Towards Country-Wide LoD1 City Model Reconstruction of from TanDEM-X Intensity Images

Michael Schmitt, Michael Recla, Christopher Ummerle, Islam Mansour

Department of Aerospace Engineering, University of the Bundeswehr Munich, Neubiberg, Germany

Keywords: Synthetic Aperture Radar (SAR), Remote Sensing, Urban Areas

Abstract

3D city models have become an important piece of geoinformation. They are available in different Levels of Detail (LoD), which determine the amount of complexity provided in the model. LoD1 city models represent simple prismatic building volumes and are typically produced by means of remote sensing. In this article, we investigate the possibility for country-wide reconstruction of LoD1 city models from TanDEM-X intensity images by utilizing deep learning-based single-image height and building footprint reconstruction. As study area, we use the land surface of the country of Denmark. Our results show the general potential of this AI-based approach of country-wide city model reconstruction, which can serve as a data-efficient pipeline that is particularly well-suited in time-critical scenarios or for the exploitation of archive imagery of satellite missions with global data coverage.

1. Introduction

3D city models enable the analysis, visualization, and simulation of urban processes, supporting decision-making in areas like urban planning, environmental management, transportation, and disaster resilience (Biljecki et al., 2015). By providing a static, spatially accurate representation of the city’s geometry and semantics, they form the foundational layer of urban digital twins. Generally, they are provided in different Levels of Detail (LoD), which describe how geometrically and semantically detailed a city model is, and which are standardized in the CityGML data model (Gröger and Plümer, 2012, Kolbe et al., 2021) (see Fig. 1). While LoD1 models are simple building volumes based on building footprints and height information, without roof shapes or textures, LoD2 models show differentiated roof structures and more accurate building geometries. LoD3 and LoD4 models already represent the building exterior and interior, respectively, realistically. Thus, while city models of LoDs 3 and 4 require high-accuracy/high-resolution input data such as airborne or terrestrial very-high-resolution photogrammetry or laser scanning, city models of LoD 1 and 2 can generally also be derived from spaceborne remote sensing data. Here, the extraction of building footprints is relatively straightforward, although still dependent on the used sensor technology and the available spatial resolution. For high- and very-high-resolution optical satellite imagery, reliable solution have existed for several decades (Li et al., 2022, Sirko et al., 2021). For medium-resolution optical data and very-high-resolution synthetic aperture radar (SAR) data, only recently meaningful and scalable building footprint extraction results have been confirmed (Feng et al., 2023, Prexl and Schmitt, 2023, Recla and Schmitt, 2024a). However, the much more critical challenge is the reconstruction of height information. Again, for VHR optical data operational solutions, mostly relying on stereo photogrammetry (Zhao et al., 2023), but also increasingly on artificial intelligence (Amirkolaee and Arefi, 2019, Liu et al., 2020), exist. For other sensor technologies and resolution regimes, however, height reconstruction is still an open issue. As an example, SAR satellite missions such as TanDEM-X (Krieger et al., 2007) allow for SAR interferometry (InSAR), enabling precise height reconstruction from coherent image pairs by exploiting their phase difference for accurate triangulation. However, due to unavoidable effects caused by the system-inherent side-looking imaging principle of SAR sensors, urban areas cannot be well reconstructed by conventional InSAR techniques (Rossi and Gernhardt, 2013). SAR tomography (TomoSAR) has been tried as an alternative and did establish a means for the three-dimensional reconstruction of urban areas (Zhu and Bamler, 2010), but is limited by the fact that many coherent images of the target scene are required – making TomoSAR a costly and time-consuming surveying tool.

Figure 1. Different Levels of Detail for a 3D building model.

