Integrating Advanced AI techniques to assist Urban Digital Twins Generation

Jiaojiao Tian*, Mareike Weishaupt, Veronika Gstaiger, Guneet Mutreja, Chaikal Amrullah, Corentin Henry, Thomas Krauß, Pablo d'Angelo, Sandeep K. Jangir, Stefan Auer, Franz Kurz, Ksenia Bittner

Photogrammetry and image analysis department,
Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany

Keywords: Digital Twin, Trees, LoD1, LoD2, Semantic Segmentation, Vectorization

Abstract

Digital twins play a crucial role in autonomous driving applications and transportation system simulations. The need for large scale and dynamic information has increased interest in generating urban digital twins from remote sensing data. Aerial high resolution imagery of urban areas serves as the one of the most important data sources for this task. Advances in deep learning and machine learning allow more accurate and automated extraction of urban elements. In recent years, we have developed and integrated advanced deep learning models to extract various land cover types surrounding road networks, including buildings, roads, and vegetation. Furthermore, we have conducted proof of concept studies aimed at detecting and delineating linear landmarks from aerial imagery, including curbstones and road borders. These developments contribute to the creation of more accurate and detailed urban digital twins, which are essential for advanced urban analytics and intelligent transportation systems.
Results from the deep learning models are presented for the Schwarzer Berg district in Brunswick, Germany, which is a test region for the development of mobility services and technologies at the German Aerospace Center (DLR). The AI models are trained using benchmark datasets from other urban regions, indicating that the proposed approaches can be readily transferred and evaluated in other European cities.

1. Introduction

Rapid urbanization and increasing climate uncertainty present significant challenges for modern cities. In response, smart city management strategies are now more critical than ever. (Schrotter and Hürzeler, 2020; Hlal et al., 2025). To support the smart city management, an urban digital twin is a powerful tool for simulation and visual replication of the physical urban environment. In the urban context, a city digital twin represents the built infrastructures in 2D and 3D (Batty, 2024) mostly enriched by additional semantic information. By integrating multi-source geospatial data together with advanced simulation and prediction models, urban digital twins are valuable for many applications, including risk management, decision support, infrastructure maintenance and traffic flow optimization (Sharifi et al., 2024; Faliagka et al., 2024; Hlal et al., 2025).

The Schwarzer Berg district of Brunswick, Germany, serves as one of the German Aerospace Center (DLR)'s prototype test regions for future mobility services and technologies. This region is currently used in on-going research and development projects, including the new mobility service for the automated transport of people and goods (DLR-KI, n.d.). A detailed and dynamic up-to-date two-dimensional (2D) and three-dimensional (3D) city models as a fundamental starting point (Batty, 2024) of Digital twin, and also serves as the fundamental component for reliable operation of autonomous mobility services. However, until now, many digital twin systems are fully or partly generated manually, which limited their applications in regional and national wide (Sun et al., 2019; Somanath et al., 2023; Badreddine et al., 2026).

Recent advancements in remote sensing technologies and Artificial Intelligent (AI) have further strengthened the potential of urban digital twins as the increasing spatial and temporal resolution of spaceborne and airborne imagery allows for more detailed and frequent urban mapping. As regards the selection of data sources, airborne survey systems have become established in a wide range of research areas within urban planning and development, as they offer a number of advantages, including a high degree of flexibility in data collection. The flexibility of these systems stems from their ability to be scheduled, re-configured, and repositioned at will, providing high-resolution, low-latency, and sub-cloud observations on a timeline and with a spatial detail that far surpasses the fixed and less adaptable nature of satellite data. The sensors are mounted on aircraft or helicopters, each of which has advantages and disadvantages depending on the area of application. With recent advances in technology, both optical and radar sensors can be easily mounted on airborne platforms and utilized for digital twin generation and related applications (Javan et al., 2025; Xu et al., 2025; Gstaiger et al., 2025). However, RGB cameras combined with photogrammetric techniques for 3D reconstruction remain the most widely used approach for city modeling.

