UDC 004.8
DOI: 10.15587/1729-4061.2026.351605

DEVISING A NEURAL-NETWORK METHOD FOR ASSESSING THE CONDITION OF DESTROYED BUILDINGS USING IMAGES FROM UNMANNED AERIAL VEHICLES

Oleksandr Mazurets
PhD, Associate Professor*
ORCID: https://orcid.org/0000-0002-8900-0650

Maryna Molchanova
Corresponding author
Doctor of Philosophy (PhD), Senior Lecturer*
E-mail: m.o.molchanova@gmail.com
ORCID: https://orcid.org/0000-0001-9810-936X

Maksim Shurypa*
ORCID: https://orcid.org/0009-0003-7025-4647

Olena Sobko
Doctor of Philosophy (PhD), Senior Lecturer*
ORCID: https://orcid.org/0000-0001-5371-5788

*Department of Computer Science
Khmelnytskyi National University
Instytutska str., 11, Khmelnytskyi, Ukraine, 29016

The process of assessing the condition of destroyed buildings using aero photographs acquired from unmanned aerial vehicles (UAVs) has been investigated in this study. The task addressed relates to the fact that post-war monitoring of territories is complicated by the volume of satellite and UAV images, which exceeds the capabilities of expert inspection while the lack of uniform interpretation tools and domain shift reduce the reproducibility of assessment.

The proposed two-stage neural-network method combines segmentation of buildings and determining the level of building destruction on a four-level damage scale: "absent", "minor", "significant", "destroyed". The "xView2" corpus was used as the source material, supplemented with authentic marked UAV images.

The "You Only Look Once" (YOLO) segmentation models, versions v8s and v11n, were used to extract building contours, and the Vision Transformer (ViT) was used to categorize damage. Experiments were performed in Google Colab (USA) applying PyTorch (USA) and Ultralytics (UK). The Mean Average Precision (mAP) was calculated for segmentation models. The mAP indicators remain acceptable even in complex urban settings.

For the ViT classification model, the Precision, Recall, and F1 values above 0.9 were obtained. The values achieved are attributed to the combination of a two-stage architecture and sample balancing.

The devised method is applicable to satellite and UAV images; unlike existing solutions, it retains stability under domain shift. The resulting models could be implemented as a basic module in geographic information systems and decision support systems, enabling practical use of this study's results. For correct operation, sufficient resolution and representativeness of the training sample are required

Keywords: UAVs, aerial photography, YOLO, ViT, neural-network classification of damage, assessment of damage to buildings

Received 04.11.2025
Received in revised form 16.01.2026
Accepted date 26.01.2026
Published date 27.02.2026

How to Cite: *Mazurets, O., Molchanova, M., Shurypa, M., Sobko, O. (2026). Devising a neural-network method for assessing the condition of destroyed buildings using images from unmanned aerial vehicles. Eastern-European Journal of Enterprise Technologies, 1 (2 (139)), 6–16. https://doi.org/10.15587/1729-4061.2026.351605*

1. Introduction

The destruction of civil and industrial infrastructure poses problems for urban planning and technical supervision systems that cannot be solved by conventional means of visual inspection. At the stage of post-war reconstruction, operational, spatially detailed, and reproducible information on the extent of damage to buildings is necessary. Such information makes it possible to justify the priorities of repair work, plan the logistics of building materials, and assess the risks of further operation of facilities.

Manned inspections under such conditions are often dangerous or economically impractical, therefore remote observation methods play a key role, primarily aerial photography using UAVs, which provide high spatial resolution and flexibility of survey routes.

Accumulation of image arrays from unmanned platforms does not guarantee the possibility of timely assessment of the state of infrastructure since manual photo interpretation of such data requires significant human resources and is characterized by subjectivity. Classical computer vision methods, focused on the analysis of textures, contours or other parameters, are characterized by instability with respect to the variety of building types, lighting changes, partial overlapping of objects, smoke and dust, typical of destruction zones. This necessitates the need for specialized algorithms capable of automatically identifying buildings in aerial images, assessing the extent of their damage, and operating under the scaled processing mode of large areas.

The evolution of deep neural-network models opens up opportunities for building interference-resistant methods for automated assessment of building damage from aerial photography data. Convolutional [1] and transformative [2] architectures are able to generalize complex spatial features, combine information from sources of different scales, take into account the context of the scene and local structural deformations. In addition, existing approaches are often limited to a binary scheme of division into damaged and undamaged objects. This underlines the relevance of devising a method for neural-network assessment of the state of destroyed buildings from UAV images.

2. Literature review and problem statement

Methods for automated assessment of building damage using remote sensing data are actively developing due to the need for operational analysis of the consequences of extreme events and military conflicts. Thus, in paper [3], approaches to identifying damaged buildings using satellite and unmanned aerial vehicles (UAVs) images are systematized and the use of optical and radar sensors of different spatial resolutions is analyzed. It is shown that approaches based on deep neural networks dominate for segmentation and classification tasks but face limitations in resolution, coverage, and generalizability of models.

In paper [4], the task of assessing building damage after the 2023 earthquake in Turkey using very high spatial resolution satellite data is considered. The authors proposed a two-stage BDANet network that performs building segmentation using "before the event" images and classification of four levels of damage using paired "before/after" images. Experimental data obtained on the xBD set confirm the high level of detail of the assessment for individual buildings, but the method fundamentally depends on the availability of paired scenes of "before/after" images.

A separate group is made up of approaches that combine highly detailed aerial photographs with satellite images and use semi-supervised learning to use large amounts of unlabeled data. In paper [5], a five-level damage scale was formed and a corpus of thousands of labeled and tens of thousands of unlabeled images was built. Despite the involvement of unlabeled data, the method remains dependent on the prior formation of building polygons and is focused mainly on scenes without massive collapses as a result of destruction.

Much of the current research is moving from satellite data to data from unmanned platforms, which provide higher detail, flexible route planning, and operational updates of information. For example, in study [6], an automated pipeline for assessing building damage from the eruption of the La Soufrière volcano was built, based on approximately 50 thousand fragments extracted from more than 2800 UAV images. The applied convolutional models perform building localization and classification by damage level, achieving F_1-indicators of about 0.81 for the localization task and 0.84 for the damage level determination task. At the same time, the model is tuned to specific volcanic scenarios, and the authors note the need for additional data and testing in other regions.

Another direction is related to the use of lightweight neural-network architectures suitable for deployment directly on board UAVs. Thus, in paper [7], the use of CNN MobileNetV3-Small for classifying the level of post-earthquake damage from images obtained from UAVs under a mode close to real time was investigated. It was found that the combination of UAVs with such Convolutional Neural Network (CNN) provides accuracy sufficient for operational support of search and rescue teams under conditions of limited computing resources.

