8, U-Net, and Mask R- CNN, taking into account their architecture, depth, accuracy, and advantages and disadvantages. ResNet, with residual feedback networks and a depth of up to 152 layers, achieves an accuracy of approximately 89%, having high accuracy and a deep architecture, but is characterised by high computational costs. VGG-16, which consists of 16 convolutional layers, achieves an accuracy of about 83% and is characterised by a simple architecture and good feature extraction, but has a high memory consumption. YOLOv8, with 24 layers and simultaneous grid segmentation, demonstrates high real-time accuracy (~97%), but may be less accurate on small objects. U-Net, with an encoder-decoder architecture and 23 layers, achieves around 96% accuracy, offering high pixel-level segmentation, but is slower in processing. Finally, the Mask R-CNN, which uses a branch of masks with regional suggestions, achieves an accuracy of about 95% and is noted for its high accuracy for multiple objects, but its use is also associated with high computational costs.

2. Research results

For the study, a dataset was collected and processed (Fig. 1 and 2), consisting of 138 images of destroyed or partially destroyed buildings, of which 107 images were used for model training and 31 for validation. This material contains up-to-date information on the state of the infrastructure in the areas of greatest destruction and serves as a basis for assessing the possibilities for further analysis. The following segmentation classes were identified as part of the data annotation: brick walls, concrete, debris, and background. It is worth noting that the class "debris" is considered in a general way, as it can contain various materials, such as brick or concrete fragments. A detailed classification of these materials within the debris is difficult due to their mixing and damage during the destruction. Also, although there are fragments of glass, wood, etc. in the images, these materials are not included in the segmentation. Glass is difficult to recognise due to its transparency and reflective properties, and the assessment and counting of these types of raw materials for recycling or disposal is less feasible in the context of building materials. The obtained images demonstrate different types of damage, which allows us to evaluate the effectiveness of deep learning models for detecting and classifying building materials suitable for reuse or recycling. For annotation, we used the CVAT (Computer Vision Annotation Tool) tool, which supports high-precision markups even in difficult conditions, although it can require significant computing resources and time for manual markup of complex images.

Fig. 1. A fragment of the selected dataset

Fig. 2. A fragment of the selected dataset

The study aims to evaluate the effectiveness of using segmented images to train semantic segmentation models, in particular the YOLO and U-Net architectures. These models will be used to automate the process of identifying building materials, which can significantly accelerate the assessment of the amount of resources suitable for reuse or recycling. This, in turn, will help to optimise reconstruction plans and reduce the cost of transporting and processing construction residues. Automation of such processes is an important step in the development of integrated recovery management systems, which will allow for a rapid response to destruction and increase the efficiency of resource use. It is important to note that the results obtained are of both theoretical and practical importance, as they can be applied to improve reconstruction and recycling processes in real-world conditions.

The U-Net architecture was used in the study. This model is based on convolutional neural networks and is characterised by a symmetrical structure consisting of two main parts: an encoder and a decoder. The encoder, or contracting part, is responsible for reducing the spatial resolution of images by applying convolution and pooling operations, while preserving important contextual information. The decoder, or expansion part, gradually restores the spatial resolution using upsampling and inverse convolution operations, returning the data to the original image size and preserving spatial information about the location of objects. An important feature of U-Net is the use of "skip connections", which allow the transfer of detailed information from early encoder layers to the corresponding decoder layers, which significantly improves the quality of segmentation.

The model was trained using the PyTorch library, which provides flexibility in configuration and the ability to use advanced optimisation methods. The Adam optimiser was used for training with an initial learning rate of 0.001, which allows for quick finding of global minima of the loss function. Dice Loss was chosen as the loss function, which is particularly effective for segmentation tasks, as it takes into account both the correctness of pixel classification and their context in the image. The model was trained on 50 epochs, which allowed to achieve stable convergence and high accuracy of the results. The data was prepared based on pre-segmented images (Fig. 3 and 4) obtained using CVAT, which ensured high accuracy of building materials recognition.