Concurrently, the rapid development of artificial intelligence (AI)-based image processing and deep learning techniques has greatly enhanced the efficiency and accuracy of urban mapping (Xu et al., 2024; Islam et al., 2025). In the last years, state-of-the art models, from convolutional neural network (CNN), vision transformers and large foundation models have been applied to a wide range of remote sensing applications (Zhu et al., 2017; Ma et al., 2019; Xie et al., 2020; Osco et al., 2021; Li et al., 2025). The advanced information retrieval and classification techniques allow the delivering of high precision results and the processing large-scale and even global datasets. However, most of these implementations in Remote Sensing remain at research and experimental stage. A significant gap still exists to bring the AI techniques to operational mapping workflows. Further efforts are needed to fully unlock the potential of AI

* Corresponding authors: Jiaojiao.tian@dlr.de

models for the automatic creation of urban digital twins.

The remainder of the paper is organized as follows. Image products and semantic layers are presented in Section II and III. Section IV describes the mapping products which have the potential to be directly integrated to a digital twin system. Finally, Section V provides a discussion and conclusions of the work.

2. Image product

2.1 Aerial image acquisition

Optical aerial image data from the 3K camera system of the German Aerospace Center (DLR) (Rosenbaum et al., 2008) is used for our study. The system is developed in-house for acquiring aerial images, processing them directly on board the aircraft and optionally transmitting the data to a ground station. The system has been in operation for many years and has provided valuable data across a wide range of research areas and measurement campaigns (Gstaiger et al., 2025).

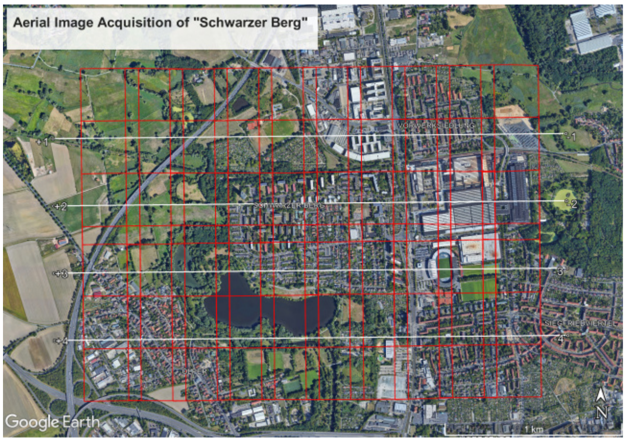

Figure 1. Flight planning for image acquisition at Braunschweig, Schwarzer Berg (white: flight lines; red: coverage of images)

On 2024-04-30, the aerial image data of the test area ”Schwarzer Berg” in Brunswick, Germany, was collected using the 3K camera system on board the DLR research aircraft Cessna 208 B Grand Caravan (D-FDLR). Figure 1 shows the flight strips (white) with the corresponding coverage of the individual images (red). The flight strips were planned in such a way that a transversal overlap of 70 % provided an optimal basis for the generation of a dense digital surface model. Due to the planned flight altitude of $700\,\mathrm{m}$ above ground, images were taken with a resulting ground sampling distance of $8.7\,\mathrm{cm}$ per pixel.

2.2 DSM and orthophoto generation

After satellite positioning service (SAPOS) correction of the flight global positioning system (GPS) data and bundle-block-adjustment of all aerial images a digital surface model (DSM) is derived using the semi-global-matching as shown in d’Angelo and Reinartz (2011) (see Figure. 2).

All aerial images are orthorectified using the calculated DSM. To generate a true-ortho-mosaic the hidden areas in each of the orthorectified images are filled with no-data-values. Afterwards all of these true-ortho-images are stacked together and the median-value at each position is selected. This allows for filling areas hidden in one image by the content of another image and furthermore this method removes also moving objects like cars or pedestrians from the true-ortho-mosaic since the number of images without an object is usually larger than the number containing the moving object.