In paper [8], a modification of the YOLO family of neural networks was used to combine high accuracy and speed when detecting damaged fragments of buildings in UAV images. The described DB-YOLOv5 method supplements the basic architecture with extended convolutions and attention modules and is oriented towards real-time operation in embedded systems. The authors note that the insufficient operability of previous solutions is a critical limitation for the use of such detectors in emergencies. Further modifications of the neural network are described, in particular YOLOv6s-GRE and its quantized variants, which provide increased values of the mAP metric without increasing the model size, which is important for onboard applications.

A number of publications consider the features of infrastructure inspection using deep learning tools and UAVs and generalize the application of such approaches to detect defects in various engineering structures. Thus, in [9], typical tasks, sensor configurations, dataset structure, and key challenges related to the lack of reference sets and the need for real-time models are analyzed. Studies [10] on cracks and defects in facades of historical buildings confirm the ability of CNN and transformer architectures to detect small damages in UAV images. At the same time, these works mainly consider relatively intact facades and almost do not cover scenarios of complete or close to complete destruction of buildings.

Summarizing the above results, several principal trends can be identified in modern approaches to assessing building damage using remote sensing data [3–6, 9]. In particular, for global mapping of natural disaster consequences, damage assessment methods based on satellite images have been devised, using paired "before/after" scenes and multi-level damage scales [3–5]. At the same time, for local disaster scenarios, approaches using UAVs are actively developing, which combine building detection and damage level classification, including under a mode close to real time [6–8]. Despite significant progress, issues related to the lack of available paired "before/after" images remain unresolved in known approaches, in particular in zones of prolonged military operations and rapidly changing urbanized areas [3, 6]. In some cases, this is due to the objective difficulties of long-term monitoring, limitations of photo archives, and the economic inexpediency of systematic observations before the start of the active phase of destruction.

Most approaches based on UAV data consider a limited number of damage classes and focus on the condition of the roof or upper floors, avoiding completely collapsed frames [6, 7, 10]. The reason for this is the difficulty of large-scale data mapping for heavily damaged areas due to the heterogeneity of the rubble and the increased flight safety requirements in such areas [6, 9]. An additional challenge is the domain shift between satellite and UAV images, which manifests itself in changes in flight altitude, viewing angles, background structure, and fine-scale details of the rubble [3, 5, 9].

Most described solutions are either tied to satellite data or work with UAVs under narrowly defined shooting conditions, without systematic adaptation to different flight configurations [5–7].

An option to overcome the outlined difficulties is a two-stage approach, in which the segmentation of the building and the classification of the degree of damage are performed by different neural network models with the possibility of separate retraining. In paper [11], a general concept of such an approach was proposed, in which the segmentation module extracts knowledge in the form of regions of interest from a large corpus of photo data, and the classification module analyzes individual fragments of buildings in these regions.

3. The aim and objectives of the study

The aim of the study is to devise a two-stage neural-network method for assessing the condition of destroyed buildings using UAV images, taking into account the domain shift between satellite and UAV data. The method generates uniform quantitative estimates of the level of destruction of buildings for further use in monitoring systems and planning for the restoration of urban infrastructure.

To achieve this goal, the following tasks were set:

– to define steps of the neural-network method for asa sessing the condition of buildings using satellite and UAV images, taking into account the domain shift;

– to develop an experimental software tool for data prepap ration, training, and application of neural-network models for automated processing of satellite and UAV images;

– to conduct an experimental verification of the effece tiveness of the proposed method on test data subsets with a quantitative assessment of the accuracy of segmentation and classification and analysis of the influence of shooting conditions.

4. Materials and methods

The object of the study is the process of assessing the condition of destroyed buildings using aerial photographs from UAVs. The principal hypothesis assumes that a two-stage neural-network approach with separation of the stages of building segmentation and classification of the degree of damage is suitable for a stable assessment of the condition of buildings.

The study assumes that the source images have sufficient spatial resolution, correct georeferencing, and show buildings with recognizable contours. It is also assumed that the available satellite bodies and samples of UAV images provide representative coverage of typical destruction scenarios in urban buildings.

Basic simplifications include limiting the analysis to four generalized damage levels, which describe mainly the condition of the external structures of buildings.

Internal damage, utility networks, and small local defects of facades are not considered due to going beyond the scope of the task and the difficulty of reliable interpretation from aerial photographs. The consideration is limited to optical RGB images without the use of radar or multispectral channels, which simplifies the requirements for the UAV sensor equipment.

The approach to neural-network assessment of the state of destroyed buildings from UAV images (Fig. 1) is designed to construct trained neural-network models capable of performing image segmentation and classification of building states by the corresponding neural networks (Fig. 2). The approach is based on combining the large-scale satellite image corpus “xView2” with a subsample of labeled UAV images that we added.

Based on this dataset, two neural-network models are sequentially trained: a segmentation detector and a classification model.

Based on a comparative analysis of modern neural network architectures, the YOLOv8s and YOLOv11n models were selected for building segmentation tasks. They provide stable selection of building contours even under variable lighting conditions, scale, and UAV shooting angle.

The ViT neural-network model was selected as a classification module, which analyzes images as a set of patches and detects local signs of damage. The use of ViT allowed for pre-training on large image corpuses and reduced the impact of the limited amount of labeled pre-training data for the destruction scale.

The YOLO-type segmentation detector forms building contour masks taking into account the domain shift between satellite and UAV images. The choice is based on the ability of these architectures to perform multi-scale object detection and segmentation in real time while remaining robust to changes in scale, perspective, and shooting conditions, which is critical for multi-scale satellite and UAV imagery.

A classification model based on a ViT vision transformer determines the belonging of objects to one of four levels of destruction based on the contours of buildings. The transformer architecture makes it possible to use pre-training on large image corpuses and adapting the model to the specifics of the target dataset. This mitigates the impact of the limited amount of labeled data for a four-level destruction scale under the conditions of domain shift between satellite and aerial data.

The input data of the approach is a dataset for training neural networks. The basis for building and verifying the proposed method is the open corpus “xView2” of the “XBD Dataset Organized” competition [12], which is implemented on the basis of the “xBD2” dataset for assessing building damage from satellite images before and after disasters. The dataset contains over 22,000 1024×1024 pixel images acquired before and after 19 events, covering an area of about 45,000 km^2 in different regions of the world. For these images, over 850,000 building polygons with damage degree annotations were prepared, covering different building densities, construction types, and variable photofixation angles [13].