Notwithstanding to the situation described above, 3D city models are often reconstructed by means of data fusion. In this case, each piece of information is derived from the best possible source. For example, building footprints can come from VHR optical imagery or existing geodatabases such as OpenStreetMap, and height data can be provided by airborne laserscanning or photogrammetric 3D reconstruction. However, most of these data sources, in particular regarding height data, are time-consuming and/or expensive. The body of literature on this topic is vast and beyond the scope of this paper. The interested reader is referred to summary articles such as (Haala and Kada, 2010, Musialski et al., 2013).

In order to provide a single-source alternative to the above-described approaches, this paper seeks to investigate the possibility to use recently introduced SAR2Height approach for simultaneous extraction of building footprints and building heights from Stripmap SAR intensity images provided by the TanDEM-X mission. SAR2Height is a new AI-powered paradigm for the reconstruction of urban height maps from single VHR SAR intensity images (Recla and Schmitt, 2024b).

Using the land surface of Denmark as case study area, the goal

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-469-2026

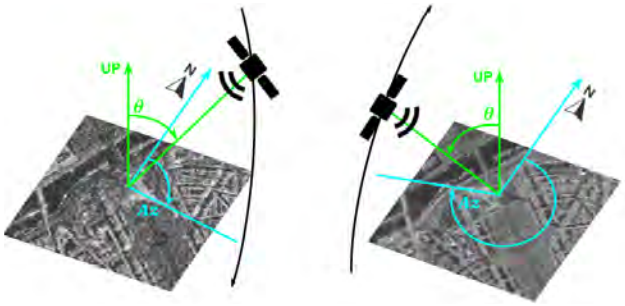


Figure 2. Conceptual illustration of the influence of the azimuth and the incidence angle on the appearance of elevated objects in VHR SAR imagery.

is to demonstrate the large-scale reconstruction of LoD1 city models from a single, weather-independent data source. On the one hand, this will enable a fruitful utilization of the data that is already available in the TanDEM-X archive, on the other hand it will provide a perspective for coarse 3D city modeling in time-critical scenarios such as disaster response or reconnaissance.

2. Country-Wide LoD1 City Model Reconstruction

2.1 Building Height and Footprint Prediction

For the simultaneous reconstruction of building heights and footprints as basis for the LoD1 city model reconstruction, we use the SARFormer model as the latest instance of a series of SAR2Height approaches designed for urban height map reconstruction from single SAR intensity images (Prexl et al., 2025). It is based on the well-established Vision Transformer (ViT) (Dosovitskiy, 2020) architecture and characterized by the following special features:

- It is enhanced by an acquisition parameter encoding (APE) module, wich allows to feed SAR-specific acquisition parameters such as spatial resolution, azimuth angle of the satellite track, and incidence angle of the acquisition directly into the deep neural network. Since these parameters are crucial for the appearance of urban objects in SAR imagery (cf. Fig. 2), the APE enables the training of models that can easily process data measured under heterogeneous acquisition conditions.
- As a sequence-to-sequence model, the SARFormer accepts an arbitrary number of input images and can theoretically output an arbitrary number of height and map layers. In the instance used in this paper, it predicts one height map in ground geometry (i.e. UTM coordinates), and a building footprint map in ground geometry.
- The SARFormer uses self-supervised pre-training of the masked autoencoder (MAE) kind (He et al., 2022). This enables the training on SAR images, for which no training data exist, and greatly reduces the need for labeled training examples.

A conceptual illustration of the SARFormer is shown in Fig. 3, while more details on its inner working are found in the original publication (Prexl et al., 2025).

The SARFormer model used for the experiments in this paper was trained on a dataset comprising 899 SAR intensity images from 125 cities. The ground truth building footprints were generated by fusing data from *OpenStreetMap* (OSM)[1] and *Microsoft Building Footprints*[2]. These fused labels, which cover all 125 locations (see Fig. 4 for the data distribution), were in parts manually curated to remove obvious errors. While footprint labels were available for all cities, corresponding height labels, derived from LiDAR measurements, were only available for a subset of 26 cities. The image data were provided by different satellite missions (Umbra, TerraSAR-X, Capella, and ICEYE) and utilizing two acqusition modes (Stripmap and Spotlight variants). Incidence angles ranged from 12° to 78°. Pre-training on an NVIDIA DGX using 7 H100 GPUs took approximately 31 hours. Fine-tuning was done in approximately 10 hours on 4 H100 GPUs.