Fig. 3. A fragment of the labelled dataset using CVAT

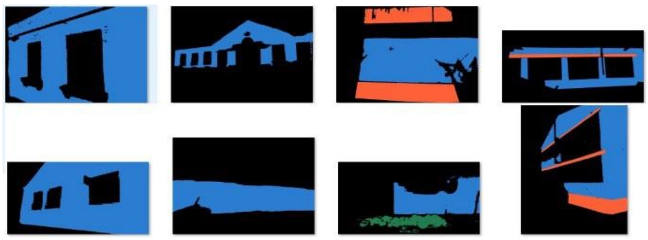


Fig. 4. A fragment of the labelled dataset using CVAT

To improve the efficiency of the segmentation process, we also investigated the effect of different input image sizes on the model's performance. The optimal size was chosen based on achieving a balance between processing speed and recognition accuracy. The images were scaled to reduce processing time without significantly losing segmentation quality. In addition, the impact of the model architecture on segmentation efficiency was investigated. Restricting the use of complex models in favour of optimised variants allowed us to achieve high performance. Experiments have shown that the combination of image size reduction and model architecture optimisation methods significantly improved segmentation efficiency without significant losses in the accuracy of building materials recognition.

Additionally, the impact of data augmentation was investigated, which allows to improve the quality of segmentation by increasing the diversity of the training data. The example shown in Figure 5 shows how the augmented image of a destroyed building (left) compares to the segmentation mask (right) generated by the model. The augmentation included techniques such as random scaling, rotation, and reflection, allowing the model to better adapt to different input variations. As a result, the augmentation reduced the likelihood of model overtraining and increased its ability to accurately segment the image, which is critical for correct building material recognition.

The segmentation mask clearly identifies different classes of objects: background areas (56.0%), brick walls (29.7%), concrete (1.2%) and debris (13.2%). This allows for a detailed analysis of the structure of the demolished building and a corresponding estimate of the amount of materials.

Evaluating the quality of semantic segmentation models is a critical step in the process of their development, as it allows to determine how accurately the model is able to recognise and classify different materials in images of destroyed buildings. In this study, the quality of the model was evaluated using such metrics as accuracy, recall, F1-score, and IoU (Intersection over Union) score.

The validation results showed that the model achieved Dice Loss of 0.23, F1-score of 0.77, and IoU score of 0.66. The F1-score of 0.77 indicates a good balance between the model's precision and recall, which is an important indicator of its ability to correctly segment different classes of objects. At the same time, the IoU score of 0.66 indicates certain limitations in segmentation accuracy, especially in the case of classifying specific materials such as concrete. The low accuracy of concrete recognition may be due to the difficulty of identifying it in images or to the insufficient amount of data of this material in the training set.

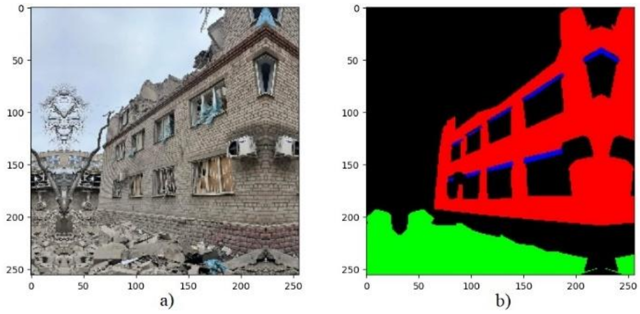


Fig. 5. a) image obtained by augmented image processing; b) image segmentation mask (background (black): 56.0%, brick (red): 29.7%, concrete (blue) 1.2%, trash (green): 13.2%)

Comparison of the segmentation masks (GT mask and PRED mask) also confirms the results of the quantitative assessment. Specifically, the GT mask contained the following areas: background areas 32.6%, brick walls 62.2%, concrete 5.2%, and debris 0%. While the PRED mask showed background areas of 37.2%, brick walls 62.8%, concrete 0%, and debris 0% (Fig. 6). These results indicate a good ability of the model to identify and segment brick walls, which is confirmed by the similar values in both masks. However, the presence of some discrepancies, such as the lack of correct segmentation of concrete, indicates possible areas for further improvement of the model, for example, by increasing the amount of training data, especially for materials that are more difficult to identify.