In the version of the dataset “xView2”, distributed via the Kaggle platform, the images are organized into “before” and “after” pairs. The accompanying markup files contain building polygons tied to both pixel coordinates and a geographic coordinate system. Damage assessment in “xBD” is based on the ordinal four-level scale “Joint Damage Scale”, which is used for disaster impact analysis tasks. Buildings are assigned four main levels: “no-damage”, “minor-damage”, “major-damage” and “destroyed”.

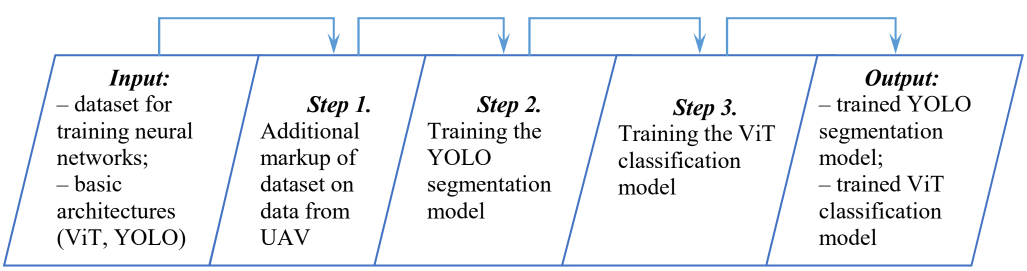

Fig. 1. Stages in the approach to neural-network assessment of the condition of destroyed buildings from images

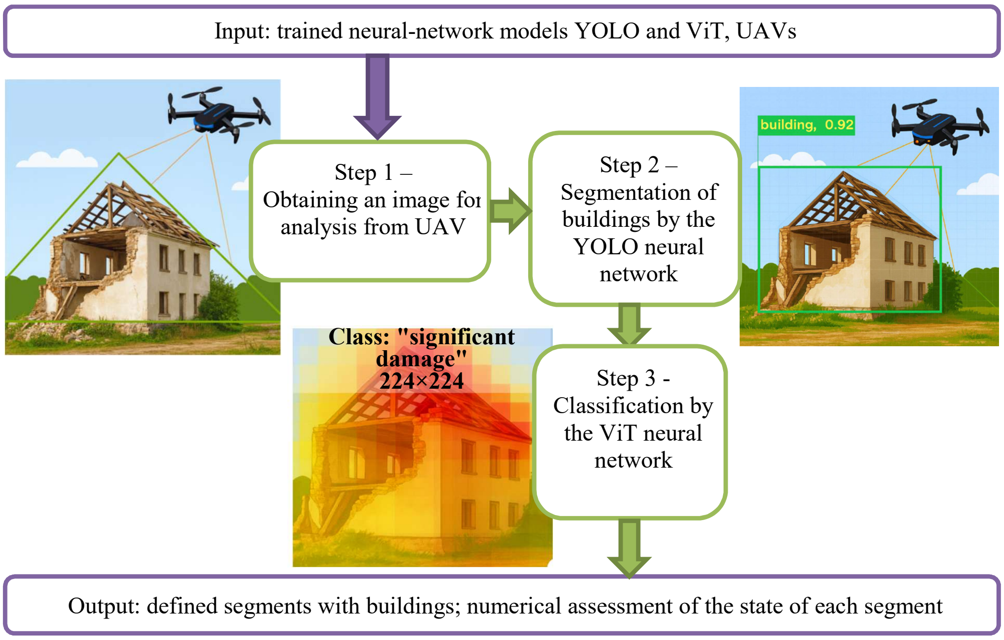

Fig. 2. Illustration and steps of the neural-network method for assessing the condition of destroyed buildings from images

Classes are interpreted as ordered levels of damage. The transition from the level of "no damage" to the level of "minor damage" corresponds to local deformations that do not lead to a loss of the load-bearing capacity of the building. Subsequently, "significant damage" characterizes significant violations of the integrity of the roof, floors and external walls. The level of "destroyed" corresponds to an almost complete collapse or loss of the main volume of the building. The interpretation of this scale is described in detail in works on the analysis of "xBD" and is used in modern damage assessment models as a basic one [14].

In order to take into account the domain shift, a corpus of high-resolution UAV photo images from urban areas was additionally formed. All images were normalized, reduced to the same size; class balancing was performed to avoid imbalances between classes.

At stage 1, the dataset is further marked up. We supplemented the specified corpus with a subsample of UAV aerial photographs with manual marking of building contours and damage levels consistent with the Joint Damage Scale. The number of added samples is 24. This combination of a large-scale satellite set and specialized UAV images is considered as input data for a two-stage neural-network method.

At stage 2, a segmentation model based on the YOLO family architecture is trained. The original satellite photographs of the xView2 corpus together with the added UAV aerial photographs are converted into a fixed-size sample of images with corresponding building/background raster masks obtained from vector polygons. The sample is divided into training and validation subsets taking into account the discontinuity of the areas, which makes it possible to assess the model's ability to generalize to new scenes. To increase the stability to domain shift and variability of shooting conditions, image transformations are applied: rotation, scaling, reflection, change of brightness and contrast. The network is trained to solve the segmentation problem for one class of objects – "building". As a result, a segmentation detector resistant to changes in scale, angle, and level of detail is formed, which superimposes building masks on input photographic images from satellites and UAVs.

At stage 3, a classification model based on the ViT transformer is trained. Using segmentation masks obtained from the detector model, fragments corresponding to individual buildings are cut out from the original images. After that, each fragment is assigned a class on a four-class scale of damage levels according to the initial annotations of the "xView2" dataset ("no damage", "minor damage", "significant damage" and "destroyed"). The ViT model is initialized with the weights of pre-training on a large set of natural images, and further training is performed on the formed sample of buildings using cross-entropy loss. For this stage, the number of samples for all classes is equalized, and is 11 thousand samples for each class. Accordingly, the total number of samples is 44 thousand.

So, the segmentation neural-network model provides the allocation of buildings under the conditions of domain shift between satellite and UAV data, and the classification neural-network model establishes the level of destruction of individual buildings.

The initial data of the proposed approach are pre-trained neural-network models, which will later be used to categorize the condition of destroyed buildings.

The results of the above-described approach are the input data for the proposed method of neural-network assessment of the state of destroyed buildings based on UAV images. The structural diagram of the method is shown in Fig. 2.

The input data for the method are the photo image of the area obtained from the UAV and the pre-trained neural-network models YOLO and ViT, which act as fixed parameterized segmentation and classification operators.

