The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium “Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation”, 25–29 August 2025, Seoul, Republic of Korea

Towards LOD-2 Building Reconstruction: Leveraging Segmentation and Roof Shape Extraction Methods from VHR Imagery

Vaibhav Rajan[1], Sander Münster[1], Jonas Bruschke[1], Ferdinand Maiwald[2]

[1] Digital Humanities, Friedrich-Schiller-Universität Jena, Germany - (vaibhav.rajan, sander.muenster, jonas.bruschke)@uni-jena.de
[2] Chair of Optical 3D-Metrology, TU Dresden, Germany - ferdinand.maiwald@tu-dresden.de

Keywords: LOD-2 Models, VHR Satellite Imagery, Roof Shape Detection, Building Segmentation, Edge Reconstruction, Feature Extraction.

Abstract

Accurate extraction of roof structures from aerial imagery is a critical step in the creation of detailed 3D models for digital heritage reconstruction. This study explores a hybrid methodology that combines prompt-based segmentation with structured vector reconstruction to enhance the extraction of roof skeletons from Very High Resolution (VHR) orthophotos. Using HEAT (Holistic Edge Attention Transformer) as the primary reconstruction model, we fine-tuned it on a domain-specific dataset containing representative gabled and hipped roofs to adapt to the unique geometries found in the city of Jena, Germany. To test whether prior roof isolation could improve reconstruction performance, we integrated mask outputs from two segmentation models — RobustSAM and LangSAM — into the HEAT pipeline. While segmentation offered visually precise results in several instances, overall evaluation revealed that prior segmentation did not consistently improve HEAT's reconstruction accuracy. These findings underscore HEAT's robustness and adaptability, especially when properly fine-tuned. Moreover, while SAM variants did not significantly boost performance here, their ease of use and potential for improvement through domain-specific fine-tuning suggest promising applications in other contexts.

1. Introduction

Digital heritage preservation plays a crucial role in protecting and safeguarding cultural identities, enabling people to experience their historical urban environment virtually (UNESCO, 2003). Through different approaches in digital reconstruction, no longer existing (or altered) buildings can be restored and visualized, facilitating research, education, and cultural engagement (Münster, 2019). Among the critical elements in achieving authentic digital urban reconstructions are accurate 3D buildings, specifically detailed roof structures, which significantly define the visual integrity of the building itself. The present research contributes to our ongoing efforts in developing a comprehensive 4D city application, aimed at supporting urban history education and research (Münster et al., 2024).

3D city models are classified into Levels of Detail (LOD-0 to LOD-4), each indicating an increasing geometric and semantic complexity (Tan et al., 2023). LOD-0 and LOD-1 provide basic terrain and extruded footprints, often insufficient for detailed spatial analysis (Krafczek and Jabari, 2021). LOD-2 adds realistic roof geometries and building parts, enabling applications such as shadow simulation, runoff analysis, and solar potential mapping (Biljecki and Dehbi, 2019; Li et al., 2025). Higher LODs (LOD-3 and LOD-4) incorporate façade and interior details, overlapping with BIM domains for advanced facility management and virtual planning (Sun et al., 2019).

Among these, LOD-2 strikes a balance between geometric realism and computational feasibility, making it particularly valuable for applications that require structural accuracy at the surface level, especially in the context of historical modeling and analysis of roof types (Tan et al., 2023; Krafczek and Jabari, 2021). The accurate extraction of building roofs is an essential step in the creation of LOD-2 models. As OpenStreetMap (OSM) does not provide roof shape types in all regions, the building footprints can be utilized for LOD-1 generation only through simple extrusions. However, LOD-2 requires the identification of the inner roof structures, which cannot be obtained from footprints alone. To achieve this, Very High Resolution (VHR) satellite imagery and aerial orthophotos can be very useful. They provide a mostly unobstructed top-down view, enabling the possibility of detecting the roof skeletons. However, extracting accurate roof structures remains challenging due to image resolution limits, occlusions, and variations in roof geometry.