Evaluation of the performance of YOLO and U-Net models in the task of building materials identification and segmentation allows us to better understand their advantages and disadvantages, as well as the possibilities for their effective application in different contexts. The YOLO (You Only Look Once) methodology is known for its ability to perform fast image processing due to its end-to-end neural network architecture, which allows simultaneous prediction of bounding boxes and class probabilities for different objects in the image. This approach delivers high performance, making it particularly useful for real-time tasks such as video analysis.

The study of YOLO's performance on a dataset consisting of images of construction sites with brick walls showed good results. In particular, the model achieved 90% precision and 89% recall, which demonstrates the model's ability to both correctly classify objects and detect them in images. The accuracy at the threshold value of 0.5 (mAP50) was 94%, and in the range of threshold values from 0.5 to 0.95 (mAP50-95) – 81%. This demonstrates the high performance of YOLO in environments where objects such as brick walls need to be identified quickly and accurately, even in difficult conditions, as demonstrated in the video of the construction site.

The U-Net model, which specialises in segmentation, demonstrated a detailed approach to determining the boundaries of objects in an image. Thanks to its architecture consisting of a symmetric encoder and decoder, U-Net efficiently processes spatial information, allowing it to accurately detect the contours of building materials. In the study, U-Net achieved an F1-score of 0.77 and an IoU score of 0.66 on the validation set, which demonstrates the model's ability to accurately recognise and classify the selected classes of building materials. The validation set consisted of 31 images that were not used during training, which provides an objective assessment of the model's ability to adapt to new data. However, some difficulties were encountered with the identification of concrete, which can be explained by the difficulty of recognising this material or the insufficient amount of data in the training set.

In general, YOLO proved to be more effective for fast identification tasks, while U-Net performed better in precise contouring tasks. The choice between these models depends on the specifics of the task: if the priority is real-time identification speed, YOLO is the better choice; if high segmentation accuracy and detailed material recognition are required, U-Net provides a more efficient solution.

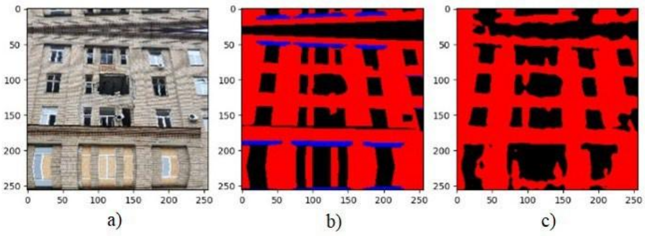


Fig. 6. a) image obtained by augmented image processing; b) image segmentation mask (background (black): 32.6%, brick (red): 62.2%, concrete (blue) 5.2%, trash (green): 0.0%); c) model prediction mask (background (black): 37.2%, brick (red): 62.8%, concrete (blue) 0.0%, trash (green): 0.0%)

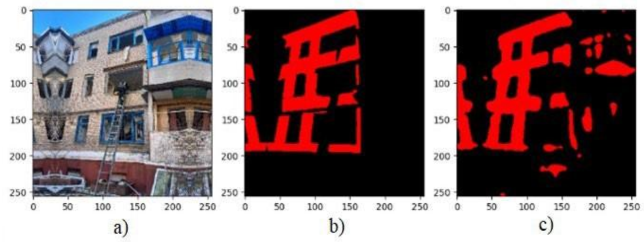


Fig. 7. a) image obtained by augmented image processing; b) image segmentation mask (background (black): 81.3%, brick (red): 18.7%, concrete (blue) 0.0%, trash (green): 0.0%); c) model prediction mask (background (black): 77.9%, brick (red): 22.1%, concrete (blue) 0.0%, trash (green): 0.0%)