In the first step, an image from the UAV is fed, which is brought to a working format using standard YOLO tools: scaling to a fixed size, brightness normalization, and removal of shooting artifacts are performed.

The prepared frame is sent to the second step where the YOLO segmentation neural-network model performs scene analysis. The neural network calculates a feature map, generates a set of hypotheses about the location of the buildings. For each hypothesis, YOLO forms a rectangular window and a corresponding segmentation mask and estimates the probability that a building is actually present in the given area. After applying the threshold, a subset of segments remains, which are interpreted as the contours of individual buildings in the image.

In the third step, a cutout of the building from the photo image is formed for each selected segment using the mask: pixels outside the contour are replaced with the background, the cutout itself is scaled to the input ViT size (224 × 224), normalized, and converted into a set of patches. Patches are separate parts of the image for separate processing by the model. The ViT classification model based on patches forms a probability vector for four damage scale classes. The class with the maximum probability is taken as the level of damage of the corresponding building, and the probability itself is used as a numerical measure of confidence.

5. Results of investigating the method for assessing the condition of destroyed buildings using images

5. 1. Results of applying the steps of the method for a neural-network assessment of the condition of destroyed buildings using images

Fig. 2 shows a structural diagram of the method of neural-network assessment of the condition of destroyed buildings using images. The key is the separation of building segmentation and classification of destruction by two pre-trained models YOLO and ViT. Fig. 3 depicts the graphical result of the method in the form of localized building objects with a defined damage class and confidence level.

The method provides for the detection of individual buildings in a satellite image and their assignment to a damage scale class. The numerical values next to the objects reflect the confidence of the classification and characterize the reliability of the obtained assessment for each building.

5. 2. Results of the development of a software prototype of a two-stage image processing pipeline

For the practical implementation of the proposed method, an experimental software tool was developed that provides a full cycle of processing photo images: from their loading to the formation of generalized damage maps.

The software was implemented in the Python programming language using the PyTorch, Ultralytics, OpenCV (USA), NumPy (USA), PIL (Sweden), Streamlit (USA) libraries.

The cloud power of Google Colab with GPU T4 architecture was used to train neural networks.

The interface is implemented in the form of a desktop application (Fig. 4), which accepts photo images as input data, performs building segmentation, classification of damage to individual buildings, and visualizes the results in the form of maps and tables. The screenshot demonstrates threshold settings, visualization of building localization, and a detection table.

As can be seen from Fig. 4, the options "Show masks (for seg models)" (show masks for segmentation models) and "Show boxes/labels" (show frames and labels). The "Detections" table generates a structured result: class_id (class identifier), label (class label) and conf (confidence index) for each object.

The visualization overlays a frame with "label" (class label) and "conf" (confidence index) for each detection on the image. The user can save the results in Comma-Separated Values (CSV) or Geographic JavaScript Object Notation (GeoJSON) formats for further possible integration into geographic information systems.

a *b*

Fig. 3. Application of the method for assessing the condition of destroyed buildings on a satellite image: *a* – input image; *b* – detected buildings with damage class and confidence

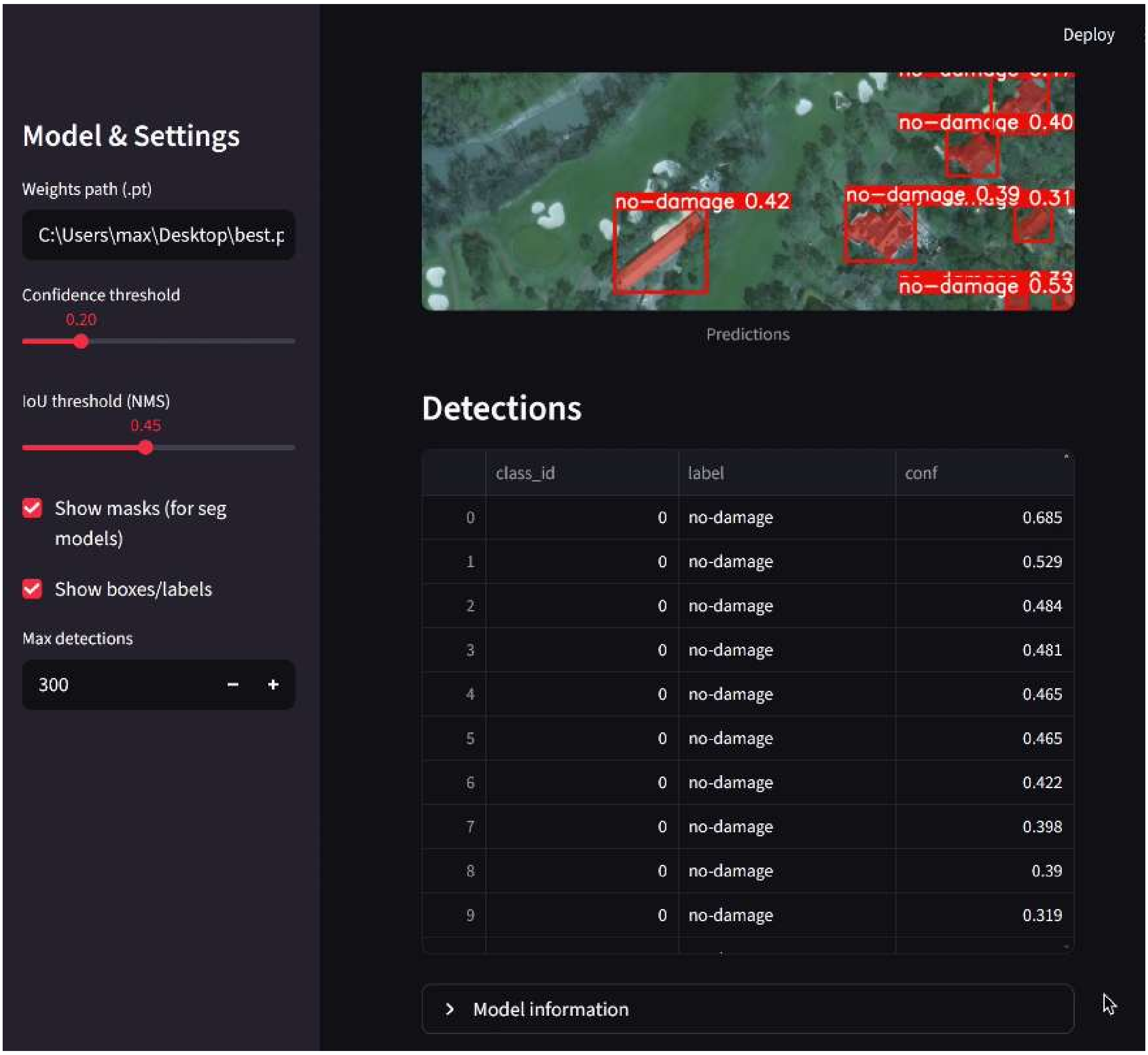

Fig. 4. Software prototype interface: visualization of detections and table of results with confidence, as well as threshold parameters

5. 3. Results of experimental evaluation of segmens tation and classification accuracy

The YOLOv8s and YOLOv11n segmentation models were trained over 100 epochs on generated samples of satellite and UAV imagery. The training results are shown in Fig. 5, 6, which demonstrate changes in accuracy and loss indicators during training. For the ViT classification model, the generalized Precision, Recall, and F_1-measure indicators were obtained, which are given in Table 1.