For building height and footprint reconstruction, SAR images of the target area are fed into the model, and the desired outputs are provided in the form of raster images, which can be stored, e.g., in GeoTiff format.

2.2 Post-Processing and Vectorization

After the prediction of the raster height and footprint maps using the SARFormer, post-processing is needed to convert these raster predictions into vector-based LoD1 building models, where each building part is represented by a single polygon with an associated representative height. To achieve this, a gradient-guided watershed segmentation (Vincent and Soille, 1991) is applied to the height map within the predicted footprint masks. This step is crucial for partitioning large, complex building conglomerates into smaller, more height-consistent parts. The algorithm first computes the spatial gradient of the height map to identify areas of abrupt height change, which act as barriers. Markers for the segmentation are then generated by identifying stable, low-gradient regions; these are typically found by performing a connected-component analysis on the non-barrier areas and seeding the watershed from the peaks of an internal distance transform. The watershed segmentation then expands from these markers, using the gradient magnitude as the topographical relief, thereby partitioning the footprints along the detected high-gradient ridges.

This segmentation approach, whose effect is illustrated in Fig. 5, allows for a tunable sensitivity to decompose complex structures into segments that better represent the underlying height variations. Following the segmentation, the resulting raster-based building parts are vectorized into polygons. For each polygon, representative height statistics (such as the mean, median, and 98th percentile) are computed from the corresponding height map pixels. To assign an absolute base elevation, a globally available digital terrain model is then sampled at each polygon’s centroid. Finally, the polygon geometries are simplified to reduce vertex complexity using the Douglas-Peucker algorithm (Douglas and Peucker, 1973), yielding the final LoD1 building models with associated height attributes.

2.3 Mosaicking

In areas with overlapping SAR acquisitions, individual LoD1 building models produced from the individual acquisitions may differ slightly. We resolve these differences by first grouping intersecting polygons via a union-find over a self spatial join

[1] https://www.openstreetmap.org
[2] https://github.com/microsoft/GlobalMLBuildingFootprints/

