GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 20 • Number 5 • 2026

https://doi.org/10.7494/geom.2026.20.5.59

Enhancing the Urban Agglomeration Management Combining UAVs, Deep Learning, and WebGIS in Ho Chi Minh City, Vietnam

Ngoc Huyen Trang Tran[1, 2] *, Van Trung Le[1], Le Phu Vo[1]

[1] Vietnam National University Ho Chi Minh City (VNU-HCM), Ho Chi Minh City University of Technology (HCMUT), Faculty of Environment and Natural Resources, Ho Chi Minh City, Vietnam

[2] Ho Chi Minh City University of Natural Resources and Environment (HCMUNRE), Faculty of Geodesy, Cartography and Construction, Ho Chi Minh City, Vietnam

* Corresponding author

Abstract: Ho Chi Minh City (HCMC), the biggest city and most urban agglomeration in Vietnam, has witnessed the rapid growth of urban expansion over the past several decades. However, the city authorities faced a wide range of constraints in handling urban management and planning issues. Among these, the management practices of urban construction activities and planning tasks are the most challenging due to the low quality of available data, lack of data synchronization between management departments, and the limitations in monitoring construction activities. These have resulted in poor performance in urban planning. In this study, Unmanned Aerial Vehicles (UAVs) were deployed to take photos of the existing construction activities in a residential area of HCMC. The traces of houses will be segmented and extracted automatically using a deep learning (DL) model. Using UAVs combined with DL is a solution that demonstrates high automation, which can create standardized and uniform datasets. A 3D map of the area was created and shared on the WebGIS platform. Comparing the construction status with the legal boundaries of the land plot, the horizontal and vertical construction permit standards will help detect illegal construction objects. The results of the study show that the combination of UAV images and DL shows strong potential for building 3D maps, with the accuracy of the ground and elevation being 4.8 cm and 11.9 cm, respectively. This result suggests a promising solution to assess construction status with an appropriate accuracy corresponding to a map scale of 1:500. Importantly, the achieved 3D map shared on WebGIS is effective for use as a data source between the line government agencies.

Keywords: building footprints, urban management, semantic segmentation, transfer learning, digital urban mapping

Received: December 11, 2025; accepted: May 23, 2026

E-mails and ORCID iDs:
tnhtrang.sdh222@hcmut.edu.vn, https://orcid.org/0000-0001-5589-2251 (N.H.T.T.);
lvtrungbk@gmail.com, https://orcid.org/0000-0002-0885-8656 (V.T.L.);
volephu@hcmut.edu.vn, https://orcid.org/0000-0002-5342-9324 (L.P.V.).

1. Introduction

1.1. Managing Urban Construction Activities in Ho Chi Minh City: Current Practices and Issues

The management practices of construction activities involve the control and supervision of a sequence of mandatory procedures, including registration, permit issuance, construction, surveillance, inspection, and handover of the works constructed. The purpose of the management of construction activities is to ensure safety, compliance with regulations, the rights of relevant parties, and sustainable development within the managed area. The management of construction activities is carried out by state management agencies, comprising Project Management Boards, Departments of Construction, Departments of Construction Management, and units managing land resources, labor safety, environmental protection, and the community [1]. The management of construction activities requires close cooperation between government agencies, investors, contractors, and other relevant parties. It also requires strict enforcement of legal regulations, including regulations on construction, labor safety, environmental protection, urban planning, and other regulations related to construction activities. In the framework of this study, we refer to the work of construction acceptance inspection; that is, to examining whether the aspects of the completed work are in accordance with the previously granted license [2–4].

Ho Chi Minh City (HCMC), a rapidly urbanizing municipality, is facing many challenges in the management of construction activities. The common challenges include illegal construction and permitted height violations, which have resulted in negative impacts on the urban planning scheme and caused nuisance and a loss of amenity for urban dwellers [5]. To effectively address these issues, the application of the latest aerial and informatics technologies is necessary, particularly for improving the timeliness, objectivity, and spatial completeness of urban monitoring data. In this context, UAVs, deep learning (DL) and WebGIS can provide an integrated technical basis for more effective management practices of urban construction [6–8].

1.2. Emerging Technologies in Supporting Spatial Urban Planning and Management

With the rapid emergence of new aerial technologies, UAVs in particular have resulted in the majority of advances in socio-economic planning and management through high-resolution imagery, real-time data collection and the capability to create 3D (three-dimensional) maps at a relatively low cost. This makes UAVs highly applicable in many fields. In the urban sector, UAVs offer significant advantages, such as monitoring construction progress, detecting violations of construction regulations, assessing the condition of trees, and monitoring traffic conditions [9, 10]. Moreover, the use of UAVs in these fields not only helps to reduce costs and save time but also ensures safety standards and minimizes risks to human resources. With these outstanding advantages, UAVs are expected to continue to play an important role in sustainable urban management and development in the future [11, 12].

The use of 3D city maps is increasing in many scientific and practical fields, including urban planning and architecture, education, design and advertising, telecommunications, tourism, transportation, meteorology, security (police, military), etc. Most large and small cities are embarking on establishing a GIS system, which is built on a base map featuring assets, infrastructure, environmental objects, etc. [13]. One of the most useful tools for building smart cities is 3D mapping, as the third dimension not only improves visualization but also supports spatial analysis, scenario evaluation, and decision-making in urban environment management. In recent studies, 3D city models have been increasingly recognized as core components of digital urban environments and smart-city applications. Currently, 3D smart city models are being deployed in many major cities around the world, such as Hamburg, Helsinki, Abu Dhabi, Chicago and London, to serve urban master planning in a more reasonable way and to assess the impact of new development projects and public services [14, 15]. With 3D mapping capabilities, UAVs enable the acquisition of topographic elevation and building height information, which can then be compared with approved spatial plans [16]. This technology is particularly useful in monitoring hard-to-reach areas and places where direct measurements cannot be made by traditional methods. Furthermore, 3D maps created from UAV data not only assist state management agencies in identifying violations of construction space (such as land encroachment, boundary crossing) but also help to detect violations of building height, an important factor for ensuring aviation safety and compliance with urban planning and urban aesthetics. In addition, 3D data obtained from UAVs can be updated periodically to monitor changes in urban structure over time, thereby assisting management agencies to promptly detect abnormal changes and effectively handle violations related to construction order [17].

DL has dramatically transformed the fields of image recognition, segmentation, and extraction, particularly with respect to images captured by UAVs. DL demonstrates outstanding efficiency in UAV image analysis thanks to its ability to process large amounts of data while maintaining a high level of accuracy. At the same time, DL is highly flexible and can adapt to a variety of image types and environmental conditions. In addition, its ability to learn representative features from large datasets has made DL an increasingly essential tool for automated spatial data extraction, including building footprint delineation, urban object detection, and thematic mapping from remote sensing imagery [18–20]. Recent review studies have further highlighted the important role of DL in automated building extraction from remote sensing data, particularly in complex urban environments [21]. The synergy between UAVs and DL technology enables the automatic analysis and identification of violations from image data and 3D maps. DL algorithms are capable of swiftly detecting unusual changes, classifying construction zones by function or land use, and thereby reducing reliance on manual monitoring processes. More importantly, the combination of UAV imagery and DL facilitates accurate, real-time insights, greatly enhancing the efficiency and effectiveness of urban management and other critical activities [22, 23].

When data from UAVs and DL are integrated into the WebGIS system, management agencies can visualize spatial information and make timely decisions. WebGIS serves as a centralized, transparent, and easily shared data management platform among stakeholders, thereby improving efficiency in the management practices of construction activities and other urban activities [24, 25]. In a broader smart-city context, such web-based spatial platforms can also support interactive digital urban environments and more responsive urban governance [8].

The objective of this study was to develop and evaluate an integrated workflow using UAV imagery and DL functions within Esri's ArcGIS Pro software environment to segment and extract building footprints and tree locations along urban routes. This study also calculated object heights from classified point clouds so that the resulting spatial and height information could be shared on a WebGIS platform. Rather than focusing solely on the individual use of UAVs, DL, or WebGIS as separate tools, this study examines their combined ability to support urban monitoring and management in a more efficient and practical way. This provides urban managers with a comprehensive and intuitive overview, offering a more complete basis for monitoring construction activities and urban greenery.

The combination of UAVs, DL techniques, and WebGIS platforms not only improves the capacity for monitoring, analyzing, and managing construction orders but also contributes positively to the process of building HCMC into a smart, modern, and sustainable city. The novelty of this study lies in the integration of these components within a unified workflow that links data acquisition, object extraction, height assessment, and web-based spatial management, thereby providing a more application-oriented solution for urban management in HCMC.

2. Materials and Methods

2.1. Study Area

Phuoc Thien resettlement area, located in Long Binh Ward, Ho Chi Minh City (HCMC), is situated in a rapidly urbanizing area in the eastern corridor of the city (Fig. 1). Formerly belonging to Thu Duc City, this area is now directly managed by HCMC after the recent rearrangement of administrative units. Nationally, HCMC plays the central role in Vietnam's economic growth and modernizing urban infrastructure [26, 27]. In recent years, the city has advocated expanding the urban space in the Eastern corridor to meet population needs, arrange resettlement and develop urban satellite areas. Phuoc Thien is a typical example; well-planned, located next to the Vinhomes Grand Park urban area, and directly connected to strategic infrastructure axes such as the highway of HCMC – Long Thanh – Dau

Giay and Ring Road No. 3. This area has invested in a modern and synchronous technical infrastructure system, including underground electricity, water, telecommunications lines, internal traffic systems, public lighting and a well-distributed network of trees. In the 3D digital spatial database and digital urban map, the main objects shown include buildings, streetlights, trees, and traffic systems, while underground structures are not visible due to the technical limitations of UAV imagery. Notably, the Phuoc Thien area is located on the corridor of Metro Line No. 1 (Ben Thanh–Suoi Tien) – the first urban railway line of HCMC and the second in Vietnam – considered a symbol of modern public transportation [28]. This Metro line plays an important role in connecting the city center with the new development area in the Eastern corridor, supporting labor mobility and population allocation, and promoting regional integration in the process of urbanization. With the advantages of geographical location, improved infrastructure, and an interregional transportation system, the Phuoc Thien area is gradually forming its role as a new urban center of HCMC. This development not only attracts residents and a qualified workforce but also reflects the trend of expanding urban space in a modern and sustainable direction [29, 30].