An example of the operation of a software prototype is shown in Fig. 7.

The best performance is observed for the extreme classes “no damage” and “destroyed”, where the value of F_1-measure exceeds 0.92. The average macro-averaged values of Precision, Recall, and F_1-measure are at 0.91, which confirms the effect of the two-stage approach. The method showed resistance to changes in flight altitude, shooting angle, and infrastructure density, which is important for further practical application.

Table 1

Result of training the ViT neural network for building condition classification

Damage class	Precision	Recall	F_1-measure
No damage	0.93	0.92	0.92
Minor damage	0.90	0.89	0.89
Major damage	0.91	0.90	0.90
Destroyed	0.92	0.93	0.92
Medium (macro)	0.92	0.91	0.91

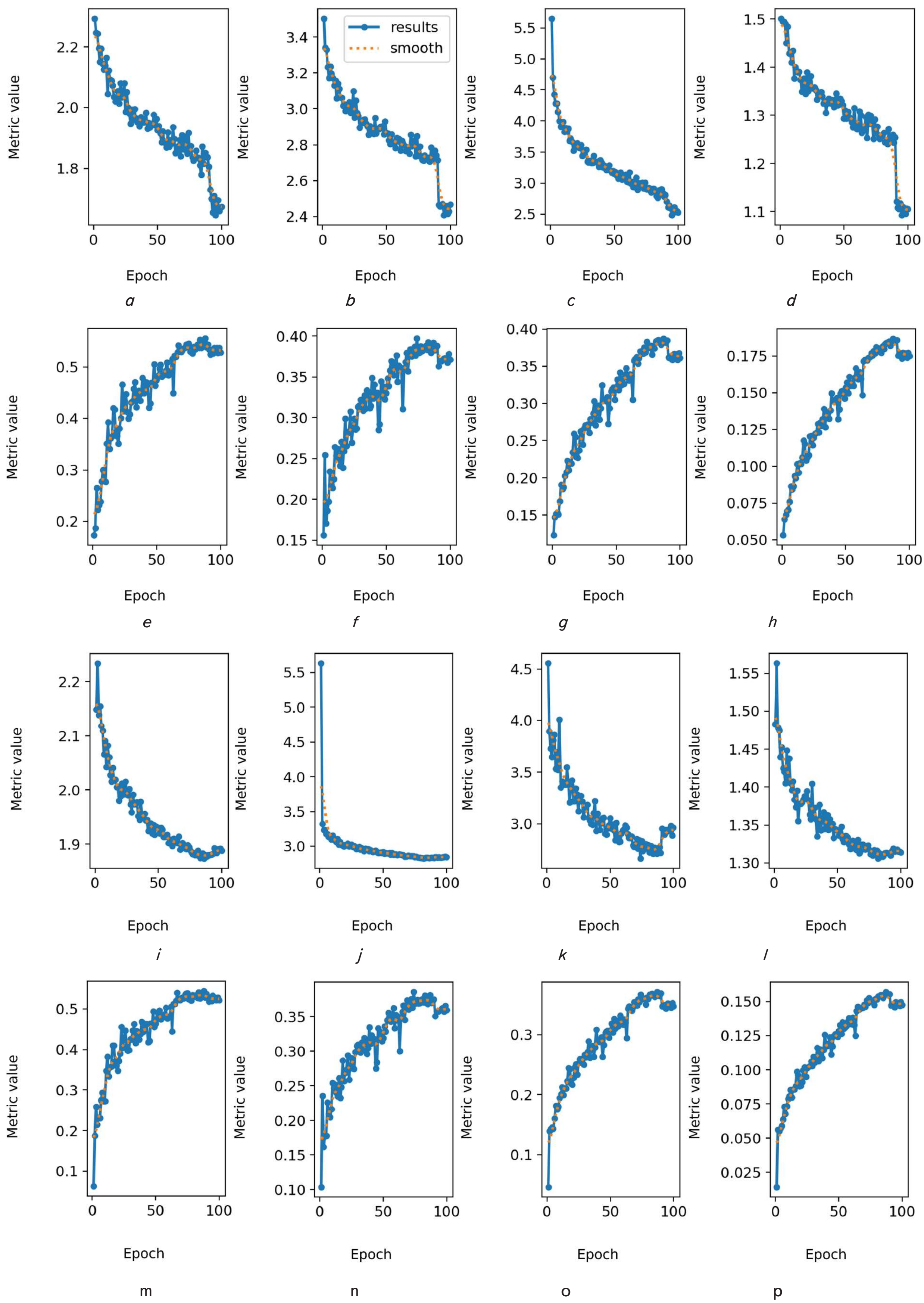

Fig. 5. Results of training the YOLOv8s neural network: *a* – train/box_ loss; *b* – train/seg_los; *c* – train/cls_loss; *d* – train/dfl_loss; *e* – metrics/precision(B); *f* – metrics/recall(B); *g* – metrics/mAP50(B); *h* – metrics/mAP50-95(B); *i* – val/box_loss; *j* – val/seg_los; *k* – val/cls_loss; *l* – val/dfl_loss; *m* – metrics/precision(M); *n* – metrics/recall(M); *o* – metrics/mAP50(M); *p* – metrics/mAP50-95(M)