2.3 DTM/nDSM generation

A digital terrain model (DTM) (Figure 2) represents the Earth’s bare surface, excluding vegetation and man-made structures like buildings, roads, and bridges. It provides essential data for fields such as surveying, construction, disaster management, hydrology, and land cover mapping. Therefore, creating a highly accurate and detailed DTM is crucial for modern technological applications. In this study, we develop a deep learning model based on the EfficientNet architecture integrated into a U-Net framework (Bittner et al., 2023). Despite its relatively low parameter count compared to conventional networks, the model effectively distinguishes non-ground pixels and estimates bare-earth elevations while preserving the geomorphological complexity of anthropogenic landscapes.

The availability of both the DSM and DTM enables the derivation of a normalized digital surface model (nDSM) (Figure 3) by substracting the DTM from the DSM. This process effectively isolates the elevation of above-ground features, such as vegetation, buildings, and other structures, by removing the influence of the underlying terrain. The resulting nDSM provides valuable information for numerous applications, including urban morphology analysis, vegetation height estimation, infrastructure monitoring, and 3D landscape modeling.

3. Semantic layers

3.1 Semantic Segmentation from 2D/3D multimodal pre-trained foundation models

Accurate and detailed Land Use/Land Cover (LULC) maps are a critical layer for digital twins, providing essential semantic context. While foundation models are powerful, many are pre-trained only on 2D (RGB) imagery, overlooking the crucial 3D elevation information from DSMs. To address this, we employ HiRes-FusedMIM (Mutreja et al., 2025), a multimodal foundation model pre-trained on a large-scale dataset (over 368k images) of paired high-resolution (0.2-$0.5\,\mathrm{m}$) RGB and DSM data. Its dual-encoder architecture learns powerful, aligned representations from both modalities, proving highly effective on diverse semantic segmentation benchmarks like LoveDA and Vaihingen. For this work, the pre-trained model is fine-tuned on the high-resolution ISPRS Potsdam dataset to recognize key urban LULC classes. This fine-tuned model is then applied to the Brunswick test area to generate a detailed semantic segmentation map, as shown in Figure 4.

3.2 Traffic Areas

The method for segmenting roads, access ways and parking areas is based on a neural network called SkipFuse-Dense-U-Net-121, which has already proven its performance in several studies (Henry et al., 2021; Merkle et al., 2022; Kurz et al., 2025). This network is characterised by a high level of detail in the segmentation results, as well as excellent generalisation capabilities. Especially when training with image material from different regions. These properties make it particularly suitable for combination with the DLR’s internal Traffic Infrastructure

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-607-2026

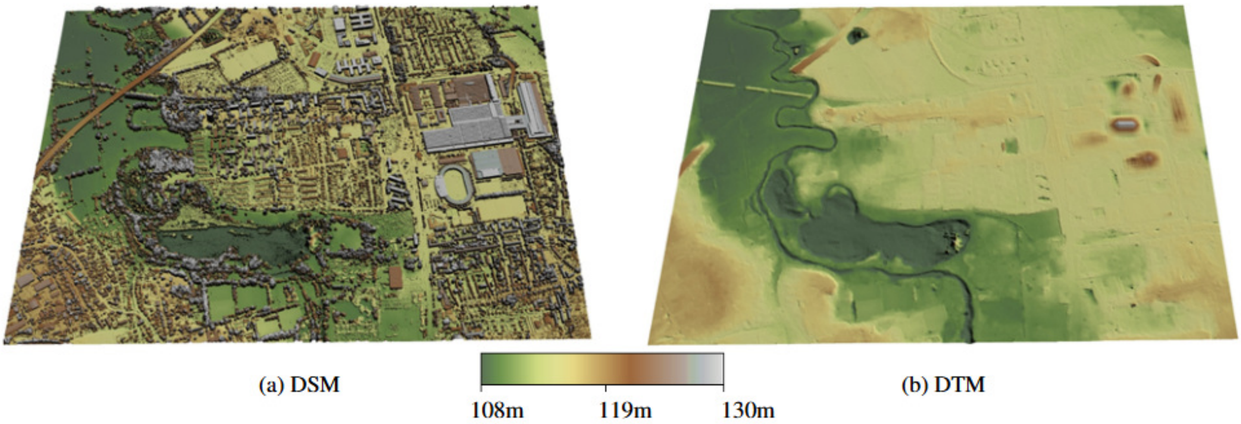

Figure 2. Example of generated DSM and DTM in Schwarzer Berg.