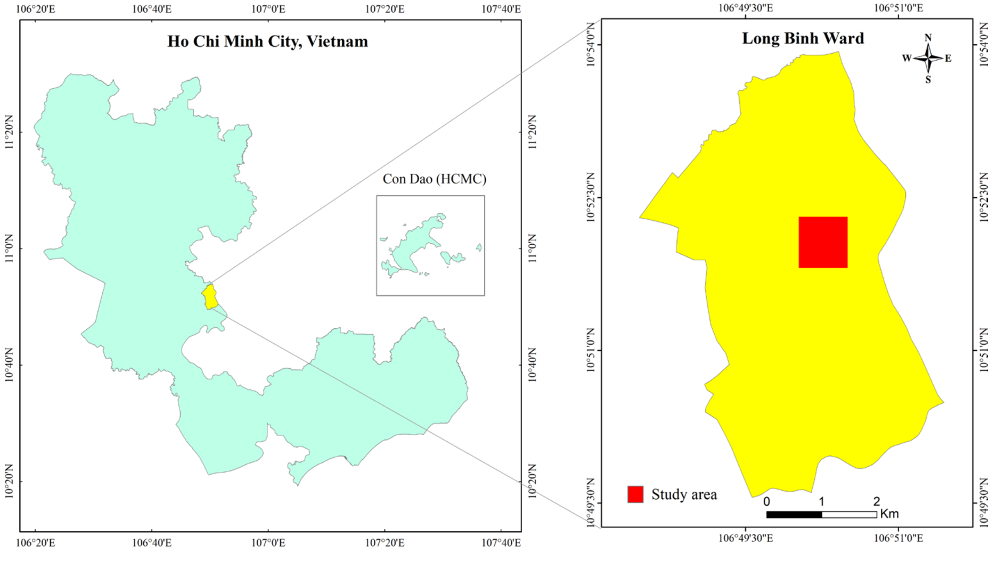

Fig. 1. Study area in Phuoc Thien resettlement area

Therefore, the establishment of a modern management system employing technologies such as UAVs, AI, and WebGIS is necessary to support managing urban construction activities, monitoring green spaces, and maintaining a healthy living environment. This system not only helps to monitor and detect violations but also sets the foundation for Thu Duc City to become a model smart city.

2.2. Conceptual Framework of Procedure

The framework of the research methodology is the integration of three main components: UAVs, DL, and a WebGIS system (Fig. 2). This approach not only ensures technical accuracy in each step but also aims at the ability to share, exploit, and update data for effective urban management.

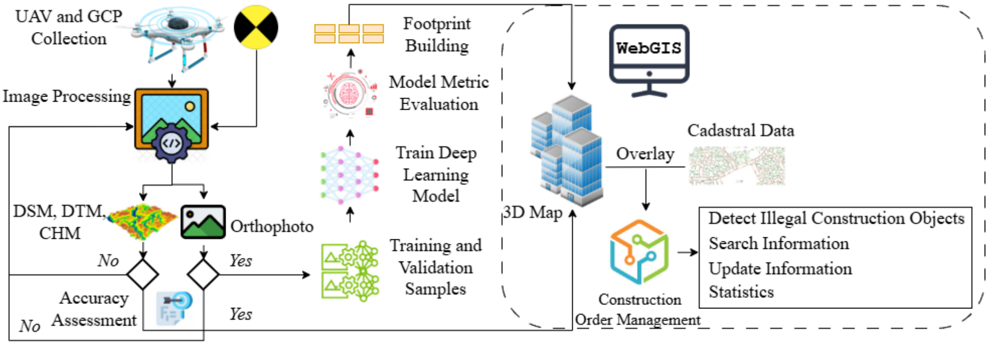

Fig. 2. Procedure of integrating UAVs, deep learning, and WebGIS in urban construction activities management

The first component, UAVs, performs the task of taking high-resolution digital images using unmanned aerial vehicles, combined with ground control points (GCPs) to ensure coordinate accuracy. The resulting images will be processed by specialized software to create orthophotos and digital elevation models (DEMs) such as digital surface models (DSMs), digital terrain models (DTMs) and normalized digital surface models (nDSMs). These models play a crucial intermediate role in determining the height, spatial structure, and layering of information according to the depth of the urban area [31, 32].

The second component is DL, in which convolutional neural network (CNN) models are trained and fine-tuned to automatically recognize and extract construction boundaries from UAV imagery. The use of DL not only replaces the manual digitization process but also improves the accuracy and efficiency of geometric data extraction. The output is a building footprint layer in the form of pre-standardized vector data for integration into the geographic information system (GIS) [33, 34].

The third component is WebGIS, which takes on the role of storing, organizing, displaying, and sharing spatial data in the form of interactive maps. All processed data (orthomosaic, digital spatial data, DEMs, cadastral data) are integrated into the WebGIS system to support the functions of generating statistics as well as looking up, updating, and checking the current construction status. This system serves as a digital platform for monitoring and managing urban construction activities.

The process of integrating UAV–DL–WebGIS in this study not only fully ensures the processing chain from data collection–processing–analysis–management but also reflects the high feasibility of deployment in developing cities. The final result is a large-scale spatial digital data layer and 3D urban maps, fully reflecting the necessary management factors such as building boundaries, height, density, management boundaries, and technical infrastructure. This research methodological framework not only contributes to technologically but also creates an academic foundation for urban management solutions based on digital data in the context of today's digital transformation.

2.3. Data Collection

Data Collection by UAV and GNSS RTK

In order to collect data for the study area, a DJI Phantom 4 Pro UAV with PPK (Post-Processed Kinematic) positioning technology was deployed, as illustrated in Figure 3. This device possesses a high-resolution optical sensor, which enables the recording of detailed, high-quality aerial images, and PPK technology helps to calibrate the post-flight positioning signal, thereby improving the positioning accuracy of the image [35, 36]. The flight shooting process is designed in two different configuration options to serve the analysis requirements of coordinate accuracy and the level of geometric detail of aerial objects.

a) b)

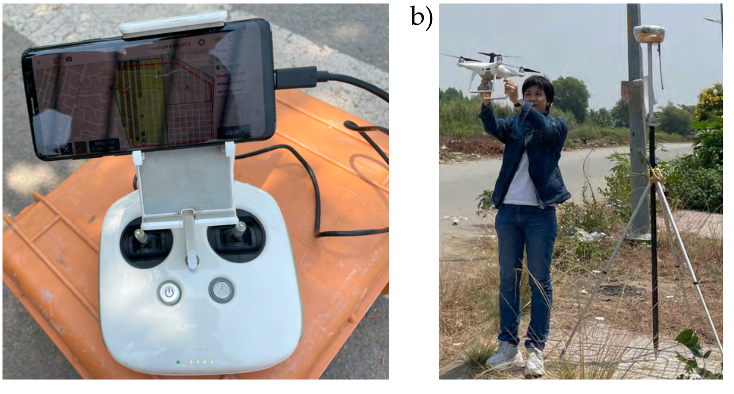

Fig. 3. UAV and remote controller used during field survey:
a) remote controller; b) DJI Phantom 4 Pro PPK

In the first flight design, the UAV was operated in waypoint flight mode, which means that the flight trajectory is pre-set and the vehicle automatically follows the programmed route. The flight altitude was fixed at 100 m, and the shooting axis is positioned perpendicular to the ground (90° angle) in order to obtain an orthogonal image with low geometric distortion. The overlapping was set at 80% for both the along-track and the across-track, ensuring the level of overlap required for geometric reconstruction and spatial modeling. With this configuration, two flights were deployed, and a total of 458 images were obtained.

In the second flight design, the UAV was operated according to the double-grid flight mode to enhance the ability to reproduce 3D spatial structures. In this configuration, the camera was adjusted to shoot at an 80° oblique angle, which helped to capture more information about the façades of the buildings and improve the quality of the 3D model with an overlapping of 70% for both the along-track and the across-track. With this option, four flight missions were carried out at an altitude of 100 m, obtaining a total of 981 UAV images.