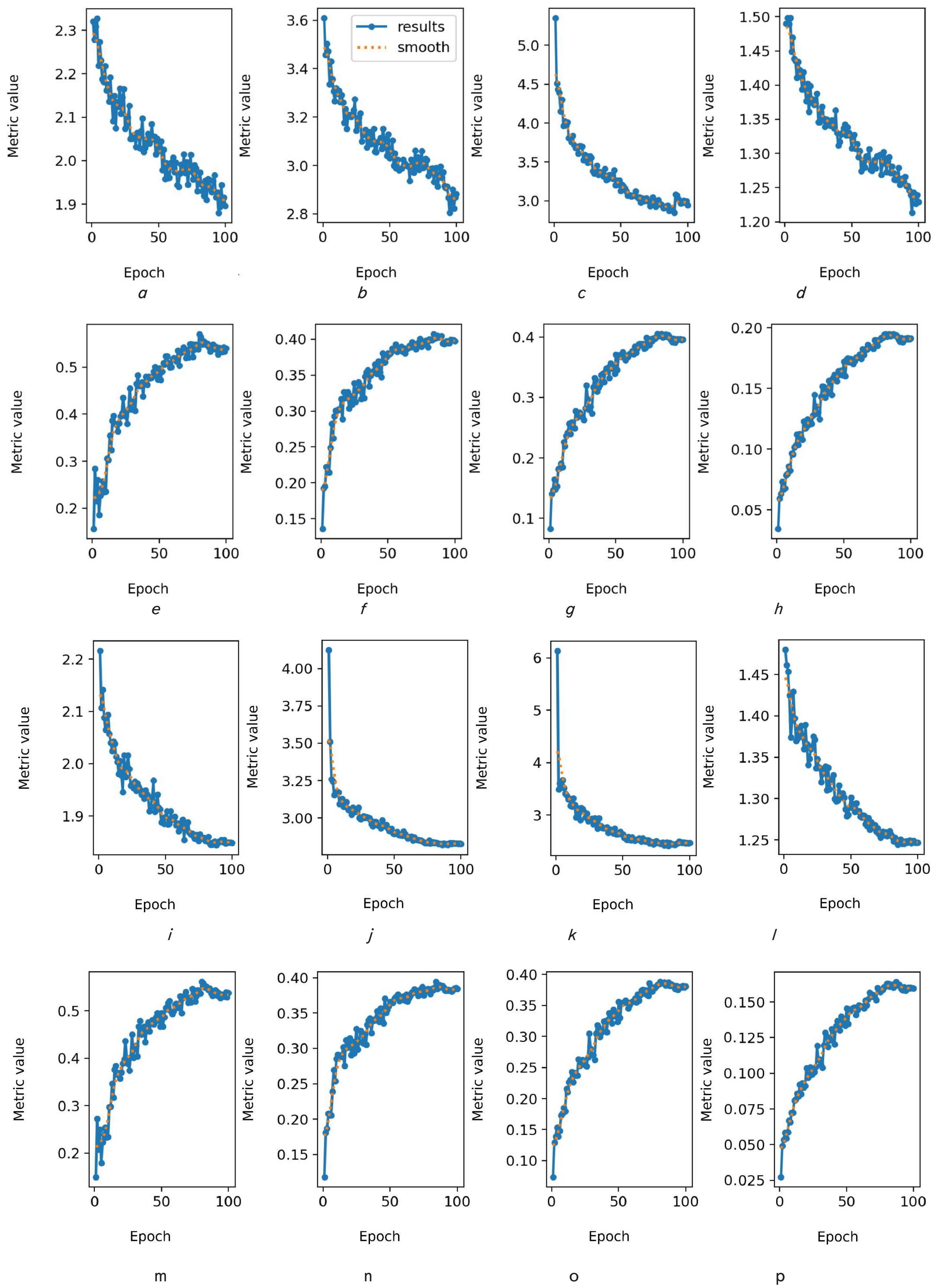

Fig. 6. Results of training the YOLOv11n neural network: *a* — train/box_ loss; *b* — train/seg_los; *c* — train/cls_loss; *d* — train/dfl_loss; *e* — metrics/precission(B); *f* — metrics/recall(B); *g* — metrics/mAP50(B); *h* — metrics/mAP50-95(B); *i* — val/box_loss; *j* — val/seg_los; *k* — val/cls_loss; *l* — val/dfl_loss; *m* — metrics/precission(M); *n* — metrics/recall(M); *o* — metrics/mAP50(M); *p* — metrics/mAP50-95(M)