2. Related Work

2.1 Methods for Roof Skeleton Detection

Recent advances in deep learning models have shown promising results in automated feature extraction, particularly in roof structure reconstruction (Campoverde et al., 2024). Most recent state-of-the-art models include an end-to-end neural system. Methods like L-CNN (Zhou et al., 2019) and HAWP (Xue et al., 2020) follow, despite their differences, a similar two-stage approach: initially detecting corner features from the input image and subsequently classifying edge candidates based on these detected corners using a score-based classifier. Among them is HEAT – Holistic Edge Attention Transformer for Structured Reconstruction (Chen et al., 2022). In this approach, edge features are initialized using trigonometric positional encoding, followed by its fusion to the edge candidates. Finally, it uses two transformer decoders (image and edge) to classify each edge candidate (Chen et al., 2022).

2.2 Methods for Segmentation

The Segment Anything Model (SAM), introduced by Kirillov et al. (2023), marked a significant advance in prompt-based segmentation. It can generate high-quality masks from simple inputs such as bounding boxes or clicks, even in zero-shot settings (Kirillov et al., 2023). This flexibility allows SAM to be adapted to diverse scenarios, including VHR imagery. Its foundational performance makes it an excellent candidate for specific tasks like roof segmentation. Building on SAM, task-specific variants like RobustSAM (Chen et al., 2024) and LangSAM (Medeiros, 2023) have emerged. Although detailed peer-reviewed studies are limited, RobustSAM emphasizes making SAM less sensitive to low-quality and degraded images, enhancing segmentation accuracy under challenging imaging conditions (Chen et al., 2024). On the other hand, LangSAM (Language Segment Anything), built on Segment Anything Model 2 and GroundingDINO, combines SAM with language prompts, guiding segmentation through intuitive text prompts offering user-friendly control (Medeiros, 2023).

Other deep learning models for building footprint extraction from VHR imagery also remain crucial. Architectures such as U-Net (Ronneberger et al., 2015), DeepLabv3+ (Chen et al., 2018), and Res-U-Net have served as effective baselines, typically yielding around 70–80 percent Intersection over Union (IoU) when adequately trained with sufficient data (Yuan, 2018; Zhang et al., 2018). Recent transformer-based approaches, including Transformer-U-Net and Swin Transformer hybrids, introduce global attention mechanisms, handling complex urban layouts better and further improving segmentation accuracy (Song et al., 2023; Xiao et al., 2022).

This study explores a methodology that integrates segmentation-based object isolation with structural vector reconstruction to improve the accuracy of roof shape extraction. It aims to develop and evaluate a pipeline combining SAM segmentation methods (RobustSAM[1] and LangSAM[2]) with structured roof reconstruction algorithm (HEAT[3]) on VHR imagery. As a preliminary step in the research, the approach presented herein was intentionally tested on simple, quadrilateral roof structures to ensure clear, interpretable results to serve as a proof of concept. By first validating our methodology against simpler roof geometries, we can rigorously assess the fundamental feasibility, accuracy improvements, and integration effectiveness of our methodology.

3. Methodology

3.1 Data Description and Pre-processing

The dataset used in the research are VHR (GSD 0.2 m) aerial orthophotos of the city of Jena, Germany from the year 2023, retrieved from the Thuringian state geoportal website (`https://geoportal.thueringen.de/`). Around 9000+ buildings were extracted (Figure 1) from the image through OSM building footprints via the Overpass API. The buildings were cropped after overlaying the polygons on the image with a 2 meter padding ensuring complete coverage of the buildings.

Since the extracted building patches were of varying sizes and in GeoTIFF format, they were not suitable for direct input to our downstream model. To meet the model's requirements, each patch was first converted to a PNG format and resized to a fixed resolution of 256×256 pixels. This standardization, however, resulted in a loss of georeferencing information. To preserve spatial context, georeferencing metadata, including bounding coordinates and scaling factors, was extracted beforehand and saved as accompanying JSON files. These would later be used to restore the geographic reference of the model's output.

