UDC 004.8

NEURAL NETWORK ASSESSMENT OF BUILDINGS CONDITION BASED ON VISUAL DATA

Mazurets Oleksandr Viktorovych
Ph. D. in Engineering Science, Associate Professor
Klimenko Valeriia Ihorivna
Teacher
Shurypa Maksim Oleksandrovych
Master student
Computer Science Department, Khmelnytskyi National University,
Khmelnytskyi, Ukraine

Introductions. Post-war reconstruction processes and the progressive aging of housing stock create an acute demand for systematic and operational monitoring of the technical condition of buildings without relying on large-scale field inspections [1]. In many regions, especially those affected by military actions or long-term underinvestment in infrastructure, the number of objects requiring assessment significantly exceeds the available human and financial resources. Classical approaches to technical diagnostics are predominantly based on on-site expert inspections, visual surveys, and localized instrumental measurements, which are costly, time-consuming, and difficult to organize at the scale of entire districts or cities. As a result, such methods are poorly suited for rapid situation assessment, prioritization of restoration efforts, and continuous monitoring of large urban areas.

At the same time, vast volumes of visual data are already being accumulated from various sources, including high-resolution satellite imagery, aerial photographs acquired by unmanned aerial vehicles, and ground-level images collected through routine inspections or public reporting channels [2]. Despite their availability, these heterogeneous data streams are rarely transformed into structured indicators that can directly support engineering decisions, urban planning, or risk assessment. Without automated analysis tools, the interpretation of such imagery remains fragmented and largely dependent on manual processing, which limits its practical value in large-scale monitoring scenarios. This gap highlights the urgency of developing neural network–based methods capable of extracting building outlines, identifying damage patterns, and assessing the overall condition of structures using visual data as the primary source of information.

Neural network approaches for visual data analysis [3] have gained particular relevance in the domain of building condition assessment due to their potential to improve monitoring efficiency, enable early detection of structural defects, and reduce the likelihood of emergency situations. In contrast to traditional techniques based on manual inspection or classical image processing algorithms, which often rely on handcrafted features and heuristic thresholds, deep learning–based solutions offer a higher degree of automation and adaptability. Manual and semi-automated methods are not only labor-intensive and subjective, but also exhibit limited robustness when applied to large and diverse datasets [4], especially in the presence of complex visual conditions typical for urban environments.

The rapid development of deep learning, and convolutional neural network (CNN) architectures in particular [5], has opened new opportunities for automated detection of cracks, corrosion traces, surface deformations, partial collapses, and other indicators of structural degradation. Modern CNN-based models demonstrate a strong ability to extract spatial, textural, and geometric features from a wide range of visual sources, including photographic images, video sequences, UAV data, and laser scanning outputs [6]. Advances in computer vision tasks such as semantic and instance segmentation [7], image classification [8], and object detection [9] make it possible to construct multi-level damage maps, estimate degrees of material deterioration, and identify zones with elevated risk of failure. These capabilities are especially important for prioritizing detailed engineering inspections and optimizing the allocation of restoration resources.

Particularly promising are integrated approaches that combine CNNs with transformer-based architectures and multimodal analysis techniques, enabling the joint processing of visual data with auxiliary information such as textual descriptions of defects, sensor measurements, or geometric building models [10]. Such hybrid frameworks enhance contextual understanding and improve the robustness of condition assessment by incorporating complementary sources of evidence. In the context of real-world applications, an important research direction is the development of algorithms capable of operating reliably under noise, varying illumination, diverse viewing angles, occlusions, and complex backgrounds, which are characteristic of real engineering structures and post-war urban environments [11]. Addressing these challenges is essential for ensuring the practical applicability and scalability of automated building condition assessment systems.

The purpose of the work is to form and experimentally substantiate a neural network approach to assessing the condition of buildings using visual data, which combines automatic segmentation of building objects in images and subsequent interpretation of their visual features to determine the degree of damage.

Methodology. The proposed methodology is aimed at the automated identification of buildings and the subsequent assessment of their technical condition based on high-resolution orthophotos and aerial imagery acquired from unmanned aerial vehicles. To address this task, single-stage convolutional neural network–based segmentation detectors from the YOLOv8, YOLOv11, and YOLOv12 families were selected as the core modeling framework. The choice of these architectures was motivated by their favorable balance between detection accuracy, inference speed, and suitability for large-scale processing scenarios. Particular emphasis was placed on lightweight model variants, namely 8s, 11n, 11s, 11m, and 12n, which are better aligned with the constraints of real-world deployment, including limited computational resources and the need for near-real-time inference.