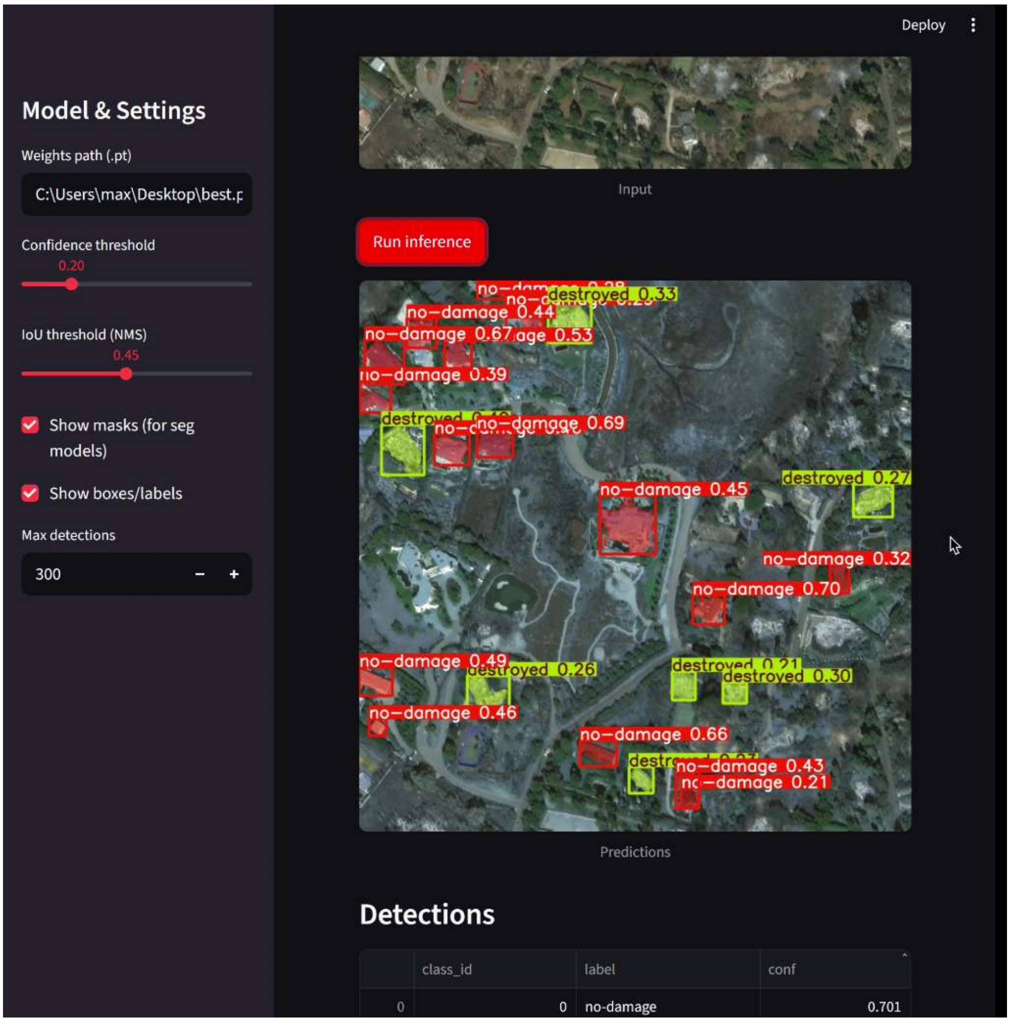

Fig. 7. Example of detection and classification of destroyed buildings

6. Discussion of the results of devising a method for assessing the condition of destroyed buildings using images

The proposed method combines segmentation of buildings, classification of the degree of damage, and contributes to the construction of spatial models of the condition of territories in a format suitable for integration into profile geographic information systems.

As a result of the proposed method, a set of segments with buildings is built for each input image, which are assigned the classes "no damage", "minor damage", "significant damage", or "destroyed". The resulting classes and numerical estimates could be used to construct damage maps and for further integration into geographic information systems and decision support systems.

Fig. 4 shows the dynamics of the loss function and integral metrics during training of the YOLOv8s segmentation model for 100 epochs. The curves for the training and validation samples decrease without sharp jumps and remain close to each other, which is consistent with the absence of pronounced overtraining. The most intense change in indicators is observed in the first 40–50 epochs, after which the plots enter a slow saturation mode with minor fluctuations around the minimum values.

A similar pattern is shown by the YOLOv11n model, the training results of which are depicted in Fig. 5. All components of the loss function monotonically decrease, and the shapes of the curves for the training and validation samples remain similar throughout the training process. This is consistent with the fact that both architectures are able to adapt to the combination of satellite and UAV images without a sharp difference in quality between the training and test parts of the training sample.

As can be seen from the results in Fig. 4, 5, the values of the integral mAP indicators for frames and masks remain moderate. This is explained by the complexity of scenes with fragmented rubble, partial destruction and variability of UAV shooting angles, as well as the influence of domain shift between data types. At the same time, the preservation of the trend towards an increase in mAP and the absence of a sharp deterioration in the validation sample indicates sufficient stability of the segmentation models.

Table 1 gives performance indicators of the ViT model for four levels of damage. The macro-averaged values of Precision, Recall, and F_1-measure exceed 0.9, which corresponds to the visual quality of classification on test fragments of buildings by a person.

Higher values of F_1-measure for classes "no damage" and "destroyed" are explained by the greater difference of their visual features compared to the intermediate states. Classes "minor damage" and "major damage" demonstrate slightly lower indicators due to the proximity of characteristic features between adjacent levels of the scale.

An example of the software prototype operation, illustrated in Fig. 5, shows the agreement of the segmentation and classification results with the reference markup.

The highlighted building contours and assigned classes correspond to the provided annotations, which confirms the operability of the method on real images.

Features of the proposed method and comparison with known approaches. Compared with studies focused on satellite data and pairs of "before/after" images [3–5], the proposed method uses UAV photographs as input data to assess the level of destruction. This makes it possible to work under conditions where images of the "before the event" state are absent or unavailable, which is typical for long-term military conflicts and is defined as an advantage of the proposed method.

Unlike solutions focused on specific types of disasters (for example, volcanic eruptions or earthquakes [4, 6]), the proe posed method is focused on a wide range of destruction in urban areas. The four-level damage scale is consistent with the approaches proposed in [4, 5] but is adapted to the conditions of post-war reconstruction with a large proportion of completely destroyed building frames. Compared to lightweight neural-network models for on-board classification [7, 8], the emphasis in our work is on combining segmentation and classification with subsequent integration into geographic information systems. The proposed method could be used to assess the condition of buildings as individual objects with the construction of damage maps at the territorial level.

The implemented separation of roles between the segmentation and classification modules is close to the concepts described in [11] but is implemented taking into account the domain shift between satellite and UAV data.

The proposed method is designed for images of sufficient spatial resolution when the contours of buildings and the main signs of destruction remain distinguishable. With a significant reduction in resolution or strong noise, the quality of segmentation and classification may deteriorate significantly. The method is focused on optical RGB images and does not use radar or multispectral channels, although they could increase the resistance to cloudiness and lighting changes.

Thus, the scope of application is currently limited to conditions where it is possible to obtain high-quality optical data. The range of flight altitudes and shooting angles at which training was conducted determines the area of correct segmentation of photographic images. In the event of a significant change in trajectories or a transition to very low or high flight, additional training of the model may be required. The four-level damage scale provides a generalized description of damage and does not cover detailed classes related to the condition of internal structures and engineering systems.

Disadvantages of the study and possible ways to eliminate them. One of the shortcomings of our work is the limited amount of UAV data that we used for post-training of segmentation and classification models. This may lead to insufficient coverage of the variety of building types, materials, and damage configurations. Another disadvantage is the lack of practical implementation of the proposed approach. In-depth experiments could more accurately quantitatively show the advantages of the selected combination of YOLO and ViT for this task. The above shortcomings can be partially eliminated by expanding the corpus of UAV images and conducting applied experiments.

Further development of the study may be associated with the use of semi-supervised and self-learning approaches to involve large arrays of unlabeled photographs from UAVs.

It is promising to refine the model to use multimodal data using radar, infrared, or thermal imaging indicators. However, this may require devising separate approaches for feature fusion and intersensor calibration.

A separate potential area of research is the adaptation of the method to on-board implementation on limited UAV computing resources. The difficulties here are related to the need to reduce the size of the models without significant loss of accuracy.

In the future, it is also possible to combine the developed method with time analysis of image series to reflect the dynamics of destruction and reconstruction.

The above areas of further research reflect the prospects for applying the proposed method to various practical tasks, which requires the adaptation of neural-network models and the organization of photo data storage.