[1] `https://github.com/robustsam/RobustSAM`
[2] `https://github.com/luca-medeiros/lang-segment-anything`
[3] `https://github.com/woodfrog/heat`

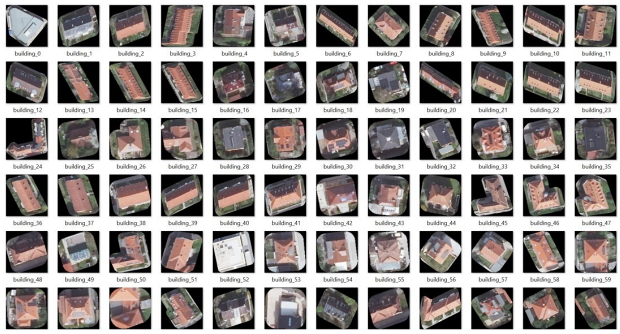

Figure 1. 9000+ buildings extracted from the aerial image via Overpass API.

3.2 HEAT: Holistic Edge Attention Transformer for Structured Reconstruction

We decided to use HEAT because one of its modules is specifically designed for outdoor structured vector reconstruction, making it particularly well suited for tasks like roof extraction. Among the available models, HEAT stands out for its state-of-the-art performance in reconstruction tasks, as demonstrated in its original benchmark results (Chen et al., 2022). The HEAT model provides two pre-trained checkpoints trained and evaluated on satellite image patches from cities such as Paris, Las Vegas, and Atlanta, sourced from the SpaceNet Challenge (Van Etten et al., 2019). Its open-source availability with pre-trained weights made it a practical choice for our study.

The HEAT model is designed to extract a structured graph of corners and connecting edges from an image, such as those found in building roof outlines. It operates in two main stages. First, it predicts potential corner points across the image using a convolutional backbone combined with position-aware encodings and attention mechanisms. These corner candidates are refined through non-maximum suppression to select high-confidence points. Next, it constructs all possible edge connections between detected corners and represents each edge as a feature vector. These edge features are enriched with contextual visual information by sampling image features along the line connecting each corner pair, using deformable attention.

HEAT then classifies which of these edge candidates are valid using two transformer decoders: one that processes image-aware features and another that uses only geometric information derived from corner positions. Both decoders share weights and jointly optimize the edge classification task. This dual-branch design ensures that structural patterns are reinforced, even in cases where image texture is ambiguous. The result is a robust prediction of planar wireframes representing the structure of building roofs.

As most buildings in the test city featured diverse roof types such as sloped and gabled structures, we fine-tuned the model to better suit our dataset, since the original model had been trained primarily on flat roofs using around 2000 image samples. To perform this fine-tuning, we used a set of 30 manually annotated roof images from our dataset, which contains primarily

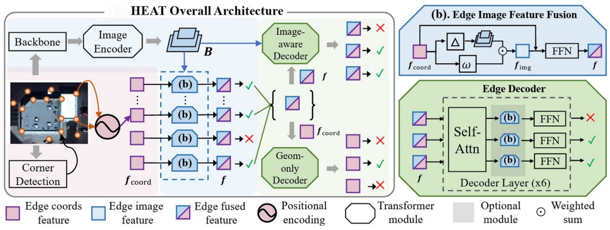

Figure 2. Network architecture of the HEAT model, illustrating the two-stage process of corner detection and edge classification using transformer decoders. Modified from Chen et al. (2022).

simple roof types comprising of gabled and hipped structures. An additional set of 20 samples was reserved for evaluation. Instead of retraining the entire model, we selectively fine-tuned only the corner detection and edge classification modules using these annotations. This strategy allowed us to adapt the model to the varying roof types present in our dataset while retaining the benefits of the learned features from the pre-trained model.

We fine-tuned the model using a batch size of 2 and a learning rate of 0.0002, optimized with AdamW. Training lasted for 400 epochs, reducing the learning rate at epoch 300. Only the parts responsible for detecting corners and classifying edges were updated, while the rest of the model remained frozen. We kept key settings from the original model, such as corner loss weights and edge sampling. Later, we evaluated HEAT, both independently and together with segmentation outputs from RobustSAM and LangSAM, to see how they affect the final roof skeleton results.