The initial dataset consisted of large-format orthophotos and UAV-acquired aerial images characterized by heterogeneous spatial resolution, varying illumination conditions, occlusions, and complex urban backgrounds. To ensure consistency and facilitate batch processing during training, all source images were subdivided into fixed-size tiles with normalized spatial resolution. This tiling strategy not only reduced memory requirements but also improved the detectability of small and partially occluded buildings, which are common in dense urban environments and post-conflict areas.

A significant challenge in building condition assessment is the inherent class imbalance between intact structures and severely damaged or destroyed ones. To mitigate this issue, class balancing techniques were applied in combination with systematic dataset augmentation. The training set was expanded using a diverse set of geometric transformations, including rotations, flips, scaling, and random cropping, as well as photometric augmentations such as brightness and contrast variation, color jittering, and noise injection [12]. These transformations were designed to improve the robustness of models to variations in viewing angle, lighting, and image quality, which are typical for aerial imagery collected under non-controlled conditions.

The annotation scheme was formulated to support joint learning of spatial localization and semantic interpretation. Each image tile was labeled with binary pixel-wise masks distinguishing "building" from "background," along with categorical labels representing the technical condition of each detected structure. Three condition classes were defined: "satisfactory," corresponding to visually intact buildings; "damaged," indicating moderate structural degradation; and "significantly damaged/destroyed," encompassing severe damage or near-complete destruction. This labeling strategy enabled formulation of task as a multi-task learning problem.

Model training was conducted using a multi-head architecture. The first head was responsible for object detection and instance segmentation, producing bounding boxes and pixel-level masks for building objects. The second head performed condition classification by aggregating deep feature representations extracted within the predicted object masks [13]. This design allowed the model to learn a shared visual representation while optimizing separate loss functions for spatial localization and condition estimation. Such joint optimization facilitates mutual reinforcement between tasks, as accurate segmentation improves condition classification.

Model performance was evaluated using a set of complementary quantitative metrics. Segmentation quality was assessed using the mean Average Precision at an intersection-over-union threshold of 0.5 for bounding boxes (mAP50(B)) and instance masks (mAP50(M)). The accuracy of building condition classification was measured using macro-averaged Precision, Recall, and F_1-score, ensuring equal importance of all condition classes regardless of their frequency in the dataset [14]. This evaluation protocol was selected to provide a balanced and comprehensive assessment of both localization and semantic classification capabilities.

Results. The experimental evaluation revealed clear performance differences among the tested model configurations, highlighting the impact of architectural complexity, data augmentation strategies, and training stability on the final results. Overall, compact and carefully tuned models demonstrated a more favorable trade-off between segmentation accuracy and computational efficiency compared to deeper and more parameter-heavy variants (Figure 1).

