Automatic Building Roof Plane Extraction in Urban Environments for 3D City Modelling Using Remote Sensing Data

Carlos E. Campoverde , Mila Koeva * , Claudio Persello , Konstantin Maslov , Weiqin Jiao , Dessislava Petrova-Antonova

Posted Date: 25 January 2024

doi: 10.20944/preprints202401.1839.v1

Keywords: roof structure extraction; image processing; deep-learning; HEAT; 3D modelling; LOD2

Article

Automatic Building Roof Plane Extraction in Urban Environments for 3D City Modelling Using Remote Sensing Data

Carlos Campoverde [1], Mila Koeva [1,*], Claudio Persello [1], Konstantin Maslov [1], Weiqin Jiao [1] and Dessilava Petrova-Antonova [2]

[1] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7521 AN Enschede, The Netherlands; campoverde40006@alumni.itc.nl (CC); m.n.koeva@utwente.nl (MK); c.persello@utwente.nl (CP); k.a.maslov@utwente.nl (KM); w.jiao@utwente.nl (WJ.)

[2] GATE Institute, Sofia University "St. Kliment Ohridski", 1113 Sofia, Bulgaria; dessislava.petrova@gate-ai.eu (D.P.-A.)

* Correspondence: m.n.koeva@utwente.nl; Tel.: +31-534874410

Abstract: Mapping building roof plane structures is an active research direction in urban-related studies, as understanding roof structure provides essential information for generating highly detailed 3D building models. Traditional deep learning models have been the main focus of most recent research endeavors aiming to extract pixel-based building roof plane areas from remote sensing imagery. However, significant challenges arise, such as delineating complex roof boundaries and invisible boundaries. Additionally, challenges during the post-processing phase, where pixel-based building roof plane maps are vectorized, often result in polygons with irregular shapes. In order to address this issue, this study explores a state-of-the-art method for planar graph reconstruction applied to building roof plane extraction. We propose a framework for reconstructing regularized building roof plane structures using aerial imagery and cadastral information. Our framework employs a holistic edge classification architecture based on an attention-based neural network to detect corners and edges between them from aerial imagery. Our experiments focused on three distinct study areas characterized by different roof structure topologies: Stadsveld – 't Zwering neighborhood and Oude Markt area, located in Enschede, The Netherlands, and the Lozenets district in Sofia, Bulgaria. The outcomes of our experiments revealed that a model trained with a combined dataset of two different study areas demonstrated superior performance, capable of delineating edges obscured by shadows or canopy. Our experiment in the Oude Markt area resulted in building roof plane delineation with an F-score value of 0.43 when the model trained on the combined dataset was used. In comparison, the model trained only on the Stadsveld – 't Zwering dataset achieved an F-score value of 0.37, and the model trained only on the Lozenets dataset achieved an F-score value of 0.32. The results from the developed approach are promising and can be used for 3D city modelling in different urban settings.

Keywords: roof structure extraction; image processing; deep-learning; HEAT; 3D modelling; LOD2

1. Introduction

The rapid urban development and limited available land in urban areas have led to increased infrastructural developments above and below the ground surface [1]. The fundamental components of an urban area are buildings. The use of 3D building models in vector format is essential for creating accurate, interoperable, and efficient representations of urban environments. These models support a wide range of applications, from urban planning and design to disaster management and environmental analysis. Building structure mapping is a topic of ongoing study being done in various industries since understanding these features helps create detailed and realistic 3D city models [3]. Traditional 2D cadastral registration systems face challenges in capturing the evolving relationship between people and property in complex 3D urban environments [2]. 3D city models have emerged as a viable solution, offering a means to record and represent the various vertical developments within the 3D Geographic Information Systems (GIS) [4]. Furthermore, 3D city models find

applications in diverse fields, including urban planning, disaster management, energy efficiency, real estate, tourism, and serve as a cornerstone step toward realizing Urban Digital Twin [5].

3D city models could be used as platforms for analytical and simulated analyses, opening avenues to unveil emergent patterns and behaviors within urban landscapes [6]. In the process of 3D model reconstruction, defining the desired level of detail (LOD) becomes essential. The concept of "Levels of Detail" (LOD) in the City Geography Markup Language (CityGML) standards offers a hierarchical division of the geometric and semantic representation of objects in a 3D city model [7]. Five LODs are defined in the Open Geospatial Consortium's (2012) CityGML 2.0 standard. Although the concept is centered mainly on buildings, it is meant for numerous thematic objects; the five cases mentioned increase in geometric and semantic complexity [8,9]. Around the world, there are many examples of 3D city models for large areas; an inventory by Santhanavanich (2020) demonstrates several examples of datasets containing building models of large cities or even nations [10]. However, the generation of these 3D city models generally relies on LIDAR point clouds combined with building footprints to generate the roof structure which includes different roof planes [6]. However, the utilization of LIDAR is costly and may be challenging for many counties, particularly those in the path of development. Therefore, our study is focused on delineating building roof plane structures exclusively from aerial imagery to derive the geometry configuration of building rooftops, for LOD2 3D city models presenting [11], without recourse to LIDAR.

In addition, the complexities associated with urban features pose substantial challenges for designing automated end-to-end frameworks that span the entire range from building information retrieval and feature extraction using remote sensing data to complete 3D model reconstruction [12]. In this context, delineating building roof planes has become a central focus of recent research, mainly when a higher Level of Details (LoDs) of 3D city models is needed [13]. For this task, an essential step in 3D building modelling is roof plane segmentation, a process crucial for reconstructing 3D building models and delivering a more realistic view of the urban landscape [14].