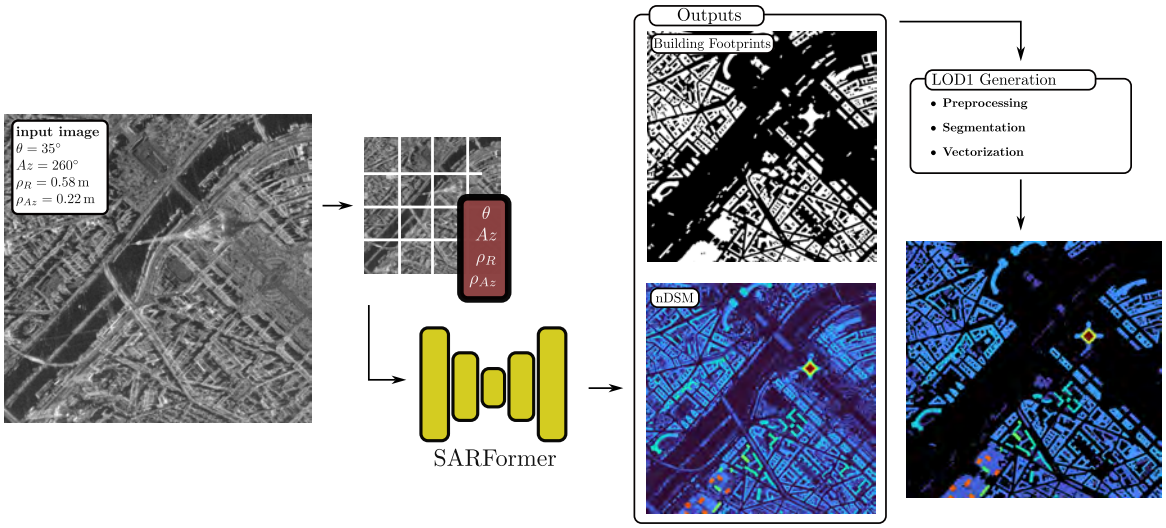


Figure 3. Conceptual illustration of the SARFormer approach. Note that while the model generally accepts an arbitrary number of different SAR acqusitions with different acqusition geometry as input (e.g. Stripmap mode, and Spotlight mode), we use only a single Stripmap image as input in this work. As shown, the model provides two data layers as output: a height map and a building footprint map in map geometry.

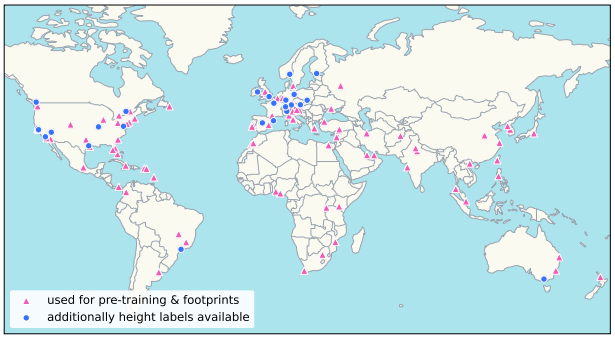
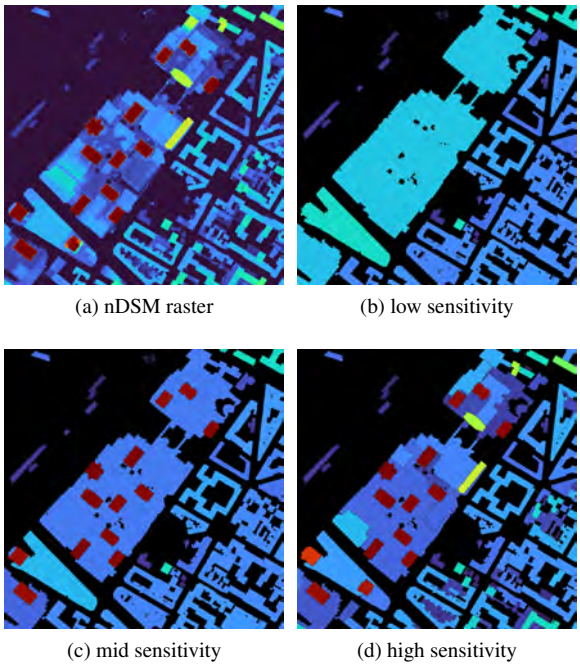


Figure 4. Global distribution of the training data used for SARFormer training. While building footprint labels have been available for all cities, LiDAR-based height reference data was only available for a subset.

and dissolving to one larger LoD1 polygon per connected component. In this step, numeric attributes (e.g., mean height, max height, elevation) were averaged from the overlapping polygons per connected component. From the boundaries of the input we construct a planar subdivision by polygonizing the global edge network. For each resulting cell we compute the count of original polygons fully containing the cell to robustly capture overlap intensity. For every dissolved footprint, we then select the highest-overlap cells in strict tiers (including all cells within a tier) until at least half of the connected component’s area is covered, and finally merge touching selections, re-averaging numeric attributes. This results in the highest-overlap LoD1 polygons that best cover the union of the individual buildings in the acquisitions while maintaining LoD1 attributes (see Fig. 6).

3. Study Area and Data

To demonstrate how the SAR2Height/SARFormer approach can be used to provide LoD1 city models for an extended area (e.g. a whole country), we selected Denmark as case study. This choice is based on the following considerations:



(a) nDSM raster (b) low sensitivity

(c) mid sensitivity (d) high sensitivity

Figure 5. Effect of the gradient-guided watershed segmentation on complex building conglomerates. (a) The input nDSM raster. (b) A low-sensitivity segmentation aggregates the complex into a single instance. (c-d) Results of our algorithm with medium (c) and high (d) sensitivity.