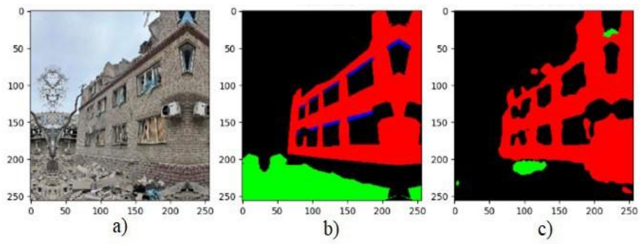


Fig. 8. a) image obtained by augmented image processing; b) image segmentation mask (background (black): 56.0%, brick (red): 29.7%, concrete (blue) 1.2%, trash (green): 13.2%); c) model prediction mask (background (black): 64.4%, brick (red): 34.2%, concrete (blue) 0.0%, trash (green): 1.3%

Fig. 9. image obtained by object segmentation using YOLO

3. Conclusions

The study conducted as part of this work allowed us to evaluate the effectiveness and features of the YOLO and U-Net models for the task of identifying and segmenting building materials based on images of destroyed buildings. The main goal was to evaluate the two models in terms of their accuracy, processing speed, and ability to adapt to different conditions, including specific material features and image structure complexity.

The main results of the study showed that the YOLO model demonstrates high speed and efficiency in identifying building materials, including brick walls, in images. Thanks to the use of an end- to-end neural network, YOLO was able to achieve an accuracy rate of 90% and a completeness rate of 89%, making it an effective tool for real-time recognition tasks. On the other hand, the U-Net model was more accurate in object segmentation, particularly in recognising brick walls and debris, but had some difficulty identifying concrete, which was reflected in its IoU score of 0.66. This indicates the need to improve methods for recognising complex materials such as concrete, which may require an increase in the size and variety of the training set.

The conclusions regarding the effectiveness of YOLO and U-Net for the task of building materials identification indicate that the choice of model should be based on the specifics of the task. YOLO is the most suitable for fast, real-time material identification, making it indispensable in applications where data processing speed is critical, such as analysing video streams from construction sites. U-Net, on the other hand, provides more detailed and accurate segmentation, which can be key in applications where high accuracy in identifying material contours and structures is required.

The prospects for further research and development of material detection systems based on UAVs include improving segmentation and identification algorithms to enhance their accuracy, reliability, and adaptability to diverse operating conditions. Special attention should be paid to the development of methods for training models on more complex, diverse, and representative datasets that encompass images of materials with various textures, shapes, and surface conditions, such as concrete, metal, or wood, to ensure high precision in classification and recognition. Algorithmic improvements should also account for variable lighting conditions, weather phenomena, and potential distortions in images obtained from UAVs, which will significantly increase the systems' versatility and applicability in real-world scenarios.

The integration of advanced models into UAV systems opens new horizons for automating processes of monitoring, analysis, and assessment of the condition of construction objects in real-time. This is particularly relevant for prompt responses to the consequences of destruction in disaster-stricken or conflict zones, where speed and accuracy in data collection are critically important. The use of YOLO and U-Net models in such systems can greatly enhance the efficiency of rescue operations, facilitate the reconstruction of damaged infrastructure, and promote the rational use of resources during recovery efforts. These solutions have the potential to become the foundation for the development of intelligent platforms that will contribute to the sustainable development of infrastructure in post-conflict regions and ensure long-term monitoring of its condition.