Meanwhile, a Hi-Target V30 RTK GNSS was also employed to measure the spatial coordinates of the ground control points (GCPs) and check points (CPs) precisely, as shown in Figure 4. Five GCPs were similarly spaced throughout the study area, and were located to ensure an average distance between points of approximately 370 m, in order to calibrate the image model and ensure geometric accuracy in the application of the UAV data. Five CPs were also employed as independent datasets to validate the accuracy of the orthomosaic and DEMs.

a) b)

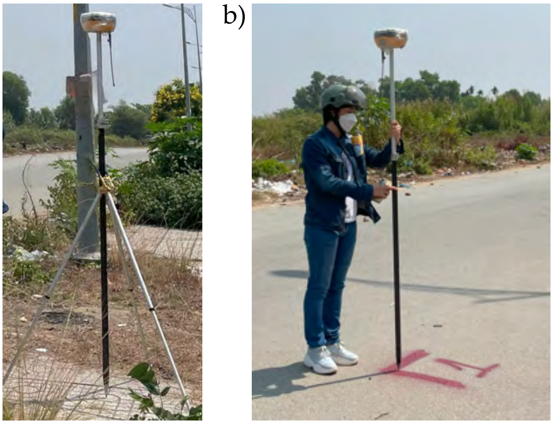

Fig. 4. Hi-Target V30 RTK GNSS: a) base station; b) rover

Processing UAV Data

The Agisoft Metashape Professional software was used to process the six flight missions (1,439 images), setting the processing level to medium. The measurement information and images were referenced in the VN-2000 National Reference System and Coordinate System, based on UTM projection with a 3° projection zone having a central meridian of 105°45′ for HCMC [37]. All outputs were then generated in this coordinate system.

A point cloud was utilized to generate a DSM (Fig. 5a). This point cloud was then further classified into ground and non-ground points. The ground points were interpolated to generate the DTM (Fig. 5b). Orthophotos were also produced at a 2.82 cm resolution. To represent a 3D map, the height of the object layers must be considered. The normalized DSM (nDSM) or canopy height model (CHM) represents the height of objects above the surface. This model is derived by subtracting the DTM from the DSM (Fig. 5c).

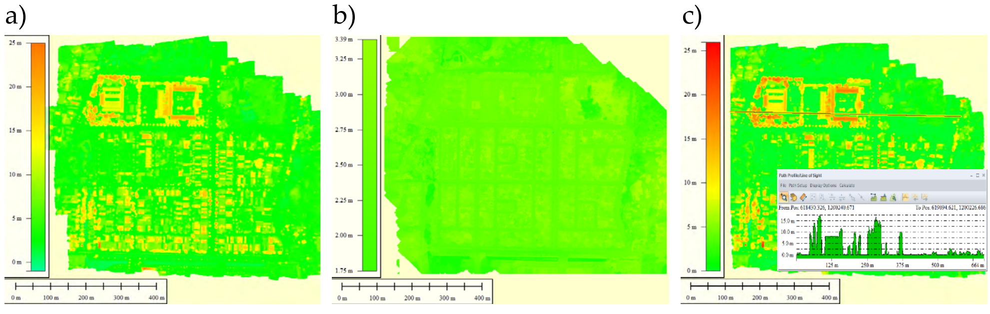

Fig. 5. Types of Digital Elevation Models: a) DSM; b) DTM; c) CHM

The heights of 15 objects in the area comprising nine houses, two light poles and four trees were randomly measured using a total station in order to verify the accuracy of the CHM. Five CPs were used to assess the accuracy of the orthomosaic. The root mean square error (RMSE) was used to calculate the accuracy as follows [38]:

$$m_X = \sqrt{\frac{\sum_{i=1}^{n} \Delta X_i^2}{n}} \tag{1}$$

$$m_Y = \sqrt{\frac{\sum_{i=1}^{n} \Delta Y_i^2}{n}} \tag{2}$$

The horizontal RMSE was calculated to evaluate the positional accuracy of the orthomosaic:

$$m_{XY} = \sqrt{m_X^2 + m_Y^2} \tag{3}$$

where: n – number of checkpoints, $\Delta X_i = X_i' - X_i$, $\Delta Y_i = Y_i' - Y_i$, X_i', Y_i' – coordinates extracted from the orthomosaic, and X_i, Y_i – coordinates of checkpoints.

The vertical RMSE was calculated to evaluate the elevation accuracy of the CHM:

$$m_H = \sqrt{\frac{\sum_{i=1}^{n} \Delta H_i^2}{n'}} \tag{4}$$

where: n' – number of test objects, $\Delta H_i = H_i' - H_i$, H_i' – height extracted from CHM, and H_i – height of test objects.

Processing Cadastral Map Data

Cadastral data was collected and processed in order to provide a legal basis for the comparison of the current state of construction in the study area. The cadastral map at 1:500 scale, reflecting the legal boundaries of land parcels under the formal land management system, is used as the main reference layer to determine the degree of conformity between the current land use status and the granted land use rights. Land parcel boundary data was transformed and standardized in terms of both geometry and properties to ensure spatial integrity and logical connectivity in GIS analysis.

The synchronous integration between spatial boundaries and attribute information, such as the owner's name, area, purpose of use, and land parcel number, enables the effective exploitation of cadastral data as a highly reliable matching base. Thereby, the process of assessing and detecting cases of boundary violations, illegal permits, or misuse of land can be carried out systematically and transparently, serving the inspection and management of construction in urban areas.

2.4. Using Deep Learning to Extract Building Footprints

The process of applying DL in this study is implemented through four main stages: (1) identifying the need for extraction and designing a strategy for refining the model, (2) building a training sample set, (3) training and refining the model according to local data, and (4) applying the model to extract objects and convert them to GIS data.

Rationale for Selecting the Training Strategy Using Transfer Learning

When the number of labeled samples is still limited, the study uses transfer learning techniques to improve the effectiveness of DL model training. This approach takes advantage of pre-trained models on a large and rich dataset, thereby shortening the training time and reducing the pressure on computational resources [39, 40]. The refinement of the model is carried out on the basis of the local dataset, which helps the model to better adapt to the specific spatial and contextual characteristics of the study area.

The training strategy chosen was to lock all the backbone layers and refine only the output layers (heads) suitable for each extraction task, thereby maintaining the generalization of the original model while learning the specific characteristics of the new data [41, 42]. The total number of parameters updated accounts for approximately 10% of the number of parameters throughout the model, in order to ensure a balance between efficiency and stability in training.

Data augmentation techniques such as rotation, flipping, stretching, and changing the brightness of the image were also systematically applied to expand the training set and improve the diversity of the sample, contributing to minimizing overfitting and improving the accuracy of the model. The selection of the above training strategy makes it possible to optimally exploit the performance capacity of the DL model with low computing costs while also enhancing the applicability of the model to UAV imagery with complex acquisition conditions and strong changes in urban landscapes.

Construction of the Training Dataset

The training dataset was constructed to support the fine-tuning of the DL model under the selected transfer learning strategy. For building-boundary segmentation, the input data were prepared from UAV imagery, with a ground sample distance of 2.82 cm/pixel, three RGB color bands, and 8-bit radiometric resolution, ensuring the ability to clearly identify the boundaries of small objects in urban areas.

The UAV imagery was collected at the Phuoc Thien resettlement area and supplemented with images from several other areas in HCMC to increase the representation of different building types. After acquisition, the data were pre-processed through radiometric adjustment, tile-size normalization to 512 × 512 pixels, and removal of noisy areas without relevant object information.

The labeling was performed manually in ArcGIS Pro using the "Object Labeling for Deep Learning" add-in. The building footprints were digitized as labeled samples to ensure that the target objects were accurately represented in the training data. The output consisted of image tiles, label layers, and associated metadata files. With segmentation networks such as DeepLabV3+, the data was organized in a "Classified Tiles" format suitable as an input format for the training procedure.

The minimum number of samples required for training was estimated based on the model size, the proportion of fine-tuned parameters, and a spatial complexity coefficient reflecting the heterogeneity of the area. Specifically, for DeepLabV3+ (ResNet-101), with a total of 48 million parameters, if 10% of the last layer is fine-tuned, 4.8 million parameters need to be learned [43, 44]. Based on this calculation, and after applying a 16-fold data augmentation factor, the minimum number of original samples required for training was determined to be 1,960.

The enhanced dataset was divided into 70% for training, 20% for validation, and 10% for testing. This division was intended to ensure an objective evaluation of the model, avoid overfitting, and help test the ability to generalize the model on actual data.

Deep Learning Model Training Using Transfer Learning

The DeepLabV3+ model with a ResNet-101 backbone was selected for this study because of its strong capacity to represent both geometric and contextual information for segmenting objects with clear boundaries, such as buildings. ResNet-101 is a deep convolutional neural network featuring skip connections that enable effective signal propagation across multiple layers, mitigating gradient vanishing, thereby supporting the learning of high-level semantic features. In addition, DeepLabV3+

incorporates ASPP (Atrous Spatial Pyramid Pooling), which expands the receptive field without reducing the spatial resolution, while capturing multi-scale contextual information. This ability is particularly useful in identifying buildings of different shapes and sizes.

In the fine-tuning process, the model learns hierarchical features that progress from simple to complex patterns, matching the geometric structure of the urban data. In the early stages of the model, the filter responses are devoted to very simple concepts: edges, corners, and boundaries, which are important for distinguishing building outlines from surrounding image textures. In the intermediate layers, it learns structural regularities and surface materials, such as the corrugated iron, concrete and brick roofs very often found in the study area. Finally, deep layers aggregate information across different regions of the image to capture higher-level morphological characterization, enabling the model to learn the shape and layout of the building as a whole, even for views that are occluded by environmental factors such as trees or shadows.

During the training process, transfer learning techniques were used to take advantage of the weights learned from a large source dataset. The early layers of ResNet-101 are preserved (frozen), while the ASPP block and final-resolution layers are fine-tuned to adapt to the UAV image characteristics of the study area. This strategy allowed faster convergence and improved training effectiveness under limited-data conditions.

The training was performed for 150 epochs with a batch size of 8 and a learning rate of 0.001. These parameters were chosen based on preliminary experiments showing stable convergence and satisfactory performance.