(a) before mosaicking (b) after mosaicking

Figure 6. Effect of the mosaicking process that is needed to combine adjacent and partially overlapping tiles.

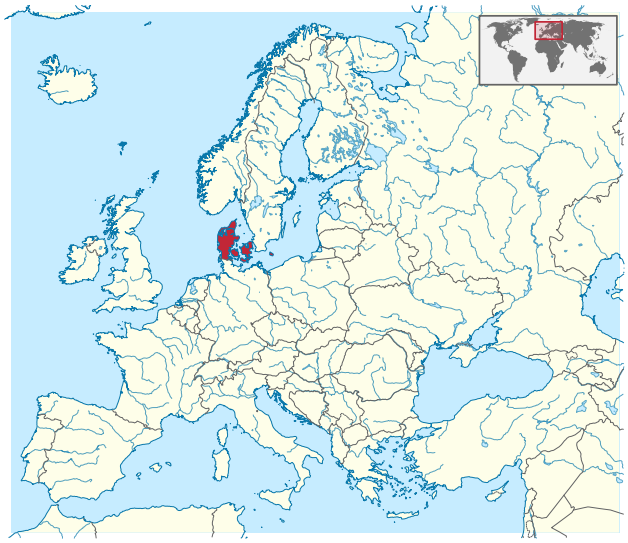


Figure 7. The country of Denmark in Europe. Source: wikimedia.org (CC BY-SA 3.0)

- With an area of 42,947 km^2, Denmark is a mid-sized European country (cf. Fig. 7) with one large metropolitan area, several medium-sized cities and abundant towns. It is thus supposed to serve as a viable example for urbanity in the western world.
- Denmark’s country-wide urbanization has been mapped before with remote sensing means in scientific publications, e.g. (Chen et al., 2020).
- As announced in 2025, Denmark is working on a country-wide digital twin of its cities [3].

To carry out the mapping, 75 TanDEM-X acquisitions have been downloaded from the TanDEM-X archive of the German Aerospace Center. While most of the images were acquired in 2023, a minor subset was acquired in 2018, while two images had to be replaced by data from 2020 due to strong radar frequency interference (see Fig. 8). All images were processed using the SARFormer to predict both building footprint masks and height maps in UTM geometry.

4. Results

An overview of the final result, i.e. a country-wide set of LoD1 building models for Denmark, produced only from archival

[3] see https://www.eng.klimadatastyrelsen.dk/green-transition-and-climate-adaptation/denmarks-digital-twin

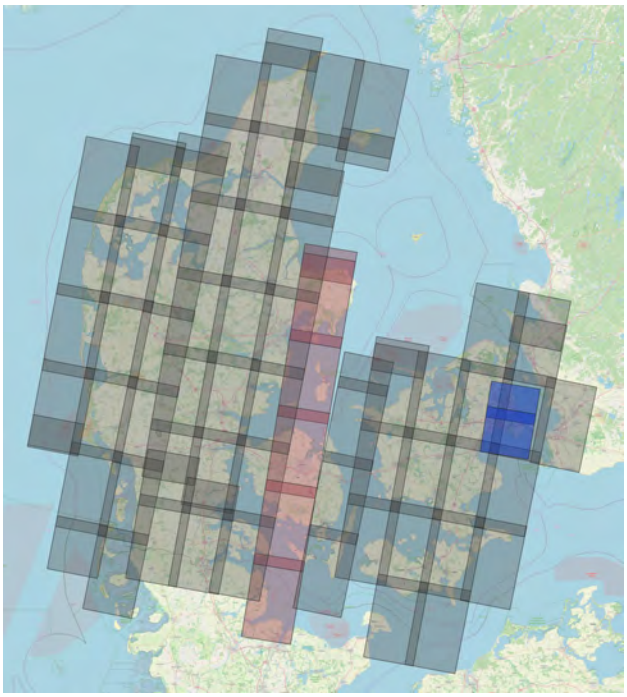


Figure 8. The footprints of the 77 utilized TanDEM-X acquisitions. The images represented by gray footprints were acquired in 2023, the images represented by the red footprints in 2018. The two blue acquisition footprints were additionally collected (from 2020) as replacements for the 2023 image, which suffers strongly from interference patterns.