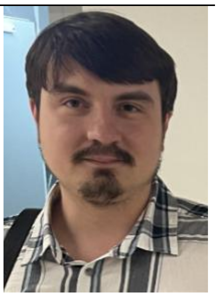
References

[1] Aimaiti Y. et al.: War Related Building Damage Assessment in Kyiv, Ukraine, Using Sentinel-1 Radar and Sentinel-2 Optical Images. Remote Sens. 14, 2022, 6239 [https://doi.org/10.3390/rs14246239].
[2] Ahmed F., et al.: Recent Advances in Unmanned Aerial Vehicles: A Review. Arabian Journal for Science and Engineering 47(7), 2022, 7963–7984.
[3] Ansari M. et al.: Significance of Color Spaces and Their Selection for Image Processing: A Survey. Recent Advances in Computer Science and Communications 15(7), 2022, 946-956.
[4] Bouguettaya A., et al.: Deep Learning Techniques to Classify Agricultural Crops through UAV Imagery: A Review. Neural Computing and Applications 34(12), 2022, 9511-9536.
[5] Calantropio A. et al.: Deep Learning for Automatic Building Damage Assessment: Application in Post-Disaster Scenarios Using UAV Data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 1, 2021, 113-120.
[6] Choi H. W. et al.: An Overview of Drone Applications in the Construction Industry. Drones 7(8), 2023, 515.
[7] Feroz S., Abu Dabous S.: UAV-Based Remote Sensing Applications for Bridge Condition Assessment. Remote Sensing 13(9), 2021, 1809.
[8] Ghandour Ali J., Jezzini A. A.: Post-War Building Damage Detection. Proceedings 2(7), 2018, 359 [https://doi.org/10.3390/ecrs-2-05172].
[9] He K. et al.: Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770-778.
[10] Jiao Z. et al.: A Deep Learning Based Forest Fire Detection Approach Using UAV and YOLOv3. 1st International Conference on Industrial Artificial Intelligence (IAI), IEEE, 2019, 1-5.
[11] Levchenko N. M., Beiner P. S., Beiner N. V.: Reconstruction of buildings using BIM technologies during city renewal in Ukraine. Physical Metallurgy and Heat Treatment of Metals 4(4), 2022, 64-70.
[12] Mahami H. et al.: Material Recognition for Automated Progress Monitoring Using Deep Learning Methods. preprint arXiv: 2006.16344, 2020.
[13] Mavroulis S. et al.: UAV and GIS Based Rapid Earthquake-Induced Building Damage Assessment and Methodology for EMS-98 Isoseismal Map Drawing: The June 12, 2017 Mw 6.3 Lesvos (Northeastern Aegean, Greece) Earthquake. International Journal of Disaster Risk Reduction 37, 2019, 101169.
[14] Myroniuk D. M., Blagitko B. Ya., Zayachuk I. M.: Computer Simulation of Deep Learning for Image Recognition. Computer Printing Technologies 42(2), 2019, 57-71.
[15] Paymode A. S., Malode V. B.: Transfer Learning for Multi-Crop Leaf Disease Image Classification Using Convolutional Neural Network VGG. Artificial Intelligence in Agriculture 6, 2022, 23-33.
[16] Ronneberger O., Fischer P., Brox T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. 18th International Conference Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany, 2015, Part III, Springer International Publishing, 2015, 234-241.
[17] Sony S. et al.: A Systematic Review of Convolutional Neural Network-Based Structural Condition Assessment Techniques. Engineering Structures 226, 2021, 111347.
[18] Sonkar S. et al.: Real-Time Object Detection and Recognition Using Fixed-Wing Lale VTOL UAV. IEEE Sensors Journal 22(21), 2022, 20738-20747.
[19] Su S., Nawata T.: Demolished Building Detection from Aerial Imagery Using Deep Learning. Proceedings of the ICA 2, 2019, 122.
[20] Wang H. et al.: YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8. IEEE Transactions on Instrumentation and Measurement, 2024.
[21] Wang, S. et al.: A Deep-Learning-Based Sea Search and Rescue Algorithm by UAV Remote Sensing. IEEE CSAA Guidance, Navigation and Control Conference (CGNCC), IEEE, 2018, 1-5.
[22] Wu W. et al.: Coupling Deep Learning and UAV for Infrastructure Condition Assessment Automation. IEEE International Smart Cities Conference (ISC2), IEEE, 2018, 1-7.
[23] Yin D. et al.: Mask R-CNN for Object Detection and Segmentation: A Comprehensive Review. Journal of Visual Communication and Image Representation 80, 2021, 103278.

M.Sc. Ruslan Voronkov
e-mail: venganza404@gmail.com

Ph.D. student Department of Computer-Integrated Technologies of Device Production, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".
Research interests: automation end robotics, control and instrumentation, automation, process automation.

https://orcid.org/0009-0000-4779-0132



Prof. Mykhailo Bezuglyi
e-mail: m.bezuglyi@kpi.ua

Doctor of Technical Sciences, professor Department of Computer-Integrated Technologies of Device Production, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic University".
Research interests: biophysics, optics, probability theory, statistics, research methodology.

https://orcid.org/0000-0003-0624-0585