Extraction of Building Footprints

On completion of the training process, the DeepLabV3+ model was deployed in ArcGIS Pro and applied to the orthomosaic imagery of the entire study area using the Classify Pixels Using Deep Learning tool. The output was a binary segmentation raster, in which each pixel was assigned a value of 1 if it belonged to a building and 0 otherwise.

This raster layer was then processed by the "Raster to Polygon" tool to convert it into vector polygons, forming a construction boundary layer in the form of digital spatial data. Post-processing steps were applied to eliminate noise, simplify boundaries, and increase geometric accuracy before integration into spatial databases.

2.5. Construction of GIS-Based Spatial Data for Urban Construction Management

The building footprint layer generated by the DL model with post-processing will be imported and managed in the GIS to become a spatial data layer for building inspection and supervision. The information is organized into two parts: the polygon spatial layer and the descriptive attributes.

Attribute fields are constructed in four functional groups:

1) Height attribute: are extracted from the CHM by assigning the z-value at the center of the building coordinates. The "Height" field can be compared with the construction permit based on the number of floors, supporting the detection of over-the-top construction.
2) Legal attributes: include information such as house number, human owner data, land use, linked from table data or cadastral maps through an attribute join or spatial join.
3) Geometric properties: are calculated by the "Calculate Geometry" tool and include area, perimeter and center coordinates. These properties support the evaluation of scale and construction density and checking the validity with the permit.
4) Architectural attributes: reflect the roof type, material, and façade characteristics extracted from the 3D mesh model. These fields support the control of urban aesthetics and architecture.

The attribute data not only serves to visualize the space but also plays an important role in checking the construction order, helping to detect violations such as building the wrong boundaries, exceeding height limits, changing the purpose of use, or wrong façade planning. Data is periodically updated from new UAV imagery and DL analysis results, contributing to the construction of a continuous, automated, and transparent monitoring system in the urban space.

2.6. Development of a WebGIS Platform for Management

Following the generation and homogenization of the digital spatial data from UAV imagery and cadastral data, a WebGIS system was developed to establish an Internet-based platform that supports monitoring and managing urban construction development. The system not only provides a multi-dimensional data visualization tool but also integrates analytical, violation-detection, and decision-making support functions in the context of rapid urbanization in resettlement areas.

The WebGIS platform was designed according to a three-layer architecture, including:

1) PostgreSQL/PostGIS spatial database, which stores all classes of normalized data from UAV imagery, altitude models, cadastral data, and results extracted from DL models;
2) GeoServer map servers, used to publish data via OGC (Open Geospatial Consortium) standard services such as WMS (Web Map Services) and WFS (Web Feature Services), thus supporting the display and querying of spatial data;
3) web application layer, based on Node.js/ExpressJS and OpenLayers (2D) and Cesium (3D) via OLCS (OpenLayers–Cesium), offers a user-friendly interactive interface for the display of both 2D maps and 3D maps concurrently.

The system was developed with the following main functional groups to meet the requirements of urban management:

- Decentralization of access and administration: Allows setting up user roles, controlling access by function and geography.
- Visualization and analysis of spatial data: The system provides tools to support the display of data by layer, monitoring of the current status, comparison with planning, and implementation of basic spatial analysis for evaluation and decision-making.
- Search and look up information: Allow queries by attribute or geographical location and support quick retrieval of information related to constructions, land parcels and areas with indicators of violation.
- Data update and editing: Support the function of editing directly on the digital map, allowing users to conveniently update field data, edit property information, or add new constructions.
- Statistics and reporting: Provide statistical tools according to time, type of construction, scale and degree of violations; support the development of quick reports for urban management and supervision.

One of the important functions of the system is to check for violations of the building order based on the matching between the current status data and the approved plan. The building footprint layer extracted from the DeepLabV3+ model is superimposed on the cadastral land parcel boundary layer, thereby identifying cases of construction that exceed the boundary or are not in the right location according to legal documents. Buildings that extend beyond the legal land boundary will be automatically marked by the system for field inspection.

In addition, to assess violations of building height, the processing steps were carried out as follows: the building footprint layer was converted into a representative point layer using the "Feature to Point" tool, then the elevation value was extracted from the CHM model using the "Extract Value to Points" tool. These height values are connected to the polygon footprint layer through a spatial join, which allowed the height of each building to be determined. According to the planning of the study area, constructions must ensure a standard height of 4 floors, equivalent to about 14 m. The system automatically detected and summarized buildings with heights exceeding or below the permissible limit, thereby supporting the detection of floor height errors.

All functions were integrated into the interactive interface, allowing users to navigate spatial data in 2D or 3D, filter objects by criteria, edit data, and extract information for inspection, planning, and licensing. The 3D interface was built using the OLCS library, which allows for a smooth transition between 2D map mode and 3D space on the same browser platform, meeting the requirements of urban visualization in the context of digital transformation.

With the integration of UAV data, DL models, and WebGIS systems, the proposed platform is not merely a visualization tool but also a decision-support system, which is scalable and integrated into modern urban management systems.

3. Results

3.1. Development of a WebGIS Platform for Management

To assess the accuracy of the orthomosaic, five CPs were used in the evaluation according to Equations (1)–(3). The results show that the plane error of the orthogonal image reaches 4.6 cm, which is summarized in Table 1. According to Article 8 of Circular No. 26, issued by the Ministry of Natural Resources and Environment (MONRE) in 2024 [37], the positional RMSE of control points for cadastral maps must not exceed 0.1 mm on the map scale. This study aimed to create a 1 : 500 scale map, where the allowable error for control points is 5 cm. Therefore, with a horizontal error of 4.6 cm as assessed in Table 1, the orthomosaic satisfies the required accuracy standards for object extraction.

Table 1. Results of the orthographic coordinate accuracy evaluation

ID	Ground truth		Orthomosaic		ΔX [cm]	ΔY [cm]	m_{xy} [cm]
	X [m]	Y [m]	X' [m]	Y' [m]			
1	1200179.031	618920.236	1200178.992	618920.258	−3.9	3.2	5.0
2	1200124.363	618523.205	1200124.320	618523.231	−4.3	2.6	5.0
3	1199855.308	618848.259	1199855.273	618848.281	−3.5	2.2	4.1
4	1200017.768	618905.917	1200017.810	618905.897	4.2	−2.0	4.7
5	1200025.596	618686.965	1200025.634	618686.946	3.8	−1.9	4.2
RMSE					4.0	2.4	4.6

Using Equation (4) to calculate the RMSE for evaluating the CHM, the results are presented in Table 2. The results indicate that the discrepancy between the field height and the height obtained from the CHM model averages 11.9 cm. This error is relatively large due to the rugged details on the surfaces of houses, where the extraction of height from the CHM model takes the average value. The unevenness in the height of the tree canopy is also a factor contributing to errors in the process of selecting tree height measurement points in the field.

According to the provisions of Circular No. 07/2021/TT-BTNMT issued by MONRE on techniques for acquiring and processing geographic image data, the allowable height error for a 1:500 scale map with an even contour interval of 0.5 m is 12.5 cm [45]. Therefore, the error of 11.9 cm in the CHM model is still within the permissible limit. As such, the height value obtained from the CHM can be used for objects in the 3D map, which will be presented in the next section.

Table 2. Results of the accuracy evaluation of the height of the objects extracted from CHM

ID	Height [m]		ΔH [m]	RMSE [cm]
	Ground truth (H)	CHM (H')		
1	11.15	11.26	0.11	11.9
2	14.80	14.92	−0.12	
3	7.48	7.38	−0.10	
4	10.40	10.50	0.10	
5	17.50	17.59	0.09	
6	12.07	11.99	−0.08	
7	7.32	7.39	0.07	
8	10.83	10.95	0.12	
9	7.42	7.38	−0.04	
10	6.20	6.29	0.09	
11	5.98	6.18	0.20	
12	6.21	6.36	0.15	
13	5.12	5.19	0.07	
14	4.94	4.74	−0.20	
15	5.07	4.95	−0.12	

3.2. Evaluation Results of Deep Learning Models and 3D Mapping

The DL model training results are plotted in Figure 6. The value of the loss function on the training set is depicted by the blue line, which portrays the decrease in the number of loops as the model obtains more information from the dataset, and its accuracy is improved gradually. The validation loss, which is a closed-door overfitting evaluation, is shown in orange. From graph observation, we can see that both the training lines and the validation lines are monotonically decreasing, which indicates that the learning procedure of the model is performed smoothly, and the parameters are fine-tuned gradually. The modest separation between the two loss curves suggests that the model is well-balanced and not significantly overfitted.

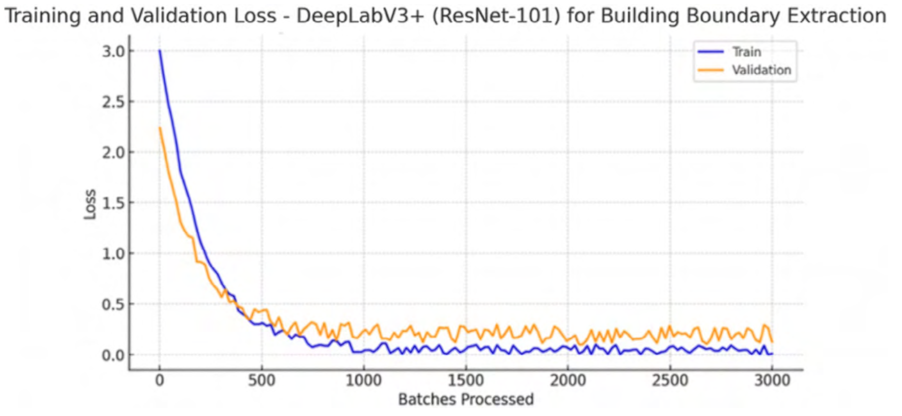

Fig. 6. Training and validation loss

In order to verify the DL model's performance, we adopted the field survey data from the practical site combined with the data of a cadastral map. The precision, recall, and Intersection over Union (IoU) were evaluated. The model had a precision of 0.90 and a recall of 0.93. The IoU was 0.91, which showed that the predicted building footprints had high spatial consistency with the reference. These evaluation results demonstrate that the proposed model is very promising for extracting building footprints from urban areas. An illustration of the predicted results with the UAV imagery is shown in Figure 7.