TanDEM-X Stripmap data is shown in Fig. 9. The high building density in the large metropolitan area of Copenhagen at the eastern border is just as visible as several medium-sized cities like Odense, Aarhus, and Aalborg and the many smaller towns spread throughout the country.

A more detailed view for three different areas with different characteristics is shown in Fig. 10. Here, the wide applicability of the presented approach is illustrated. Figure 11 shows a rendered 3D view of a part of Denmark’s capital Copenhagen. While overall visually reasonable results for dense city areas, residential suburbs and rural settlements alike are achieved, due to a lack of ground truth information, no quantitative results can be reported.

5. Discussion

As the results shown in Section 4 show, the SARFormer-based approach can generally be used to reconstruct country-wide LoD1 building models from TanDEM-X stripmap data, which is available for the whole world. Building heights and footprints are reconstructed from single SAR intensity images, polygonized and eventually mosaicked to cover large areas beyond individual satellite observations. However, a detailed view at the results shown in Figs. 9 and 10, as well as the 3D rendering depicted in Fig. 11 also indicates the two most critical limitations of the current workflow:

- In its current implementation, the workflow does not yet involve any geometrical regularization. Thus, buildings can take arbitrary shapes and do not necessarily have rectangular corners. While this on the one hand allows geometrical flexibility, it on the other hand limits geometrical accuracy and fidelity.