3.3 Segmentation Methods: RobustSAM and LangSAM

We used SAM variants primarily due to their adaptability. SAM’s support for prompt-based interaction allows targeted segmentation through manual or automated prompts. RobustSAM, fine-tuned and tested on degraded images, has been put to test for handling noise and low-quality images. Alternatively, we also test LangSAM. It’s powerful text-based prompts significantly aid in targeting specific semantic classes.

RobustSAM is a fine-tuned variant of the SAM model, developed to perform better on degraded, low-resolution, or noisy images. In our study, we tested both supported prompt types — point-based and bounding box — to guide segmentation. Point-based prompts offered greater precision by allowing interactive selection of specific roof regions, while bounding boxes provided broader spatial context. This allowed us to compare how each prompt type influenced segmentation quality.

Built on top of Segment Anything Model 2 and Grounding-DINO, LangSAM enables intuitive text-based commands to guide the mask generation. Two modes of operation were available for LangSAM: 1) an interactive web-based interface and 2) a Python library. While the interface allowed easy visualization and adjustment of prompt sensitivity (e.g., box and text thresholds), it lacked flexibility in exporting segmentation masks in a format suitable for downstream use in HEAT. In contrast, the library version offered full control over mask extraction and enabled batch processing across multiple images. Both segmentation outputs were later integrated with the HEAT model for downstream structured reconstruction and performance evaluation.

4. Results and Evaluations

This section presents the results of our experiments, divided into two major components: 1) the fine-tuning evaluation of the HEAT model and 2) a comparative analysis of segmentation performance using RobustSAM and LangSAM.

4.1 HEAT Fine-Tuning

To evaluate the effectiveness of the HEAT fine-tuning process, we compared the model performance before and after training on our annotated dataset. Key metrics — precision, recall, and F-1-score — were used to quantify improvements in corner detection and edge classification. The refined model demonstrated a notable reduction in false corners and spurious edges (Figure 3). It accurately captured simple roof geometries, such as gabled and hipped structures, with significantly improved coherence and fewer geometric artifacts. Table 1 shows the quantitative evaluation of the original versus the fine-tuned HEAT model.

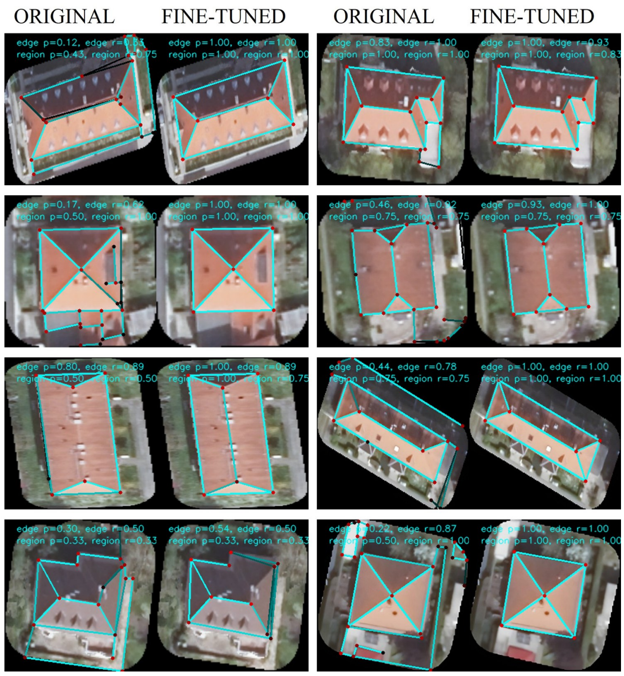

Figure 3. Results for HEAT showing the improvement in reconstruction for simple geometries. Left: result with original model; Right: result after fine-tuning.

The consistent improvement across all metrics highlights the model’s flexibility in learning domain-specific features and its capacity to produce more structured and reliable outputs. The results were achieved without any prior segmentation of the input imagery, indicating that the HEAT model, when fine-tuned, can effectively distinguish roof features from surrounding content in cases of simple roof types. Nevertheless, we proceed to evaluate whether integrating pre-segmented masks from RobustSAM and LangSAM contributes further to reconstruction accuracy or not.