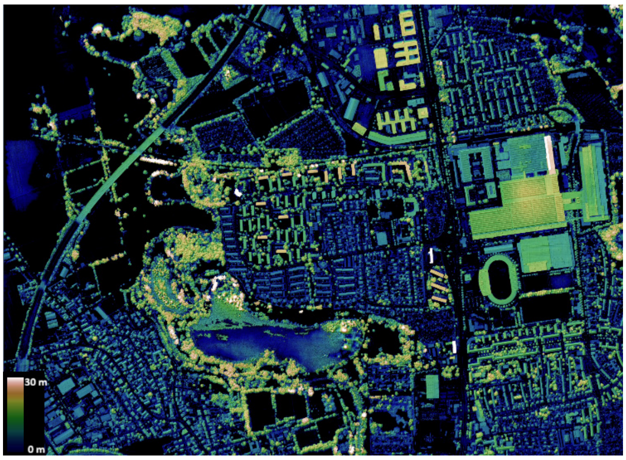

Figure 3. Example of automatically generated nDSM in Schwarzer Berg.

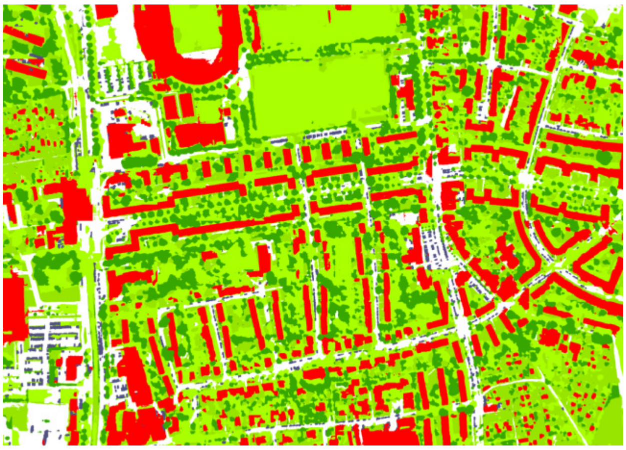

Figure 4. Example of semantic segmentation results in Schwarzer Berg. Class colors are: impervious surfaces, buildings, low vegetation, trees, cars.

And Surroundings (TIAS) dataset, which contains manual annotations for various traffic area classes on aerial images and shows scenes from over 20 German cities. The model also uses information from OpenStreetMap in the form of rasterised masks of traffic-related objects (roads, access ways, parking areas, etc.) and surrounding object types such as footpaths, cycle paths, buildings or water surfaces in order to better filter out errors automatically. On the test set of the TIAS dataset, it achieves 80% IoU on roads, 59% IoU on access ways and 63% IoU on parking areas, which shows its good to excellent generalization capability to our region of interest. An example of the segmentation results is shown in Figure 5.

Figure 5. Example of traffic areas segmentation results in Schwarzer Berg. Class colors are: parking area, road, access way.

4. Mapping Products

4.1 Pixel-based Roof Instance Segmentation with 3D Reconstruction

In the context of 3D city and infrastructure modeling, level of detail (LoD)1 represents buildings as simplified volumetric blocks generated by extruding their footprints to the corresponding height, typically assuming flat roofs. This level provides a computationally efficient yet meaningful abstraction of the built environment, supporting large-scale analyses such as energy estimation, urban planning, and environmental simulations. We perform LoD1 reconstruction following the approach of Schuegraf et al. (2023), which employs a SkipFuse-DenseNet121-U-Net architecture for building and separation line segmentation. The network output is post-processed using the watershed transform to obtain individual building instances. Each instance is polygonized and assigned the mean DSM height within its footprint, resulting in block-shaped LoD1 (see Figure 6) models suitable for large-scale digital twin reconstruction.