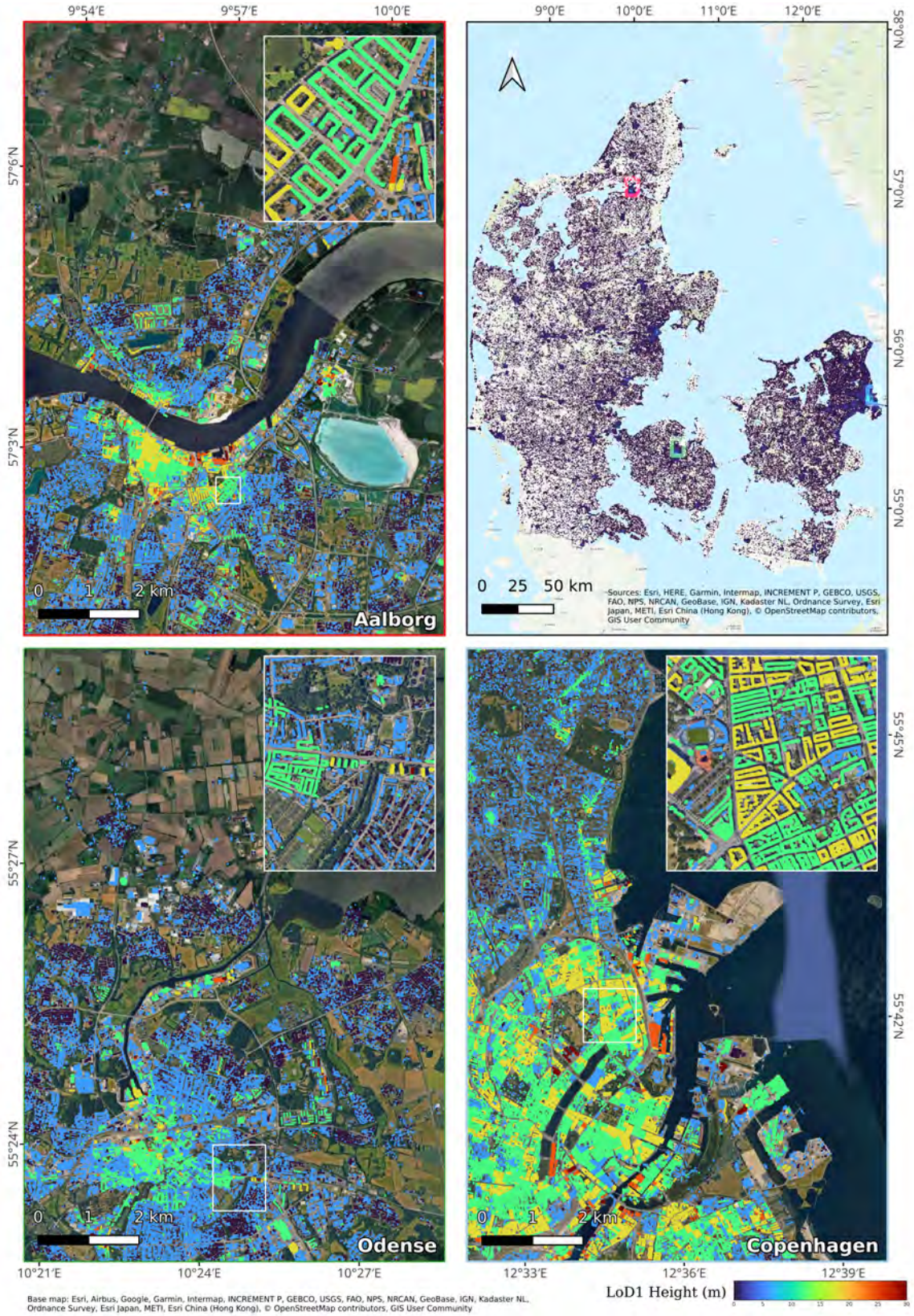


Figure 9. Country-wide reconstruction of LoD1 building models for Denmark from TanDEM-X Stripmap imagery using the SARFormer workflow. Shown are representative examples for Aalborg (top left), Odense (bottom left), and Copenhagen (bottom right), overlaid on optical basemaps with predicted building heights color-coded from low (blue) to high (red). The map inset (top right) indicates the national coverage and locations of the detailed views. Insets in each city panel highlight local variations in height prediction and building delineation quality.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-469-2026 | © Author(s) 2026. CC BY 4.0 License.

- The restriction to building models of LoD1 is only providing a coarse representation of the actual 3D scene characteristics. While this coarse representation might be sufficient for many use cases and applications, a future upgrade to LoD2 would certainly be interesting.

In addition to that, a detailed, qualitative evaluation will also be necessary to provide a full picture of the applicability of the approach in production settings.

6. Summary & Conclusion

In this paper, we have demonstrated an experimental workflow to use deep learning-based single-image height estimation and building footprint extraction for country-wide 3D city model reconstruction in LoD1. The backbone of this workflow is the SARFormer approach, a ViT-based multi-task network tailored to the peculiarities of very-high-resolution SAR data and trained on global multi-mission SAR imagery in a self-supervised manner. Using archival Stripmap SAR data of the TanDEM-X satellite mission, the feasibility of the presented workflow is demonstrated. Results show the potential of the workflow, but also illustrate its current limitations with respect to the geometrical realism of the reconstructed building models. Future work will have to focus on this aspect, potentially upgrading the output to regularized LoD2 building models.

Acknowledgments

The authors would like to thank the German Aerospace Center (DLR) for providing TerraSAR-X and TanDEM-X data under the data grants MTH3753 and OTHER7739. They would also like to thank the companies Umbra, Capella, and ICEYE for providing data through their open data programs. This project was partially funded by the German Research Foundation (DFG) under grant SCHM 3322/3–1.