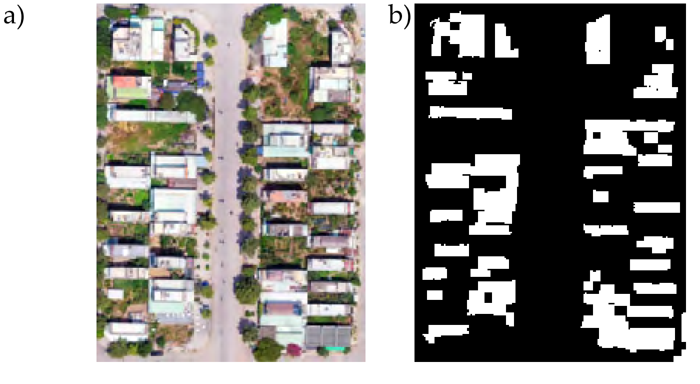

Fig. 7. Building footprint extraction results from DeepLabV3+ model:
a) UAV imagery of buildings; b) predicted building footprints

After extracting building footprints from UAV images by the DeepLabV3+ model, the results were converted into vectors and then imported into GIS to create a building boundary shapefile. Figure 8 displays the created GIS data layer where the raster of the digitized building footprints with complete geometry is clearly visible. Then, the level of detail 1 (LoD1) 3D city model was automatically derived from the DSM, CHM and the detected boundary layer by using the extrusion method (refer to Figure 9). It is the result of integrating DL and GIS technology to create 3D urban spatial data for city management.

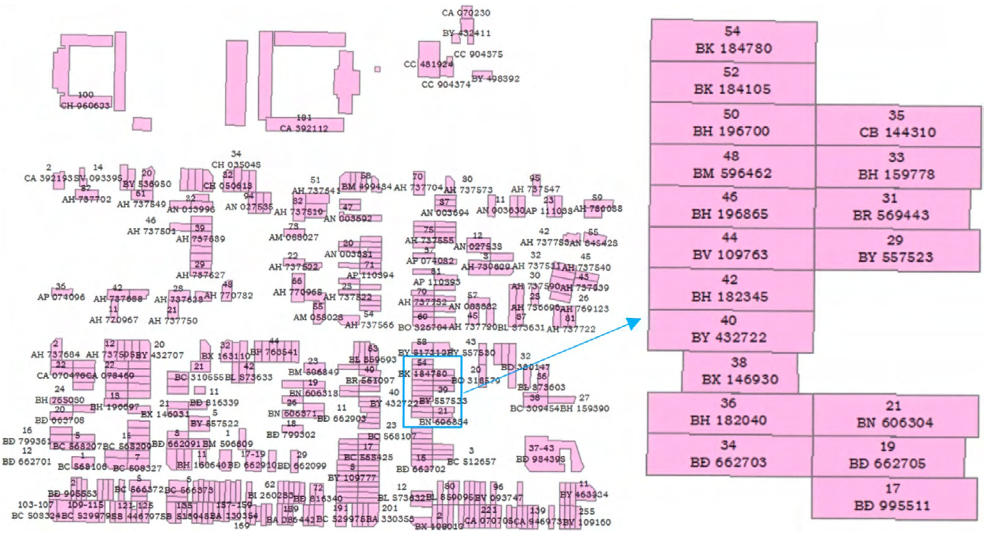

Fig. 8. Spatial data of the building layer

Fig. 9. 3D urban maps displayed in GIS environment

Using RTK (Real-Time Kinematic) technology, seven randomly selected coordinates as the corners of houses in the study area were measured and extracted from the 3D map. The accuracy of the horizontal coordinates of the 3D map was evaluated using Equations (1)–(3), with the results presented in Table 3. Because the 3D building boundaries were derived from the DL model, this positional accuracy also reflects the accuracy of the extracted building-footprint data. The evaluation data from Table 3 shows that the ground error of the 3D map is 4.8 cm, which is larger than the coordinate error of the previous orthophoto (4.6 cm). The reason is that additional errors appear in the process of automatic object recognition and extraction using the DL model, leading to the total error of the result.

Table 3. Results of the 3D map coordinate accuracy evaluation

ID	Ground truth		3D map		ΔX [cm]	ΔY [cm]	m_{xy} [cm]
	X [m]	Y [m]	X' [m]	Y' [m]			
1	1200167.267	618923.237	1200167.267	618923.211	2.4	−2.6	3.5
2	1200192.511	618832.323	1200192.511	618832.365	−3.1	4.2	5.2
3	1200180.498	618764.023	1200180.498	618764.046	4.3	2.3	4.9
4	1200189.568	618633.847	1200189.568	618633.822	−3.5	−2.5	4.3
5	1200190.808	618508.782	1200190.808	618508.825	1.2	4.3	4.5
6	1200125.401	618523.743	1200125.401	618523.777	3.2	3.4	4.7
7	1199876.929	618533.279	1199876.929	618533.229	3.6	−5.0	6.2
RMSE					3.1	3.6	4.8

In addition, a tape measure was used to measure the length of five houses, which were then compared with their lengths when extracted from the 3D map. The findings are recorded in Table 4. The results in Table 4 show that the error in the length of the houses is quite large (10.3 cm) because the error due to the curvature of the earth and the difference in terrain height was ignored in the process of measuring and processing the data.

Table 4. Result of evaluating the length accuracy of the object on the 3D map

STT	Length [m]		ΔL [m]	m_L [cm]
	Ground truth	3D map		
1	5.000	5.099	0.099	10.3
2	7.000	7.122	0.122	
3	20.000	19.898	−0.102	
4	20.010	19.917	−0.093	
5	20.020	20.118	0.098	

3.3. Results of Developing and Operating WebGIS in Urban Construction Order Management

The website interface includes a header section displaying the name of the website, contact information, and the website version. The sidebar contains the website content divided into sections: data, statistics, AI (DL), and system login. The main content area consists of two parts: a 3D display area and a 2D display area. The footer displays the author's information (Fig. 10).

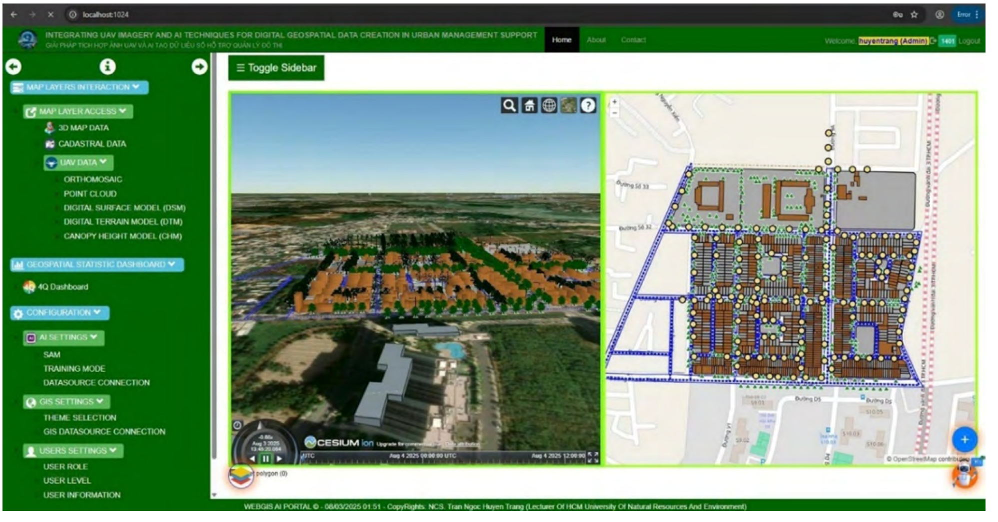

Fig. 10. WebGIS interface consists of two windows displaying 2D and 3D maps

The WebGIS for urban construction order management has been developed with the following functional groups:

1) Authorization and management functions

   The system includes two levels of user accounts:

   – Level 1. Accounts that allow interaction with data (moving, zooming in, zooming out, displaying information), searching (by keyword or spatial query), generating statistics, and exporting results into tables and reports.

   – Level 2. Administrative accounts with the highest access level, allowing function authorization, account creation and deletion, data interaction, updating, and editing.

2) Display and search functions

   Data layers are clearly organized, including cadastral data, 3D themes, and UAV imagery, supporting 2D and 3D visualization. Each object uploaded to WebGIS includes spatial and attribute parts (Fig. 11). The ability to intelligently search by keyword or spatial query optimizes the information retrieval process (Fig. 12).