Figure 6. Automatic generated LoD1 and LoD2 building models.

While LoD1 models depict buildings as uniform extrusions of their footprints with flat roofs, LoD2 introduces geometric complexity that more accurately reflects real-world architectural forms. This additional detail significantly improves both the geometric and semantic realism of the models, allowing for more precise analyses of urban environments. In (Schuegraf et al., 2024) we developed a method, which employs a depth attention-based fully convolutional network (FCN) that jointly segments building sections and roof planes from RGB and DSM inputs. The network's depth attention module effectively integrates height information, enabling accurate roof plane extraction. Subsequently, the roof planes are converted into polygons, and their heights are estimated using RANSAC (Fischler and Bolles, 1981) to generate precise LoD2 models (see Figure 6). This approach ensures topologically consistent and geometrically accurate reconstructions suitable for advanced urban analyses.

4.2 Graph-based Roof Reconstruction

Although our pixel-based roof and building instance segmentation from Section 4.1 method produces satisfactory results, the resulting polygon shapes often remain irregular, which limits the overall geometric realism of the reconstructed models. To address this issue, we are currently investigating graph-based algorithms for roof polygon reconstruction that have the potential to produce sharp representations of roof structures.

We integrate the recently proposed PolyRoof model (Amrullah et al., 2025) with an R2AU-Net backbone and an edge-aware graph neural network (GNN) for roof polygon reconstruction (see Figure 7). The CNN-based backbone extracts latent features from very-high resolution (VHR) imagery for point detection and feature representation, while the edge-aware GNN refines vertex connectivity to produce geometrically consistent roof outlines.

To further improve reconstruction accuracy, we augment training with generative synthetic data that introduces additional roof texture variability while preserving precise geometric supervision (Amrullah and Bittner, 2026). Specifically, synthetic RGB imagery is generated from accurately defined roof labels, ensuring consistent and noise-free label–image pairs that allow the model to learn richer roof appearance patterns without compromising annotation accuracy.

Figure 7. Graph-based 2D roof planes vectorization.

4.3 Landmark line object

The object recognition and the acquisition of road space information for the landmark maps was performed using the latest Visual Transformer models (Liu et al., 2021), suitable road space elements were specifically selected (landmark classes, e.g. road edges, kerbs) and training data was professionally created to learn the appearance of the classes in the remote sensing data with AI. Four line-shaped categories are selected to enrich the urban digital twin, including curbstone, fence, Guardrail and Level edge between paved and unpaved area

We adopt the two-step vectorization approach described by Tian et al. (2023) for curbstone and other landmark lines extraction. The workflow combines a Vision Transformer for initial feature extraction with a subsequent graph-based vectorization step. For the training of the AI models, 40 TrueOrtho image (TrueDOP-20) images with a spatial resolution of 20 cm were professionally annotated on an external order. Images from the areas of Munich and Lindau, Bavaria, Germany, were acquired from the State Office for Broadband, Digitization and Surveying and selected for annotation. Figure. 8 shows details of curbstone detection result.

4.4 Poles

Accurate detection of poles such as traffic signs and traffic lights is important for applications such as autonomous driving and urban planning. Thus it also serves as one of the essential elements for city digital twin. However detecting poles from aerial imagery is rather a challenging task, as the footprints of pole objects are reduced to only 1-2 pixels. Zhuo and Tian (2025) proposed the you only look once (YOLO)-Pole deep learning framework for pole localization. It is an end-to-end deep learning model based on YOLOv7 to predict pole location from 20 cm aerial directly. A specialized point-wise loss function based on Euclidean distance is implemented to enhance the accuracy of localization. We apply the trained model proposed

Figure 8. Extracted curbstone lines.

in paper (Zhuo and Tian, 2025) to 3K orthophoto data over Schwarzer Berg. Part of the result is shown in Figure 9, where successfully detected poles are highlighted with yellow dots.

Figure 9. Detected poles, highlighted with yellow dots.