References

Amirkolaee, H. A., Arefi, H., 2019. Height estimation from single aerial images using a deep convolutional encoder-decoder network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 50–66.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Çöltekin, A., 2015. Applications of 3D City Models: State of the Art Review. *ISPRS International Journal of Geo-Information*, 4(4), 2842–2889.

Chen, T.-H. K., Qiu, C., Schmitt, M., Zhu, X. X., Sabel, C. E., Prishchepov, A. V., 2020. Mapping horizontal and vertical urban densification in Denmark with Landsat time-series from 1985 to 2018: A semantic segmentation solution. *Remote Sensing of Environment*, 251, 112096.

Dosovitskiy, A., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929*.

Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2), 112–122.

(a) Inner city area with dense built-up structure and larger buildings.

(b) Suburban area with mostly residential and some larger commercial buildings.

(c) Rural small town with mostly residential detached homes.

Figure 10. Three zoomed-in views of randomly selected areas representative for three different settlement types. In all three cases reasonable LoD1 building models have been produced.

Feng, L., Xu, P., Tang, H., Liu, Z., Hou, P., 2023. National-scale mapping of building footprints using feature super-resolution semantic segmentation of Sentinel-2 images. *GIScience & Remote Sensing*, 60(1), 2196154.

Gröger, G., Plümer, L., 2012. CityGML – Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 12–33.

Haala, N., Kada, M., 2010. An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 570–580.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. *Proc. of the CVPR*, 16000–16009.

Kolbe, T. H., Kutzner, T., Smyth, C. S., Nagel, C., Roensdorf, C., Heazel, C., 2021. OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard. *OGC Document: 20-010*.

Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., Zink, M., 2007. TanDEM-X: A Satellite Forma-

Figure 11. Rendered 3D view of a part of Copenhagen.

tion for High-Resolution SAR Interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11), 3317–3341.

Li, J., Huang, X., Tu, L., Zhang, T., Wang, L., 2022. A review of building detection from very high resolution optical remote sensing images. *GIScience & Remote Sensing*, 59(1), 1199–1225.

Liu, C.-J., Krylov, V. A., Kane, P., Kavanagh, G., Dahyot, R., 2020. IM2ELEVATION: Building Height Estimation from Single-View Aerial Imagery. *Remote Sensing*, 12(17).

Musialski, P., Wonka, P., Aliaga, D. G., Wimmer, M., Van Gool, L., Purgathofer, W., 2013. A survey of urban reconstruction. *Computer Graphics Forum*, 32(6), 146–177.

Prexl, J., Recla, M., Schmitt, M., 2025. SARFormer – an acquisition parameter aware vision transformer for synthetic aperture radar data. *Proc. of the CVPR Workshops*, 2225–2234.

Prexl, J., Schmitt, M., 2023. The potential of Sentinel-2 data for global building footprint mapping with high temporal resolution. *Proc. of Joint Urban Remote Sensing Event*.

Recla, M., Schmitt, M., 2024a. Deep learning-based building footprint mapping using high-resolution SAR data. *Proc. of IEEE International Geoscience and Remote Sensing Symposium*, 9983–9986.

Recla, M., Schmitt, M., 2024b. The SAR2Height framework for urban height map reconstruction from single SAR intensity images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 104–120.

Rossi, C., Gernhardt, S., 2013. Urban DEM generation, analysis and enhancements using TanDEM-X. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85, 120–131.

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., Quinn, J., 2021. Continental-Scale Building Detection from High Resolution Satellite Imagery. *arXiv:2107.12283*.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(6), 583–598.

Zhao, L., Wang, H., Zhu, Y., Song, M., 2023. A review of 3D reconstruction from high-resolution urban satellite images. *International Journal of Remote Sensing*, 44(2), 713–748.

Zhu, X. X., Bamler, R., 2010. Very high resolution spaceborne SAR tomography in urban environment. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12), 4296–4308.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-469-2026 | © Author(s) 2026. CC BY 4.0 License.