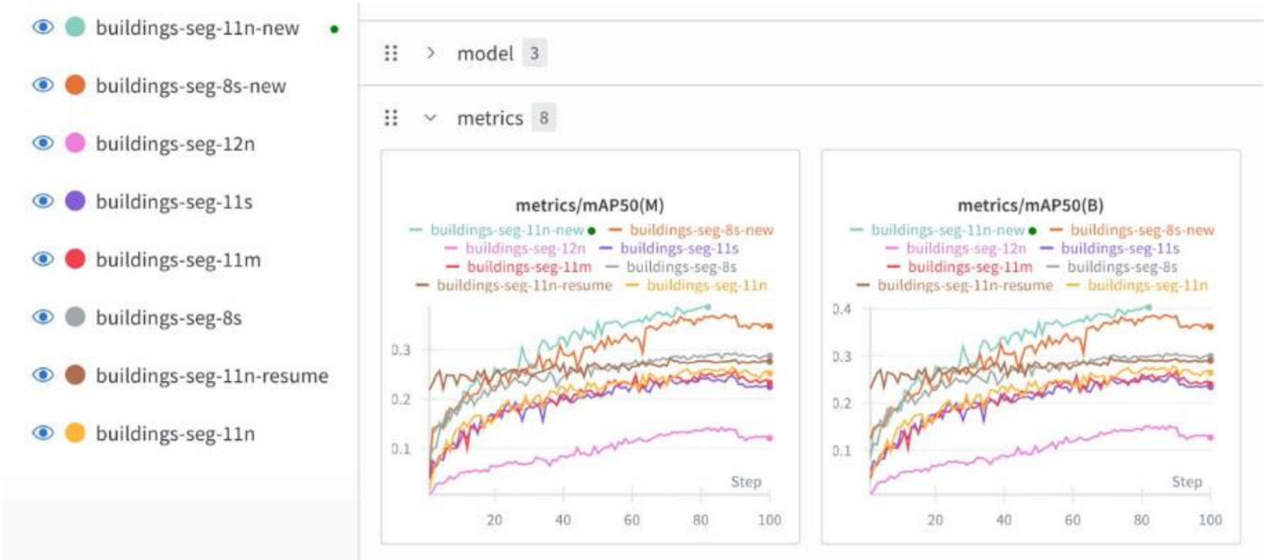

Fig. 1. YOLO training graphs of different versions

Among all tested configurations, the buildings-seg-8s-new model achieved the highest segmentation performance on the validation dataset. This configuration incorporated an extended augmentation pipeline and refined hyperparameter settings, which significantly improved its generalization capability. The model reached an mAP50(M) of approximately 0.35 for instance masks and an mAP50(B) of around 0.39 for bounding boxes. These results indicate a reliable ability to delineate building contours and localize structures even in visually complex scenes with partial occlusions and background clutter.

A comparable level of performance was observed for the buildings-seg-11n-new configuration, which achieved mAP50(M) and mAP50(B) values of approximately 0.34 and 0.37, respectively. Although slightly inferior in terms of raw accuracy, this model exhibited improved inference speed and reduced memory consumption. As a result, it represents a practical alternative for deployment on resource-constrained platforms, such as edge devices or mobile GPU nodes, where computational efficiency is a critical factor.

In contrast, baseline versions of the models that were trained without advanced augmentation strategies – namely buildings-seg-8s, buildings-seg-11n, and buildings-seg-11s – demonstrated noticeably lower performance. Their mAP50 values ranged between 0.24 and 0.30, indicating a higher rate of missed detections and imprecise segmentation, particularly for small buildings or structures partially covered by vegetation, debris, or shadows. This performance gap underscores the importance of data augmentation and sample diversity in improving model robustness for real-world aerial imagery.

The weakest results were obtained for the larger and more complex architectures, specifically buildings-seg-11m and buildings-seg-12n. Despite their increased representational capacity, these models failed to achieve satisfactory performance, with mAP50(M) values remaining below 0.15. Analysis of the training dynamics revealed slow convergence and early saturation of the learning curves, suggesting that the available training dataset was insufficient to fully exploit the capacity of these deeper networks. This observation highlights the risk of overparameterization in scenarios where labeled data are limited, and supports the preference for compact architectures in practical applications.

Overall, the experimental results demonstrate that lightweight single-stage segmentation models, when combined with appropriate data preprocessing and augmentation strategies, are capable of providing reliable building detection and condition assessment from aerial imagery. The obtained performance levels are sufficient to support downstream tasks such as automated damage mapping, prioritization of on-site engineering inspections, and large-scale monitoring of urban infrastructure in post-war reconstruction and long-term asset management scenarios.

In the second stage of the proposed pipeline, after selecting and fixing the optimal segmentation architecture, the focus was shifted to training the classification head responsible for assessing the technical condition of detected buildings. At this stage, the segmentation weights were retained, and the model was fine-tuned to discriminate between different levels of structural damage based on features aggregated within the predicted building masks. For the configuration based on the buildings-seg-8s-new model, the macro-averaged F_1 score across the three defined condition classes reached approximately 0.78, indicating a balanced performance in terms of precision and recall for multi-class condition assessment. The highest classification accuracy was achieved for the “destroyed” class, which can be explained by the presence of pronounced visual cues such as collapsed roofs, missing walls, and irregular building outlines. In contrast, the majority of misclassifications occurred at the boundary between the “damaged” and “significantly damaged” categories, reflecting the gradual and visually ambiguous nature of structural degradation processes [15].

A detailed error analysis further revealed that the classification performance is sensitive to several external factors inherent to aerial imagery. In particular, strong cast shadows, seasonal variations in illumination and surface appearance, and partial occlusion of facades by vegetation were identified as major sources of uncertainty. These findings suggest that incorporating additional sources of information, such as multispectral or infrared channels, as well as explicit textural and structural descriptors, could enhance the robustness of condition assessment, especially in challenging observation conditions.

Practical significance. From an applied perspective, the proposed approach enables the automated generation of damage maps from series of aerial or UAV-acquired images, in which both the spatial contours of buildings and their approximate technical condition are explicitly represented. Such maps support the efficient identification of priority objects for on-site engineering inspections, facilitate evidence-based decision-making in the planning and scheduling of restoration and reconstruction works, and can be seamlessly integrated into existing geographic information systems used by municipal authorities, emergency services, and urban planners.