4.5 City ITC vectorization

Trees play a crucial role in urban environments due to their widespread distribution across cities. In spatial analyses, tree canopies can act as confounding factors by obscuring urban features such as roads and rooftops, while also serving as indicators of ecological conditions. Tree is a class that often exhibits great intra-class variance. Even the same tree can have a different appearance at different phenological stages. A previous study on large-scale urban tree recognition has emphasized the importance of height information, without which trees can easily be misclassified as low vegetation and other classes (Yuan et al., 2023).

To better distinguish tree crowns from other land cover types, we first apply super-resolution to enhance image spatial resolution from 20 to 5 cm. The U2D2 framework is a two-stage method for blind super-resolution and enhancement of aerial and satellite images (Jangir et al., 2026). It first utilizes a lightweight 2× super-resolution network trained on simulated degradations, and subsequently feeds both this output and a bicubic interpolated image into a diffusion model for refinement, resulting in sharp, artifact-free images. This helps to increase the performance of the pretrained Mask RCNN model (Tian et al., 2025). This model was originally trained on dense natural forest data with a very high resolution with a spatial resolution of 2 cm. Due to the domain shift between the training and testing data, few false detections occur. Therefore, in the post-processing step we refine the tree segmentation with the height information provided by nDSM. We also use our building segmentation to reduce misclassifications of trees and buildings. Figure 10 shows the tree instance segmentation that identifies individual trees over a part of our test region. These trees are selected with the nDSM by using a minimal height for city trees above 2 m, which is commonly used in related work (Bolte et al., 2024; Hou and Walz, 2014; Yang et al., 2019) and excludes shrubs and bushes.

Figure 10. Tree instance segmentation results.

5. Discussion and conclusion

An urban digital twin serves as a fundamental platform for various of research and management related applications, including urbanization studies, transportation planning and environmental related analysis (Xu et al., 2024; Wu et al., 2024; Islam et al., 2025). The generation of urban digital twins represents a complex task that cannot be addressed by a single method alone, but rather requires the integration of multiple complementary approaches. This paper presents and organizes a collection of methods that together enable the generation of urban digital twins by combining remote sensing techniques for data acquisition and deep learning for image interpretation.

The urban digital twin products are presented in three categories: image products, semantic layers, and mapping products. The image products primarily support visualization and provide source data for a wide range of analytical and operational applications. The semantic layers represent major urban elements. More importantly, the mapping products deliver vector layers that can be readily integrated into standard GIS software: these developments pave the road to fully automatic urban feature extraction and the creation of precise and up-to-date urban digital twins. This paper exemplifies the effective combination of deep learning approaches / products, training data, and variable optical Remote Sensing sources for the modeling of a district in Brunswick, Germany, with the requirement of being transferable to other urban areas.

Among the available remote sensing data sources, aerial imagery is selected in this study partly because of their advantages in generating accurate 3D models of buildings and the extraction of small-sized street objects. Nevertheless, with the continuous improvement in the spatial and spectral resolution of satellite imagery, we intend to explore the potential of satellite data for digital twin modeling in the future. Additionally, we will integrate further applications based on the established digital twin framework.

References

Amrullah, C., Bittner, K., 2026. Graph-based roof reconstruction with synthetic data supervision from misaligned labels. M. Keuper, F. Locatello (eds), *Pattern Recognition*, Springer Nature Switzerland, Cham, 535–549.

Amrullah, C., Panangian, D., Bittner, K., 2025. Polyroof: Precision roof polygonization in urban residential building with graph neural networks. *2025 Joint Urban Remote Sensing Event (JURSE)*, CFP25RSD-ART, 1–4.

Badreddine, O., Radoine, H., Hajji, R., 2026. TwinCity: An Urban Digital Twin Framework for Data-Scarce Environments—A Case Study of Benguerir, Morocco. *Smart Cities*, 9(2), 23.

Batty, M., 2024. Digital twins in city planning. *Nature Computational Science*, 4(3), 192–199.

Bittner, K., Zorzi, S., Krauß, T., d’Angelo, P., 2023. DSM2DTM: An End-to-End Deep Learning Approach for Digital Terrain Model Generation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 925–933.