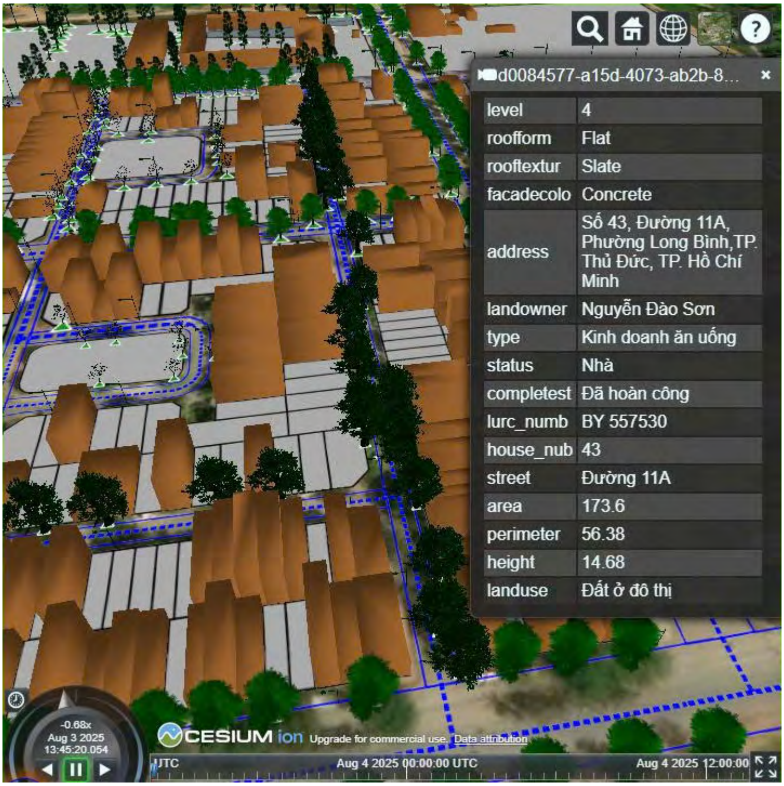

Fig. 11. Display object information

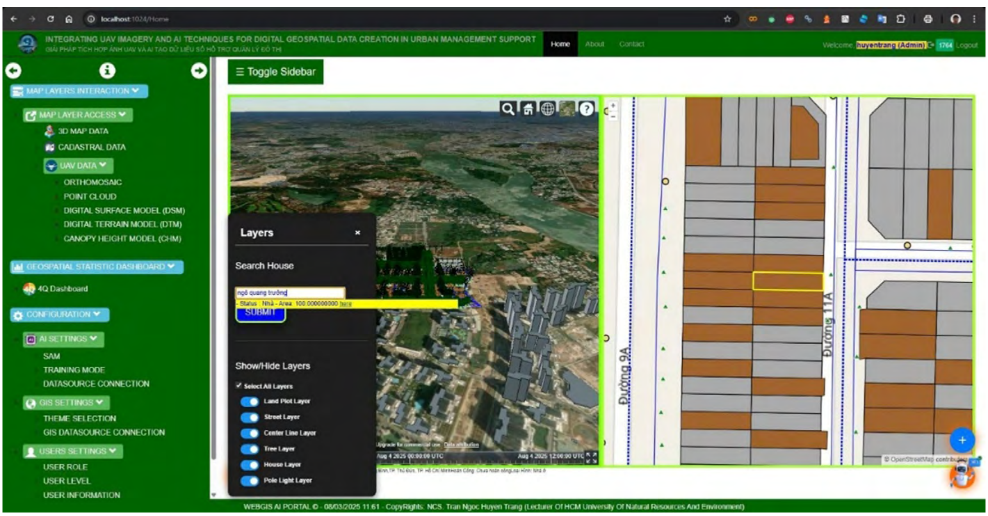

Fig. 12. Query object by attribute

3) Update and statistics functions
These functions were designed to support continuous data maintenance and updating, while providing detailed statistical tools on construction status, land use right certificates, as well as compliance with construction planning (Fig. 13). They generate statistics and export results including the number of construction works by type (residential, business, rental, etc.); the number of construction works with/without land use rights certificates and other assets on the land (pink book); and the number of construction works built in compliance or non-compliance with ground plans and height regulations.

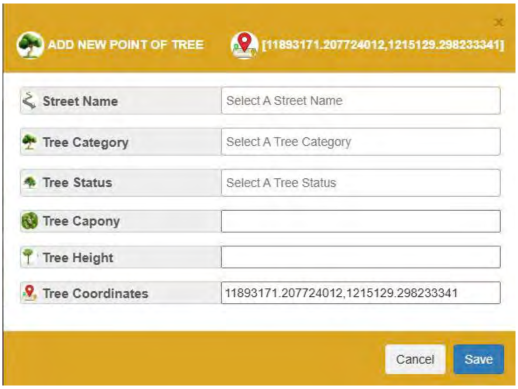

Fig. 13. Update information

4) Violation detection functions
Used to detect construction works violating ground plans and height regulations. By comparing with cadastral map data at a scale of 1:500, construction works violating the land boundary are indicated in red. Blue indicates construction works within the legal boundary, and white is the unbuilt area (Fig. 14).

a) b)

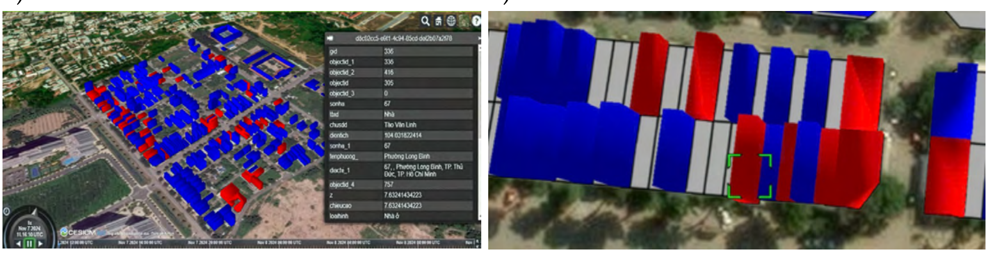

Fig. 14. Detection of construction works outside legal land boundaries:
a) red violating construction (overall); b) red violating construction (details)

The buildings are also checked for compliance with height regulations in the construction plan. Green buildings meet the height requirements and are built to the regulated height of 4 floors. In contrast, red and yellow buildings do not comply with the regulations. Red buildings exceed the allowed height (more than 4 floors), while yellow buildings fall short (less than 4 floors). Only buildings that adhere to both the site boundary and the height plan are permitted to complete construction (Fig. 15).

Fig. 15. Detect constructions violating height regulations

The WebGIS was developed to integrate data layers collected from UAVs, which are then processed to serve as input data for AI to recognize and extract layers of houses, trees, and lighting poles. It includes 2D and 3D intuitive visualization, search functions for objects based on different attribute fields, and layer overlay functions to detect violations. Statistical and update functions were also incorporated to manage the evolving database over time.

4. Discussion

The performance of the DeepLabV3+ model in this study was compared to the study by Dabove et al. [46], who also applied DL techniques to extract building footprints from remote sensing images. According to the published results, their model achieved Recall = 0.925 and IoU = 0.873, while our model achieved Recall = 0.93 and IoU = 0.91. The equivalent recall shows that the ability to correctly detect constructions in the two studies is quite close. However, the significantly higher IoU index in this study reflects the ability to identify and demarcate the building more accurately. This can be explained by two main factors. First, we used very high-resolution UAV imagery (2.82 cm/pixel), while Dabavo's study used imagery at a lower resolution (25 cm/pixel), which reduces the model's ability to learn the subtle boundary details of the structure. Secondly, the buildings in our research area have clear boundaries and stable architectural shapes, which makes it easy for the model to distinguish between backgrounds and objects, especially in contiguous regions, putting great pressure on segmentation problems.

The model results show that although the overall accuracy is high, there are still some local errors that affect the recall and precision metrics. In some locations, especially in clusters of adjacent houses with similar architecture, the DeepLabV3+ model tends to combine many houses into a single object. This phenomenon stems from the fact that the boundaries between the buildings are too narrow, the contrast is low, and the geometric features on the roofs are repetitive, making it difficult for the model to clearly separate each individual unit. This error significantly reduces the recall due to a decrease in the number of correctly identified objects (Fig. 16).

a) b) c)

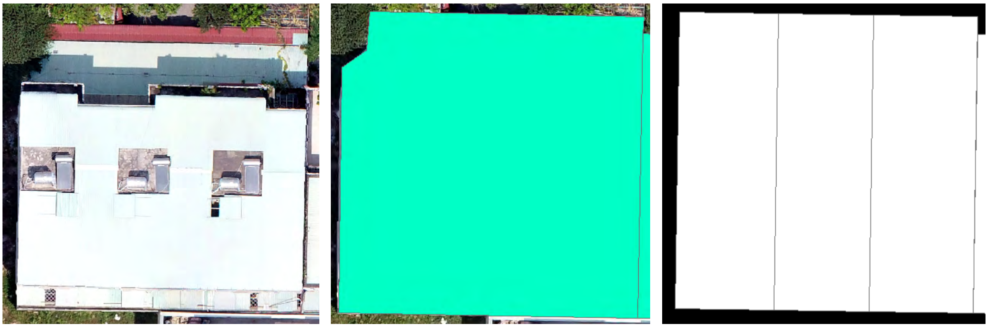

Fig. 16. Model error prediction: a) orthophoto; b) prediction from the model; c) ground truth

In addition, in the study area, there are also a number of non-residential constructions, such as substations, temporary houses for construction, and sheltered shacks for construction workers. These structures have a relatively similar shape and surface material to the roofs, resulting in the model misidentifying the object to be extracted. This increases the number of false predictions, which in turn drags down the precision.