7. Conclusions

1. Steps in a two-stage neural-network method for assessing the state of destroyed buildings using satellite and UAV images, taking into account domain displacement, have been defined. The method includes input image preparation, segmentation of buildings using the YOLO model with the formation of building contours (masks), construction of cutouts using the masks, and classification of the level of destruction using the ViT model using a four-level scale. An input data sample was formed that combines satellite images from the "xBD" and "xView2" datasets, and aerial photographs from UAVs that we created, marked using a four-level scale of building damage. The resulting sample contains over 22 thousand images and, accordingly, over 850 thousand building contours, which provides a variety of scenes and types of destruction for training neural-network models. This combination of data sources differs from typical data samples in its focus on the domain displacement between satellite and UAV images.

The feasibility of a two-stage method has been substantiated, in which the segmentation models YOLOv8s and YOLOv11n highlight the buildings while the classification model ViT assesses the level of destruction of buildings. Such separation of the roles of neural networks provides better resistance to changes in scale, perspective, and background than using one common architecture for both tasks. In the classification task by the level of damage, the proposed method provides Precision, Recall, and F_1-measure values at the level of more than 0.9, which confirms its suitability for uniform ranking of building conditions. The obtained effect is explained by the possibility of separate additional training of the segmentation module for the features of photofixation by specific UAVs and focusing the classifier on local signs of damage.

2. An experimental software tool has been developed, based on the devised method, which provides four key procedures: image loading, building segmentation, damage classification, and output of results. The results are represented in two formats – graphical and tabular. The prototype was implemented in Python using modern deep learning libraries and interface building tools, which simplifies further integration into geographic information systems. A feature of the re-

sulting solution is the combination of stages of model training and practical application in one tool, which simplifies the use of the software tool.

3. An experimental evaluation of the method has been carried out: the segmentation models achieved acceptable values of the mAP integral indicators for frames and masks in complex urban scenes with domain displacement. The ViT classification module provided macro-averaged values of Precision, Recall, and F_1-measure at the level of about 0.91–0.92 with the best indicators for the classes "without damage" and "destroyed". The combination of a two-stage architecture, pre-training, and balanced sampling ensures the achieved quality and relative uniformity of performance across damage levels.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The research was financially supported by Khmelnytskyi National University within the framework of project No. 125U003585 "Development of methods and algorithms for identification and classification of remains of destroyed buildings based on neural network analysis of visual data."

Data availability

All data are available, either in numerical or graphical form, in the main text of the manuscript.

Use of artificial intelligence

The authors used the ChatGPT-5.1 model to create illustrative elements of the diagram shown in Fig. 2. The use of AI was aimed at visually presenting the idea to simplify the reader's perception and did not affect the main conclusions of the study.

Authors' contributions

Oleksandr Mazurets: Conceptualization, Methodology, Supervision, Project administration; **Maryna Molchanova:** Conceptualization, Methodology, Writing – original draft, Investigation, Software; **Maksim Shurypa:** Validation, Investigation, Software, Visualization. **Olena Sobko:** Formal analysis, Writing – review & editing.

References

1. Mazurets, O., Sobko, O., Dydo, R., Zalutska, O., Molchanova, M. (2025). Augmented reality audiostream creation using CNN: boosting inclusion and safety for visually impaired people. CEUR Workshop Proceedings, 4004, 347–361. Available at: https://ceur-ws.org/Vol-4004/paper26.pdf
2. Krak, I., Sobko, O., Molchanova, M., Tymofiiev, I., Mazurets, O., Barmak, O. (2025). Method for neural network cyberbullying detection in text content with visual analytic. CEUR Workshop Proceedings, 3917, 298–309. Available at: https://ceur-ws.org/Vol-3917/paper57.pdf
3. Gu, J., Xie, Z., Zhang, J., He, X. (2024). Advances in Rapid Damage Identification Methods for Post-Disaster Regional Buildings Based on Remote Sensing Images: A Survey. Buildings, 14 (4), 898. https://doi.org/10.3390/buildings14040898
4. Xia, H., Wu, J., Yao, J., Zhu, H., Gong, A., Yang, J. et al. (2023). A Deep Learning Application for Building Damage Assessment Using Ultra-High-Resolution Remote Sensing Imagery in Turkey Earthquake. International Journal of Disaster Risk Science, 14 (6), 947–962. https://doi.org/10.1007/s13753-023-00526-6
5. Singh, D. K., Hoskere, V. (2023). Post Disaster Damage Assessment Using Ultra-High-Resolution Aerial Imagery with Semi-Supervised Transformers. Sensors, 23 (19), 8235. https://doi.org/10.3390/s23198235
6. Tennant, E., Jenkins, S. F., Miller, V., Robertson, R., Wen, B., Yun, S.-H., Taisne, B. (2024). Automating tephra fall building damage assessment using deep learning. Natural Hazards and Earth System Sciences, 24 (12), 4585–4608. https://doi.org/10.5194/nhess-24-4585-2024
7. Alsaaran, N., Soudani, A. (2025). Deep Learning Image-Based Classification for Post-Earthquake Damage Level Prediction Using UAVs. Sensors, 25 (17), 5406. https://doi.org/10.3390/s25175406
8. Zhuang, X., Tran, T. V., Nguyen-Xuan, H., Rabczuk, T. (2025). Deep learning-based post-earthquake structural damage level recognition. Computers & Structures, 315, 107761 https://doi.org/10.1016/j.compstruc.2025.107761
9. Lyu, C., Lin, S., Lynch, A., Zou, Y., Liarokapis, M. (2025). UAV-based deep learning applications for automated inspection of civil infrastructure. Automation in Construction, 177, 106285. https://doi.org/10.1016/j.autcon.2025.106285
10. Zha, Q., Yao, Y., Zheng, Y., Ma, W., Zhang, W. (2025). A dataset of building surface defects collected by UAVs for machine learning-based detection. Scientific Data, 12 (1). https://doi.org/10.1038/s41597-025-06318-5
11. 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. Available at: https://ceur-ws.org/Vol-3970/PAPER14.pdf
12. xView2 Challenge Dataset - train and test. Available at: https://www.kaggle.com/datasets/tunguz/xview2-challenge-dataset-train-and-test/data
13. Sobko, O., Mazurets, O., Molchanova, M., Krak, I., Barmak, O. (2025). Method for analysis and formation of representative text datasets. CEUR Workshop Proceedings, 3899, 84–98. Available at: https://ceur-ws.org/Vol-3899/paper9.pdf
14. Asif, A., Rafique, H., Jadoon, K., Zakwan, M., Mahmood, M. H. (2024). Change-centric building damage assessment across multiple disasters using deep learning. International Journal of Data Science and Analytics, 20 (3), 1915–1931. https://doi.org/10.1007/s41060-024-00577-y