4.2 Segmentation Methods

We first examined the standalone performance of the two segmentation models RobustSAM and LangSAM on our image patches. These models were applied to isolate roof regions from

Metric	Original	Fine-tuned	HEAT+RS	HEAT+LS
Corner Prec	0.38	0.65	**0.73**	**0.73**
Corner Recall	**0.85**	0.83	0.78	0.72
F-1 Score	0.52	0.73	**0.75**	0.72
Edge Prec	0.49	**0.79**	0.78	0.76
Edge Recall	0.61	**0.71**	0.69	0.60
F-1 Score	0.54	**0.75**	0.73	0.67
Region Prec	0.62	0.72	**0.75**	0.67
Region Recall	0.61	0.60	**0.62**	0.52
F-1 Score	0.61	0.66	**0.67**	0.58

Table 1. Quantitative evaluation of original vs. fine-tuned HEAT model.

the background using prompt-based inputs: point clicks and bounding boxes for RobustSAM and natural language prompts for LangSAM. A total of 50 building samples were used for evaluation representing a diverse set of roof geometries.

During evaluation, we observed notable differences between the two prompt modes available in RobustSAM. As shown in Figure 4, bounding box prompts sometimes captured additional elements near the target structure, resulting in unwanted isolated segments (highlighted with green circles). In contrast, point-based prompts provided finer control over the segmentation, enabling more precise isolation of the roof region without accidental inclusion of background objects. Based on these observations, we adopted point-based prompting as the preferred mode for subsequent reconstruction tasks.

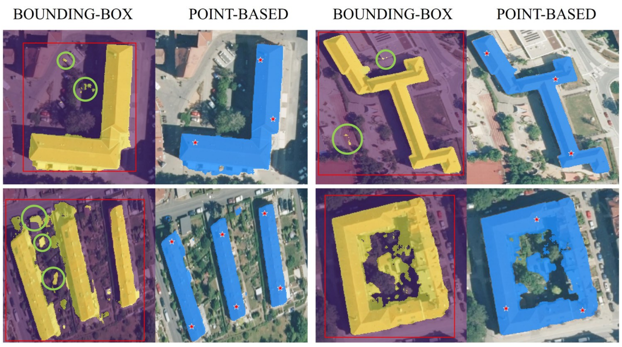

Figure 4. Comparison of RobustSAM segmentation using bounding box (left) vs. point-based prompts (right). Green circles indicate undesired isolated segments in the bounding box results.

We then evaluated the segmentation quality of LangSAM using natural language prompts. The text prompt "roof" was consistently used throughout the entire processing pipeline to extract roof regions from each image (Figure 5). At first inspection, LangSAM delivered promising and visually coherent segmentation masks. The use of language-based prompts enabled intuitive and semantically meaningful region selection, and in most cases, the segmentation performance was remarkably effective, even when the text prompt was highly unspecific like "roof".

Based on both qualitative visual assessments and quantitative evaluation metrics, the performance of RobustSAM and LangSAM appears comparably strong. The IoU and pixel accuracy values suggest that there is no significant difference between the two methods in terms of overall segmentation quality (Table 2). While both models demonstrate the ability to successfully isolate roof structures in most cases, challenges remain in more complex scenarios — particularly where low contrast between roof and surrounding regions exists. Finally, when these segmented outputs were integrated into the HEAT model to guide vector reconstruction, the overall improvements were found to be limited.

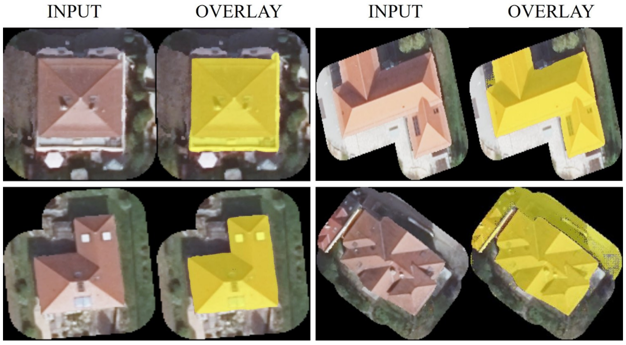

Figure 5. Overlay results from prompt-based segmentation model LangSAM. The prompt "roof" was used as input prompt.