The above errors show that the performance of the DL model is significantly influenced by architectural morphology, boundary separation, as well as the presence of interfering objects in the urban environment. Identifying and analyzing these cases helps to clarify the generalization of the model and forms the basis for improvements in the next phase of the study.

A noteworthy point in this study is the efficiency brought about by the choice of the DeepLabV3+ architecture, with the ability to capture multi-scale spatial context thanks to the ASPP module while maintaining geometric detail in the boundary regions. The application of transfer learning techniques has contributed to enhancing the ability to generalize models when working with moderate-sized field data.

Considering the challenges of both bigger data and more complex urban environments, introducing more sophisticated DL architectures, such as Vision Transformer (ViT), Swin Transformer, or attention mechanism-based models, would be a notable improvement [47–49]. These models can learn long-term spatial dependencies, and therefore can better extract objects when they are occluded, interfered with, or have fuzzy boundaries. But the procedure for application has to be formulated in an appropriate manner to suit the technical demands, the computing power, or the particular features at issue. At the same time, the combination with transfer learning techniques still plays a key role, given that labeled data is still limited, ensuring the feasibility and effectiveness of the model on actual datasets.

5. Conclusions

The present study's results showed that the urban management-based UAV and WebGIS application is highly adequate. Specifically, the orthomosaic achieved 4.6 cm precision and the CHM model achieved 11.9 cm precision, which is enough for producing 1:500 scale 3D maps with a high level of detail. Although the DL model does not reach integer precision (0.9), it significantly decreases the manpower required for the interpretation and digitization of objects. The detected building footprints with an accuracy of 4.8 cm can be considered a trustworthy basis for establishing the state management of constructions. These results demonstrate that the proposed approach is not only technically feasible but also has important implications in improving the efficiency of urban data acquisition, updating, and management.

Creating 3D maps on a WebGIS platform would provide an interactive and vivid experience with user-friendly operating tools. The ability to identify differences at ground level and in height is central to its functionality, providing urban managers with vital information and meaningful data to assist their decision-making process. Therefore, this integrated framework has clear implications in supporting construction monitoring, detecting spatial and height-related inconsistencies, and enhancing evidence-based urban management practices.

The application of advanced technology to support urban management, especially the state management of urban construction activities, has been well promoted in cities around the world. Being the most vibrant and active city in Vietnam, HCMC is leading the field of all-around digitalization; however, action is still disjointed and partial. HCMC's smart city transformation policy is a strategic move with far-reaching consequences in sustainable urban development. The development of UAV technology and WebGIS, which are integrated with smart mobile devices, plays a pivotal role in producing application-based and advanced tools for smart cities. Beyond HCMC, this approach can also be applied to other urban areas, especially cities that are experiencing rapid urbanization, requiring timely spatial data updates, construction supervision, and digital urban information management. To transfer this solution to other cities, it is necessary to ensure a suitable UAV survey plan, a consistent geospatial reference standard, a training dataset adapted to local urban characteristics, and a WebGIS database structure capable of integrating spatial and thematic information.

The development of a comprehensive web-based GIS platform (WebGIS) for a 3D map framework enables data loading, searching, real-time updating, and statistical analysis, thereby providing urban administrators with more powerful and holistic information. Leading-edge technologies, including the Internet of Things (IoT) and Artificial Intelligence (AI), play a critical role in developing the six components of smart cities. It is significant to promote advanced smart urban management with an analytical and AI-based forecasting tool supporting digital economic growth, sustainable urban development, and climate change adaptation in HCMC and other similar cities. In future work, the proposed framework can be further improved by integrating AI-based analytics functions, expanding managed urban object categories, and enhancing its applicability in broader urban governance contexts.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Author Contribution

N.H.T.T.: conceptualization, methodology, data collection and processing, software, validation, formal analysis, resources, data curation, writing – original draft preparation, writing – review and editing, visualization.

V.T.L.: conceptualization, methodology.

L.P.V.: conceptualization, writing – original draft preparation, writing – review and editing.

All authors have read and approved the published version of the manuscript.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The UAV image data used in this study were collected by the authors directly at the study area, serving the purpose of building footprints and integrating them into the WebGIS system for state management agencies. Due to data related to flight range, privacy, and local legal requirements, all raw data and derivative products are not publicly available.

However, the data may be considered for partial sharing for academic research purposes, subject to the approval of the local regulatory authority and upon appropriate request. Please contact the corresponding author for more information.

Use of Generative AI and AI-Assisted Technologies

No generative AI or AI-assisted technologies were employed in the preparation of this manuscript.

Acknowledgments

We would like to express our thanks to Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for the support of time and facilities for this study.

References

[1] National Assembly of the Socialist Republic of Vietnam: *Construction law, 50/2014/QH13*. June 18, 2014. https://vanban.chinhphu.vn/?pageid=27160&docid=175348 [access: October 10, 2025].

[2] Vasylchenko Y.M.: *State supervision of urban construction: Modern realities*. Journal of Sustainable Construction Engineering and Technology, vol. 13(3), 2022, pp. 50–58. https://doi.org/10.30880/ijscet.2022.13.03.005.

[3] Topchiy D., Tokarskiy A.: *Formation of hierarchies in the system of organization of state construction supervision in case of reorientation of urban areas*, [in:] Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R. (eds.), *Advances in Computing and Data Sciences: ICACDS 2019*, Communications in Computer and Information Science, vol. 1046, Springer, Singapore 2019, pp. 142–151. https://doi.org/10.1007/978-981-13-9942-8_13.

[4] Tokarsky A., Topchiy D., Ekba S., Bolgov V., Gorochov A.: *Methodological foundations of assessment of the impact of State construction supervision during reprofiling big urban areas*. E3S Web of Conferences, vol. 164, 2020, 04037. https://doi.org/10.1051/e3sconf/202016404037.

[5] Hajer A.S., Ishaqee A., Albazaz A.E.: *The impacts of building contraventions and informal residential land subdivision on the quality of life in Baghdad*. IOP Conference Series: Materials Science and Engineering, vol. 518(2), 2019, 022072. https://doi.org/10.1088/1757-899X/518/2/022072.

[6] Kumar R.: *WebGIS map for smart city development and decision support system: A case study of Dehradun Smart City, Uttarakhand, India*. Information System and Smart City, vol. 4(1), 2024, 1454. https://doi.org/10.59400/issc1454.

[7] Důbravová H., Bureš Lukas V., Velfl V.: *Review of the application of drones for smart cities*. IET Smart Cities, vol. 6(4), 2024, pp. 312–332. https://doi.org/10.1049/smc2.12093.

[8] White G., Zink A., Codecá L., Clarke S.: *A digital twin smart city for citizen feedback*. Computers, Environment and Urban Systems, vol. 110, 2021, 103064. https://doi.org/10.1016/j.cities.2020.103064.

[9] Nex F., Remondino F.: *UAV for 3D mapping applications: A review*. Applied Geomatics, vol. 6(1), 2014, pp. 1–27. https://doi.org/10.1007/s12518-013-0120-x.

[10] Nex F., Armenakis C., Cramer M., Cucci D.A., Gerke M., Honkavaara E., Kukko A., Persello C., Skaloud J.: *UAV in the advent of the twenties: Where we stand and what is next*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 184, 2022, pp. 215–242. https://doi.org/10.1016/j.isprsjprs.2021.12.006.

[11] Mohsan S.A.H., Othman N.Q.H., Li Y., Alsharif M.H., Khan M.A.: *Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends*. Intelligent Service Robotics, vol. 16(1), 2023, pp. 109–137. https://doi.org/10.1007/s11370-022-00452-4.

[12] Quamar M.M., Al-Ramadan B., Khan K., Shafiullah M., Ferik S.E.: *Advancements and applications of drone-integrated geographic information system technology – A review*. Remote Sensing, vol. 15(20), 2023, 5039. https://doi.org/10.3390/rs15205039.

[13] Billen R., Cutting-Decelle A.F., Marina O., Almeida J.P., Caglioni M., Falquet G., Leduc T., Métral C., Moreau G., Perret J., Rabino G., Jose R.S., Yatskiv I., Zlatanova S.: *3D City Models and Urban Information: Current Issues and Perspectives*. EDP Sciences, Les Ulis Cedex A, France 2014.

[14] Lafioune N., St-Jacques M.: *Towards the creation of a searchable 3D smart city model*. Innovation & Management Review, vol. 17(3), 2020, pp. 285–305. https://doi.org/10.1108/INMR-03-2019-0033.

[15] Biljecki F., Stoter J., Ledoux H., Zlatanova S., Çöltekin A.: *Applications of 3D city models: State of the art review*. ISPRS International Journal of Geo-Information, vol. 4(4), 2015, pp. 2842–2889. https://doi.org/10.3390/IJGI4042842.

[16] Gu L., Zhang H., Wu X.: *Surveying and mapping of large-scale 3D digital topographic map based on oblique photography technology*. Journal of Radiation Research and Applied Sciences, vol. 17(1), 2024, 100772. https://doi.org/10.1016/j.jrras.2023.100772.

[17] Kamarulzaman A.M.M., Jaafar W.S.W.M., Said M.N.M., Saad S.N.M., Mohan M.: *UAV implementations in urban planning and related sectors of rapidly developing nations: A review and future perspectives for Malaysia*. Remote Sensing, vol. 15(11), 2023, 2845. https://doi.org/10.3390/rs15112845.

[18] Tan C., Chen T., Liu J., Deng X., Wang H., Ma J.: *Building extraction from Unmanned Aerial Vehicle (UAV) data in a landslide-affected scattered mountainous area based on Res-Unet*. Sustainability, vol. 16(22), 2024, 9791. https://doi.org/10.3390/su16229791.