Conclusions and prospects. The conducted study demonstrates that lightweight YOLO-seg–based models, provided that data preparation, annotation, and augmentation are carefully organized, are capable of delivering an acceptable level of accuracy for automated building extraction and preliminary condition assessment using visual data. Among the evaluated configurations, buildings-seg-8s-new and buildings-seg-11n-new exhibited the most favorable trade-off between accuracy, inference speed, and computational efficiency. Future work will focus on expanding and diversifying the training dataset by incorporating imagery from different sensors, acquisition seasons, and geographic regions. Additional research directions include the integration of multispectral data, refinement of the condition classification scale in accordance with building codes and engineering standards, and the development of explainability modules that will allow domain experts to interpret and validate the decisions produced by the neural network system in the context of post-war reconstruction planning.

REFERENCES

1. Demian, P., Hassan, T. M., Kalmykov, O., Demianenko, I., & Makarov, R. (2024). BIM implementation in post-war reconstruction of Ukraine. Buildings, 14(11), 3495.

2. Chashyn, D., Khurudzhi, Y., & Daukšys, M. (2024). Directions for the formation of city intelligent models using artificial intelligence for the post-war reconstruction of historical buildings. Budownictwo i Architektura, 23(1), 73–86.

3. Zalizko, V. D., Cherniak, A. M., Nowak, D., & Artemov, V. Y. (2024). Assessing Ukrainian education security in the context of artificial intelligence integration for accelerated post-war recovery. Scientific Bulletin of National Mining University, (6).

4. Alshafei, I. A., & Almashhour, R. T. N. (2025). Integrating artificial intelligence in urban re-design: A collaborative approach to post-war design phases. International Journal of Human Capital in Urban Management, 10(2).

5. Mazurets, O., Uspenska, K., Vit, R., & Tyschenko, O. (2024). Intelligent

system for determining the object attributes values by neural networks means by graphic images in databases. In: Current trends in the development of scientific research in today's conditions. Proceedings of the XXV International Scientific and Practical Conference, pp. 86–91.

6. Mazurets, O. V., Klimenko, V. I., Molchanova, M. O., & Sultanov, A. V. (2024). Object-oriented intelligent system for neural network detection of sugar crystallization zones. In: Global science: Prospects and innovations. Proceedings of the 10th International Scientific and Practical Conference, pp. 198–207.

7. Zharnovskyi, O., Mazurets, O., & Sobko, O. (2024). Approach to identification of artificial intelligence-generated people images by means of machine learning. In: Key aspects of the development of scientific research in modern conditions. Proceedings of the XLV International Scientific and Practical Conference, pp. 69–73.

8. Pokhytun, A., Mazurets, O., Molchanova, M., & Tyschenko, O. (2024). Method for neural network detecting changed images of people's faces using CNN. In: New horizons in scientific research: Challenges and solutions. Proceedings of the 1st International Scientific and Practical Conference, pp. 35–40.

9. Kharysh, I., Sobko, O., & Mazurets, O. (2024). Designing CNN neural network model for detecting fractures of lower extremities by X-ray images. In: The impact of scientific research on the development of the modern world. Proceedings of the XLIV International Scientific and Practical Conference, pp. 91–96.

10. Bas, I. S., Kadynska, V. D., Klimenko, V. I., & Mazurets, O. V. (2025). Convolutional neural network transfer learning method for aircraft image classification. In: Scientific method: Reality and future trends of researching. Proceedings of the VI International Scientific and Theoretical Conference, pp. 147-155.

11. Ostapchenko, N., Tyschenko, O., Denysenko, B., & Mazurets, O. (2025). Semantic search of relevant images using vector databases. In: Modern scientific research: Theoretical and practical aspects. Proceedings of the II International Scientific and Practical Conference, pp. 161–165.

12. Mushtyn, O., Sobko, O., Molchanova, M., & Mazurets, O. (2025). Convolutional neural network architecture for image-based architectural style recognition. In: Evolving science: Theories, discoveries and practical outcomes. Proceedings of the 4th International Scientific and Practical Conference, pp. 130-143.

13. Malaydakh, V., Molchanova, M., Shevchuk, P., & Mazurets, O. (2025). Deep learning neural network architecture for determining sunflower growth stage from visual data. In: Modern scientific research: Theoretical and practical aspects. Proceedings of the II International Scientific and Practical Conference, pp. 143–148.

14. Kok, I. A., Kadynska, V. D., Zalutska, O. O., & Mazurets, O. V. (2025). Object-oriented intelligent system for automated control of smoking by video data. In: Current scientific goals, approaches and challenges. Proceedings of the IV International Scientific and Theoretical Conference, pp. 156–164.

15. Molchanova, M., Didur, V., Mazurets, O., Sobko, O., & Zakharkevich, O. (2025). Method for construction and demolition waste classification using two-factor neural network image analysis. CEUR Workshop Proceedings, 3970, 168–182.