Bolte, A.-M., Niedermann, B., Kistemann, T., Haunert, J.-H., Dehbi, Y., Kötter, T., 2024. The green window view index: automated multi-source visibility analysis for a multi-scale assessment of green window views. *Landscape Ecology*, 39(71). https://doi.org/10.1007/s10980-024-01871-7.

d’Angelo, P., Reinartz, P., 2011. Semiglobal Matching Results on the ISPRS Stereo Matching Benchmark. *International Archives of Photogrammetry and Remote Sensing*, XXXVIII-4/W19, 79–84.

DLR-KI, n.d. Modular mobility: one vehicle for all transport tasks. `https://www.dlr.de/en/ki/research-transfer/projects/imoger`. Accessed: 2026-03-16.

Faliagka, E., Christopoulou, E., Ringas, D., Politi, T., Kostis, N., Leonardos, D., Tranoris, C., Antonopoulos, C. P., Denazis, S., Voros, N., 2024. Trends in digital twin framework architectures for smart cities: A case study in smart mobility. *Sensors*, 24(5), 1665.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Gstaiger, V., Köhler, C. H., Hochstaffl, P., Bachmann, M., de los Reyes, R., Holzwarth, S., Tian, J., Gege, P., Paxa, O., Krauß, T., Merkle, N., Kurz, F., 2025. Airborne remote sensing for environmental and disaster management applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-G-20, 315–322.

Henry, C., Hellekes, J., Merkle, N., Azimi, S., Kurz, F., 2021. Citywide estimation of parking space using aerial imagery and OSM data fusion with deep learning and fine-grained annotation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 479–485.

Hlal, M., Baraka Munyaka, J.-C., Chenal, J., Azmi, R., Diop, E. B., Bounabi, M., Ebnou Abdem, S. A., Almouctar, M. A. S., Adraoui, M., 2025. Digital Twin Technology for Urban Flood Risk Management: A Systematic Review of Remote Sensing Applications and Early Warning Systems. *Remote Sensing*, 17(17), 3104.

Hou, W., Walz, U., 2014. Extraction of small biotopes and ecotones from multi-temporal RapidEye data and a high-resolution normalized digital surface model. *International Journal of Remote Sensing*, 35(20), 7245–7262. http://dx.doi.org/10.1080/01431161.2014.967890.

Islam, M. R., Subramaniam, M., Huang, P.-C., 2025. Image-based deep learning for smart digital twins: a review. *Artificial Intelligence Review*, 58(5), 146.

Jangir, S. K., Mühlhaus, M., Merkle, N., Henry, C., Bahmanyar, R., 2026. A Generalized Two-Stage Framework for Blind Super-Resolution of Real-World Aerial and Satellite Imagery. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Submitted.

Javan, F. D., Samadzadegan, F., Toosi, A., 2025. Air pollution observation—bridging spaceborne to unmanned airborne remote sensing: a systematic review and meta-analysis. *Air Quality, Atmosphere & Health*, 18(8), 2481-2549. https://doi.org/10.1007/s11869-025-01771-y.

Kurz, F., Merkle, N., Henry, C., Bahmanyar, R., Rauch, F., Hellekes, J., Gstaiger, V., Rosenbaum, D., Reinartz, P., 2025. Generating Training Data for Deep Learning-Based Segmentation Algorithms by Projecting Existing Labels onto Additional Aerial Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 189–195.

Li, L., Han, L., Ye, Y., Xiang, Y., Zhang, T., 2025. Deep learning in remote sensing image matching: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 225, 88–112.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166–177.

Merkle, N., Henry, C., Azimi, S. M., Kurz, F., 2022. Road Condition Assessment from Aerial Imagery using Deep Learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2022, 283–289. https://isprs-annals.copernicus.org/articles/V-2-2022/283/2022/.