RobustSAM		LangSAM	
IoU	PA	IoU	PA
0.7718	0.9036	0.7176	0.8842

Table 2. Segmentation performance of RobustSAM and LangSAM.

5. Conclusion

The primary objective of this study was to investigate whether the integration of prior segmentation — specifically using mask outputs from prompt-based models like RobustSAM and LangSAM — could enhance the performance of the HEAT model in roof skeleton reconstruction. By isolating roof regions in advance and supplying these refined inputs to the HEAT pipeline, we aimed to assess whether this preprocessing step would lead to more accurate and structurally coherent vector reconstructions compared to using raw imagery alone. However, on average, the segmentation process did not substantially affect the quality of vector reconstructions. In cases where the segmentation masks were accurate and covered the entire roof, the resulting vector reconstructions were largely unaffected (Figure 6 a, b). In instances where segmentation failed to capture the full extent of the roof — such as omitting parts of the structure — HEAT struggled to generate complete wireframes due to missing visual information (Figure 6 e). Similarly, when non-roof regions were incorrectly included in the mask, the model occasionally misinterpreted the context, leading to inaccurate edge constructions (Figure 6 c).

Overall, these outcomes indicate that segmentation had a limited impact on performance and in some cases even introduced challenges rather than resolving them. This suggests that the benefit of prior segmentation is highly dependent on the quality and precision of the mask, and that HEAT's performance, especially on simpler roof geometries, remains relatively robust even without explicit object isolation.