[19] Tang G., Ni J., Zhao Y., Gu Y., Cao W.: *A survey of object detection for UAVs based on deep learning*. Remote Sensing, vol. 16(1), 2024, 149. https://doi.org/10.3390/rs16010149.

[20] Ayala C., Sesma R., Aranda C., Galar M.: *A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery*. Remote Sensing, vol. 13(16), 2021, 3135. https://doi.org/10.3390/rs13163135.

[21] Luo L., Li P., Yan X.: *Deep learning-based building extraction from remote sensing images: A comprehensive review*. Energies, vol. 14(23), 2021, 7982. https://doi.org/10.3390/rs13163135.

[22] Oladipo I.D., AbdulRaheem M., Awotunde J.B., Bhoi A.K., Adeniyi E.A., Abiodun M.K.: *Machine learning and deep learning algorithms for smart cities: A start-of-the-art review*, [in:] Nath Sur S., Balas V.E., Bhoi A.K., Nayyar A. (eds.), *IoT and IoE Driven Smart Cities*, EAI/Springer Innovations in Communication and Computing, Springer, Cham 2022, pp. 123–145. https://doi.org/10.1007/978-3-030-82715-1_7.

[23] Heidari A., Navimipour N.J., Unal M.: *Applications of ML/DL in the management of smart cities and societies based on new trends in information technologies: A systematic literature review*. Sustainable Cities and Society, vol. 85, 2022, 104089. https://doi.org/10.1016/j.scs.2022.104089.

[24] Ginardi R.V.H., Gunawan W., Wardana S.R.: *WebGIS for asset management of land and building of Madiun City government*. Procedia Computer Science, vol. 124, 2017, pp. 437–443. https://doi.org/10.1016/j.procs.2017.12.175.

[25] Randazzo G., Italiano F., Micallef A., Tomasello A., Cassetti F. P., Zammit A., D'Amico S., Saliba O., Cascio M., Cavallaro F., Crupi A., Fontana M., Gregorio F., Lanza S., Colica E., Muzirafuti A.: *WebGIS implementation for dynamic mapping and visualization of coastal geospatial data: A case study of BESS project*. Applied Sciences, vol. 11(17), 2021, 8233. https://doi.org/10.3390/app11178233.

[26] Katzschner A., Waibel M., Schwede D., Katzschner L., Schmidt M., Storch H. (eds.): *Sustainable Ho Chi Minh City: Climate Policies for Emerging Mega Cities*. Springer, Cham 2016. https://doi.org/10.1007/978-3-319-04615-0.

[27] Gibert-Flutre M., Musil C.: *Ho Chi Minh City, a metropolitan playing field*. EchoGéo, vol. 52, 2020, 22079. https://doi.org/10.4000/echogeo.22079.

[28] Le V.T., Vo L.P., Tran N.H.T., Ha Q.K.: *Opportunities and challenges of UAV application for monitoring the construction progress and updating the geographic database in urban area of Ho Chi Minh City, Vietnam*. IOP Conference Series: Earth and Environmental Science, vol. 1170, 2023, 012014. https://doi.org/10.1088/1755-1315/1170/1/012014.

[29] Downes N.K., Storch H., Viet P.Q., Diem N.K., Dinh L.C.: *Assessing peri-urbanisation and urban transitions between 2010 and 2020 in Ho Chi Minh City using an urban structure type approach*. Urban Science, vol. 8(1), 2024, 11. https://doi.org/10.3390/urbansci8010011.

[30] Nguyen M.T., Le T.H.: *Science – technology development: A case of Ho Chi Minh City, Vietnam*. Journal of Law and Sustainable Development, vol. 11(1), 2023, e407. https://doi.org/10.37497/sdgs.v11i1.407.

[31] Enoc S.A., Chandler J.H., José R.R.P., Ordóñez C.: *Accuracy of Unmanned Aerial Vehicle (UAV) and SfM photogrammetry survey as a function of the number and location of ground control points used*. Remote Sensing, vol. 10(10), 2018, 1606. https://doi.org/10.3390/rs10101606.

[32] Fissore F., Pirotti F.: *DSM and DTM for extracting 3D building models: Advantages and limitations*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2/W13, 2019, pp. 1539–1546. https://doi.org/10.5194/isprs-archives-XLII-2-W13-1539-2019.

[33] Zhang Z., Guo W., Li M., Yu W.: *GIS-supervised building extraction with label noise-adaptive fully convolutional network*. IEEE Geoscience and Remote Sensing Letters, vol. 17(12), 2020, pp. 2135–2139. https://doi.org/10.1109/LGRS.2019.2963065.

[34] Braik A.M., Koliou M.: *Automated building damage assessment and large-scale mapping by integrating satellite imagery, GIS, and deep learning*. Computer-Aided Civil and Infrastructure Engineering, vol. 39(15), 2024, pp. 2389–2404. https://doi.org/10.1111/mice.13197.

[35] Famiglietti N.A., Cecere G., Grasso C., Memmolo A., Vicari A.: *A test on the potential of a low cost unmanned aerial vehicle RTK/PPK solution for precision positioning*. Sensors, vol. 21(11), 2021, 3882. https://doi.org/10.3390/s21113882.

[36] Cho J.M., Lee B.K.: *GCP and PPK utilization plan to deal with RTK signal interruption in RTK-UAV photogrammetry*. Drones, vol. 7(4), 2023, 265. https://doi.org/10.3390/drones7040265.

[37] Ministry of Natural Resources and Environment (Vietnam): *Circular No. 26/2024/TT-BTNMT on Technical regulations on cadastral surveying and mapping. Appendix 01*. November 26, 2024. https://chinhphu.vn/?pageid=27160&docid=212227&classid=1&typegroupid=6 [access: October 15, 2025].

[38] Federal Geographic Data Committee: *Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy*. FGDC-STD-007.3-1998, FGDC, Reston, Virginia 1998. https://www.fgdc.gov/standards/projects/accuracy/part3/index_html [access: October 15, 2025].

[39] Bipul N., Aryal J., Rajabifard A.: *Fine-tuning-based transfer learning for building extraction from off-nadir remote sensing images*. Remote Sensing, vol. 17(7), 2025, 1251. https://doi.org/10.3390/rs17071251.

[40] Li Z., Dong J.: *A framework integrating DeepLabV3+, transfer learning, active learning, and incremental learning for mapping building footprints*. Remote Sensing, vol. 14(19), 2022, 4738. https://doi.org/10.3390/rs14194738.

[41] Vrbančič G., Podgorelec V.: *Transfer learning with adaptive fine-tuning*. IEEE Access, vol. 8, 2020, pp. 196197–196211. https://doi.org/10.1109/ACCESS.2020.3034343.

[42] Gupta N.: *A pre-trained vs fine-tuning methodology in transfer learning*. Journal of Physics: Conference Series, vol. 1947(1), 2021, 012028. https://doi.org/10.1088/1742-6596/1947/1/012028.

[43] Atik S.O., Atik M.E., Ipbuker C.: *Comparative research on different backbone architectures of DeepLabV3+ for building segmentation*. Journal of Applied Remote Sensing, vol. 16(2), 2022, 024510. https://doi.org/10.1117/1.JRS.16.024510.

[44] Cao Q., Li M., Yang G., Tao Q., Luo Y., Wang R., Chen P.: *Urban vegetation classification for unmanned aerial vehicle remote sensing combining feature engineering and improved DeepLabV3+*. Forests, vol. 15(2), 2024, 382. https://doi.org/10.3390/f15020382.

[45] Ministry of Natural Resources and Environment (Vietnam): *Circular No. 07/2021/TT-BTNMT on Technical regulations for the acquisition and processing of digital image data from Unmanned Aerial Vehicles (UAVs) for the development and updating of the national geospatial database at scales of 1:2,000 and 1:5,000, and for the creation of topographic maps at scales of 1:500 and 1:1,000. Rule number 9*. June 30, 2021. https://tulieuvankien.dangcongsan.vn/he-thong-van-ban/van-ban-quy-pham-phap-luat/thong-tu-so-072021tt-btnmt-ngay-30062021-cua-bo-tai-nguyen-va-moi-truong-quy-dinh-ky-thuat-thu-nhan-va-xu-ly-du-lieu-anh-7742 [access: October 15, 2025].

[46] Dabove P., Daud M., Olivotto L.: *Revolutionizing urban mapping: Deep learning and data fusion strategies for accurate building footprint segmentation*. Scientific Reports, vol. 14, 2024, 13510. https://doi.org/10.1038/s41598-024-64231-0.

[47] Xu R., Xiang H., Tu Z., Xia X., Yang M.-H., Ma J.: *V2X-ViT: Vehicle-to-everything cooperative perception with Vision Transformer*, [in:] Avidan S., Brostow G., Cissé M., Farinella G.M., Hassner T. (eds.), *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, Lecture Notes in Computer Science, vol. 13699, Springer, Cham 2022, pp. 107–125. https://doi.org/10.1007/978-3-031-19842-7_7.

[48] Pinasthika K., Laksono B.S.P., Irsal R.B.P., Shabiyya S.H., Yudistira N.: *SparseSwin: Swin transformer with sparse transformer block*. Neurocomputing, vol. 580, 2024, 127433. https://doi.org/10.1016/j.neucom.2024.127433.

[49] Niu Z., Zhong G., Yu H.: *A review on the attention mechanism of deep learning*. Neurocomputing, vol. 452, 2021, pp. 48–62. https://doi.org/10.1016/j.neucom.2021.03.091.