Mutreja, G., Schuegraf, P., Bittner, K., 2025. HiRes-FusedMIM: A High-Resolution RGB-DSM Pre-trained Model for Building-Level Remote Sensing Applications. *arXiv preprint arXiv:2503.18540v1*. https://arxiv.org/pdf/2503.18540v1.

Osco, L. P., Junior, J. M., Ramos, A. P. M., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., Li, J., 2021. A review on deep learning in UAV remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102456.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-607-2026 | © Author(s) 2026. CC BY 4.0 License.

Rosenbaum, D., Kurz, F., Thomas, U., Suri, S., Reinartz, P., 2008. Towards automatic near real-time traffic monitoring with an airborne wide angle camera system. *European Transport Research Review*, 1, 11–21. https://elib.dlr.de/56538/.

Schrotter, G., Hürzeler, C., 2020. The digital twin of the city of Zurich for urban planning. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1), 99–112.

Schuegraf, P., Shan, J., Bittner, K., 2024. PLANES4LOD2: Reconstruction of LoD-2 building models using a depth attention-based fully convolutional neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 425-437.

Schuegraf, P., Zorzi, S., Fraundorfer, F., Bittner, K., 2023. Deep learning for the automatic division of building constructions into sections on remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 7186–7200.

Sharifi, A., Beris, A. T., Javidi, A. S., Nouri, M., Lonbar, A. G., Ahmadi, M., 2024. Application of artificial intelligence in digital twin models for stormwater infrastructure systems in smart cities. *Advanced Engineering Informatics*, 61, 102485.

Somanath, S., Naserentin, V., Eleftheriou, O., Sjölie, D., Wästberg, B. S., Logg, A., 2023. On procedural urban digital twin generation and visualization of large scale data. *arXiv preprint arXiv:2305.02242*.

Sun, T., Di, Z., Che, P., Liu, C., Wang, Y., 2019. Leveraging crowdsourced gps data for road extraction from aerial imagery. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7509–7518.

Tian, J., Panangian, D., Fan, W., Siegmann, B., Yuan, X., 2025. Deep learning based individual tree crown delineation from panchromatic aerial imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 885–892.

Tian, J., Zhuo, X., Auer, S., Kurz, F., Reinartz, P., 2023. Fusion of stereo aerial images and official surveying data for mapping curbstones using ai. *2023 8th International Conference on Signal and Image Processing (ICSIP)*, IEEE, 176–180.

Wu, P., Zhang, Z., Peng, X., Wang, R., 2024. Deep learning solutions for smart city challenges in urban development. *Scientific Reports*, 14(1), 5176.

Xie, Y., Tian, J., Zhu, X. X., 2020. Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geoscience and remote sensing magazine*, 8(4), 38–59.

Xu, H., Omitaomu, F., Sabri, S., Zlatanova, S., Li, X., Song, Y., 2024. Leveraging generative AI for urban digital twins: a scoping review on the autonomous generation of urban data, scenarios, designs, and 3D city models for smart city advancement. *Urban Informatics*, 3(1), 29.

Xu, Y., Wang, T., Skidmore, A. K., 2025. Fusing aerial photographs and airborne LiDAR data to improve the accuracy of detecting individual trees in urban and peri-urban areas. *Urban Forestry Urban Greening*, 105, 128696.

Yang, Q., Su, Y., Jin, S., Kelly, M., Hu, T., Ma, Q., Li, Y., Song, S., Zhang, J., Xu, G., 2019. The Influence of Vegetation Characteristics on Individual Tree Segmentation Methods with Airborne LiDAR Data. *Remote Sensing*, 11(23), 2880. https://doi.org/10.3390/rs11232880.

Yuan, X., Tian, J., Krauß, T., Zhuo, X., Reinartz, P., 2023. Multi-layer thematic map representation for urban understanding. *2023 Joint Urban Remote Sensing Event (JURSE)*, IEEE, 1–4.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine*, 5(4), 8–36.

Zhuo, X., Tian, J., 2025. YOLO-Pole: A Deep Learning Framework for Precise Pole Localization in Aerial Orthophotos. *IEEE Geoscience and Remote Sensing Letters*, 22, 6008905.