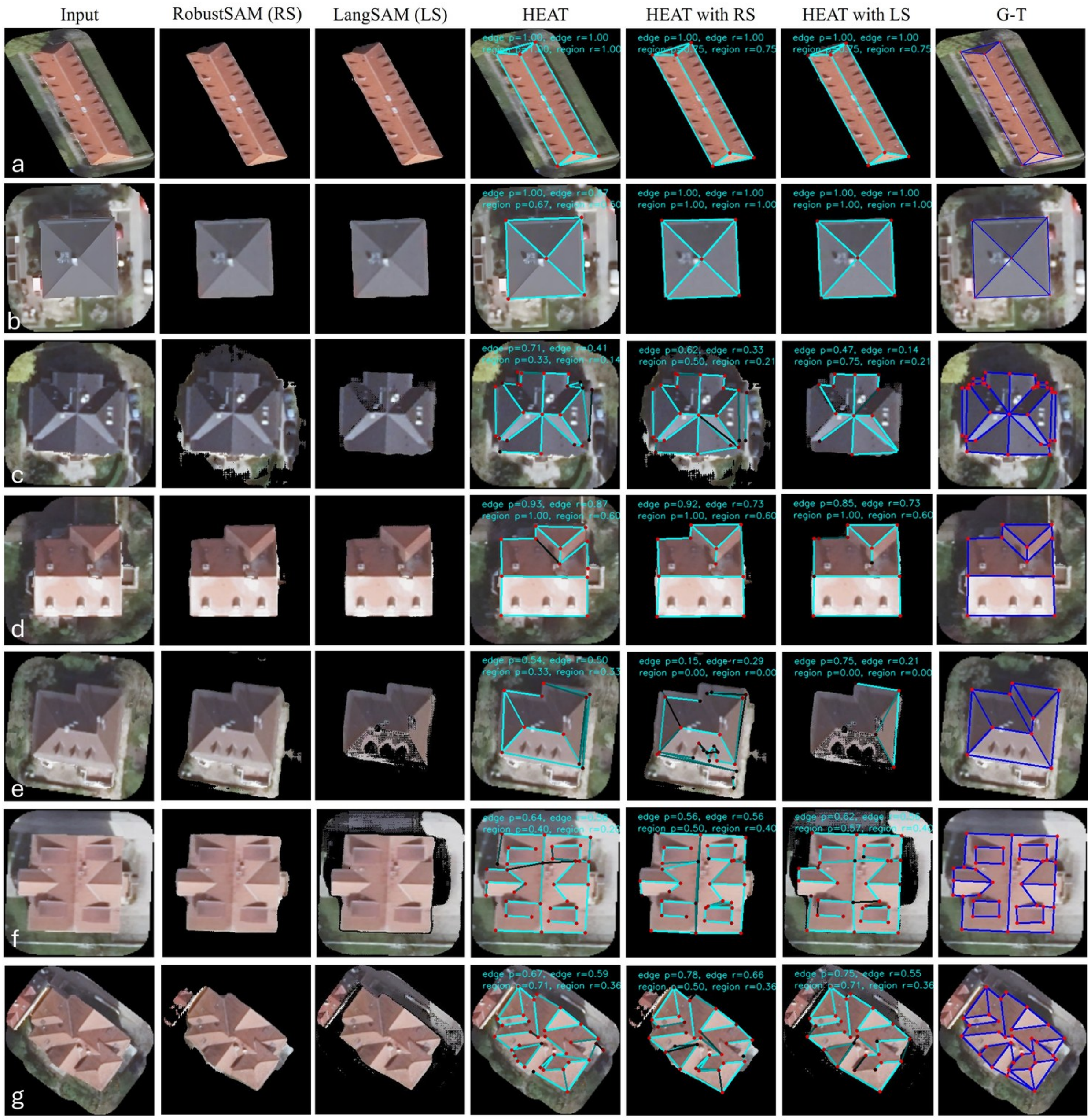

Figure 6. Qualitative evaluations of roof structure reconstruction with and without segmentation.

While segmentation did not yield significant improvements in this study, it still remains a powerful tool. Our findings suggest that with further fine-tuning on diverse roof types and varying image conditions, prompt-based segmentation models have strong potential in various applications. Although the pretrained SAM variants did not show major benefits in our specific pipeline, it is important to acknowledge that they were not trained or fine-tuned on our dataset. This gives a reasonable benefit of doubt, and it cannot be concluded that these models lack utility in general. Trained and refined on custom datasets, such models may still offer valuable contributions to future roof reconstruction workflows, underscoring their broader potential.

These findings also suggest that much of the HEAT model's performance is owed to its fine-tuning process. By exposing the model to a wide range of representative roof geometries during training, it learns to internally differentiate roofs from non-roof regions, reducing its dependence on external segmentation cues. Therefore, the model's success is tied to how well it has been fine-tuned to recognize domain-specific patterns. With enough exposure to representative samples, particularly those that capture complex roof geometries, the HEAT model becomes capable of distinguishing relevant structural cues — even in the presence of background noise. Thus, rather than relying heavily on prior segmentation, further performance gains may be achieved by expanding and diversifying the fine-tuning dataset.

6. Future work

In the preprocessing stage of our pipeline, geospatial metadata, such as bounding coordinates and scaling factors, were extracted from the original GeoTIFF image patches prior to their conversion and resizing for the HEAT model. This step ensured that the georeferencing information was preserved, allowing us to later reassign real-world spatial extents to the model's output predictions. This geospatial anchoring sets the stage for the next phase of our research — fusing the HEAT-generated 2D roof skeletons with 3D data such as LiDAR point clouds or Digital Surface Models (DSMs) to create detailed and accurate 3D wireframe representations of rooftops. While this work was mainly targeted on improving the models of the 4D city application regarding urban history education, the implications extend well beyond this scope. By enabling geo-accurate and semantically precise roof reconstructions, the proposed methodology can potentially benefit a range of urban analytics and planning domains like photovoltaic potential analysis, green roof planning and other structural inspection tasks.

Acknowledgments

We gratefully acknowledge the authors and developers of the HEAT (Chen et al., 2022), RobustSAM (Chen et al., 2024), and LangSAM (Medeiros, 2023) models for openly sharing their source code via GitHub. Their contributions have been instrumental in enabling this research and facilitating its reproducibility.

The research which this paper is based on was carried out in the EU projects INDUX-R (Grant No. 101135556), and 3DBigDataSpace (Grant No. 101173385).

References

Biljecki, F., Dehbi, Y., 2019. Raise the roof: Towards generating LOD2 models without aerial surveys using machine learning. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-4/W8, 27–34. doi.org/10.5194/isprs-annals-IV-4-W8-27-2019.

Campoverde, C., Koeva, M., Persello, C., Maslov, K., Jiao, W., Petrova-Antonova, D., 2024. Automatic Building Roof Plane Extraction in Urban Environments for 3D City Modelling Using Remote Sensing Data. *Remote Sens.*, 16(8), 1386. doi.org/10.3390/rs16081386.

Chen, J., Qian, Y., Furukawa, Y., 2022. HEAT: Holistic edge attention transformer for structured reconstruction. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3856–3865.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proc. European Conference on Computer Vision (ECCV)*.

Chen, W.-T., Vong, Y.-J., Kuo, S.-Y., Ma, S., Wang, J., 2024. RobustSAM: Segment anything robustly on degraded images. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4081–4091.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything. *arXiv preprint arXiv:2304.02643*.

Krafczek, M., Jabari, S., 2021. Generating LOD2 city models using a hybrid-driven approach: a case study for New Brunswick urban environment1. *Geomatica*, 75(3), 130–147. doi.org/10.1139/geomat-2021-0016.

Li, Q., Taubenböck, H., Zhu, X. X., 2025. Identification of the potential for roof greening using remote sensing and deep learning. *Cities*, 159, 105782. doi.org/10.1016/j.cities.2025.105782.

Medeiros, L., 2023. Language Segment-Anything. `https://github.com/luca-medeiros/lang-segment-anything`. Accessed: 2025-06-22.

Münster, S., 2019. Digital Heritage as a Scholarly Field—Topics, Researchers, and Perspectives from a Bibliometric Point of View. *J. Comput. Cult. Herit.*, 12(3), 22. doi.org/10.1145/3310012.

Münster, S., Bruschke, J., Dworak, D., Komorowicz, D., Rajan, V., Ukolov, D., 2024. 4D geo modelling from different sources at large scale. *Proc. 6th Workshop on the AnalySis, Understanding and ProMotion of HeritAge Contents*, SUMAC '24, ACM, New York, NY, USA, 13–17.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv:1505.04597*.

Song, J., Zhu, A.-X., Zhu, Y., 2023. Transformer-Based Semantic Segmentation for Extraction of Building Footprints from Very-High-Resolution Images. *Sensors*, 23(11), 5166. doi.org/10.3390/s23115166.

Sun, J., Mi, S., Olsson, P.-o., Paulsson, J., Harrie, L., 2019. Utilizing BIM and GIS for Representation and Visualization of 3D Cadastre. *ISPRS Int. J. Geo-Inf.*, 8(11), 503. doi.org/10.3390/ijgi8110503.

Tan, Y., Liang, Y., Zhu, J., 2023. CityGML in the Integration of BIM and the GIS: Challenges and Opportunities. *Buildings*, 13(7), 1758. doi.org/10.3390/buildings13071758.

UNESCO, 2003. Convention for the safeguarding of the intangible cultural heritage. Circular letter CLT-2003/WS/21, reprinted 2009. Adopted at the 32nd session of the General Conference of UNESCO.

Van Etten, A., Lindenbaum, D., Bacastow, T. M., 2019. SpaceNet: A Remote Sensing Dataset and Challenge Series. *arXiv preprint arXiv:1807.01232*.

Xiao, X., Guo, W., Chen, R., Hui, Y., Wang, J., Zhao, H., 2022. A Swin Transformer-Based Encoding Booster Integrated in U-Shaped Network for Building Extraction. *Remote Sens.*, 14(11), 2611. doi.org/10.3390/rs14112611.

Xue, N., Wu, T., Bai, S., Wang, F., Xia, G.-S., Zhang, L., Torr, P. H., 2020. Holistically-attracted wireframe parsing. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2785–2794.

Yuan, J., 2018. Learning Building Extraction in Aerial Scenes with Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(11), 2793–2798. doi.org/10.1109/TPAMI.2017.2750680.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.*, 15(5), 749–753. doi.org/10.1109/LGRS.2018.2802944.

Zhou, Y., Qi, H., Ma, Y., 2019. End-to-end wireframe parsing. *ICCV*.