Despite many research attempts to delineate building rooftop planes from remote sensing data, many of those attempts yield results in raster format, and limited exploration has been made in the automated extraction of roof plane structures in vector format [15]. However, vector formats represent geometric objects with mathematical precision, allowing for accurate representation of building shapes and dimensions. In addition, vector formats facilitate the integration of building models with other geospatial data layers and are generally more efficient in terms of storage and processing compared to raster formats. To have precise vector 3D building models including their complex roof structures is crucial in 3D city modeling, where the accuracy of building footprints and heights is important for various applications, such as urban planning and disaster management. In that vein, the feasibility of manual extraction is compromised for large-scale projects due to significant investments in time and cost [16]. Therefoe, machine learning and deep learning methods have emerged as promising solutions to address this challenge, enabling the automation of object delineation across diverse features, including buildings, roads, roofs, and land parcels, while ensuring efficient and accurate feature extraction [17]. However, the ability to mimic human-level perception for comprehensive geometric structures from images, especially in areas with complex topology or when a canopy or different obstacles obscure roof edges, remains a significant challenge in computer vision research [18].

Recent advancements are oriented towards achieving accurate 3D vector buildings in LOD2 with regularized otlines and straight edges [15]. Efforts are underway to explore end-to-end frameworks for accurate planar graph reconstruction of buildings [19]. Despite the diversity in input remote sensing datasets, the inherent challenge persists due to varying roof topology configurations in different study areas [20]. Primarily, this challenge arises in achieving the capability to perform holistic structural reasoning, such as graph reconstruction derived from corners and edges—a formidable task for end-to-end neural networks [21].

A persistent challenge for developing automated feature extraction frameworks lies in obtaining, processing, and preparing suitable datasets. Thus, the complexity of build-up areas may lead to inaccuracies in extracted features, introducing challenges such as occlusions, imprecise

 doi:10.20944/preprints202401.1839.v1

borders, and other issues [22], as highlighted in recent studies [23][24]. Understanding the configuration of building roof plane structures is paramount in developing detailed 3D models.

The presented research introduces a novel multi-stage framework for delineating and extracting roof plan structures from RGB images into a polygon vector format. The approach is using as a bases HEAT [21] to detect corners and edges on the RGB input samples. Once all features have been detected, a planar graph of the identified structure on the input RGB image is obtained, which derives the framework's next stage related to the planar graph's vectorization. One of the achievements is that the proposed framework manages to extract even the invisible rooflines located under the vegetation. This is followed by a subsequent 3D modelling stage that combines the vector planar roof structures with digital elevation models to generate a LOD2 3D model of the applied study area. We have evaluated our framework in three different study areas: Stadsveld – 't Zwering area and Oude Mark area, both areas located in Enschede, The Netherlands, and the neighborhood of Lozenets, Sofia, in Bulgaria. The qualitative and quantitative evaluations demonstrated that a model trained on a dataset from two different areas outperforms during the testing stage from all of the models trained in their specific areas.

The present study is organized as follows. Section 2 describes the study areas, presents the datasets employed in this study, and describes the two main stages utilized in our framework. Section 3 describes all the steps implemented in this study. Section 4 presents the quantitative and qualitative results obtained in our experiments. Following this, Section 5 presents a discussion of the main findings. Finally, Section 6 presents the conclusions of our study.

2. Materials and Methods

This section outlines the resources and methodology used in our study, offering an open framework for the entire research. The proposed framework aims to extract roof planes from RGB images to convert and combine those outputs for LOD2 3D modelling in two stages: (1) Roof plane extraction and (2) LOD2 3D modelling.

2.1. Study area

The study area of Stadsveld – 't Zwering, located in the southern urban area of Enschede, The Netherlands, covers an area of around 153 hectares. The dataset contains 1972, 123, and 370 building samples for training validation and testing, respectively. The extent and distribution of the building's samples are shown in Figure 1.

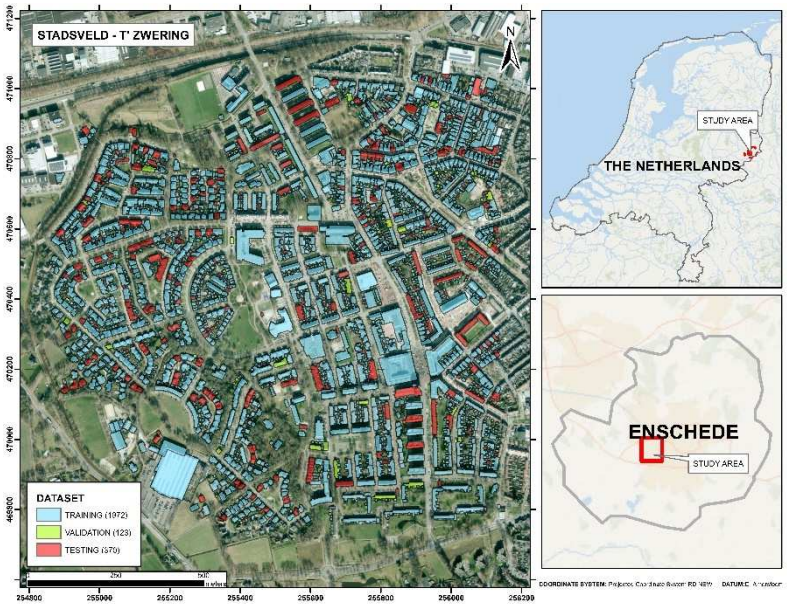

Figure 1. The building distribution for the Stadsveld – 't Zwering area, Enschede, The Netherlands.

 doi:10.20944/preprints202401.1839.v1

The study area of Oude Markt, located in the central urban area of Enschede, The Netherlands, covers an area of around 6 hectares. The dataset contains 119 building samples for testing the whole workflow. The extent and distribution of the building's samples are shown in Figure 2.

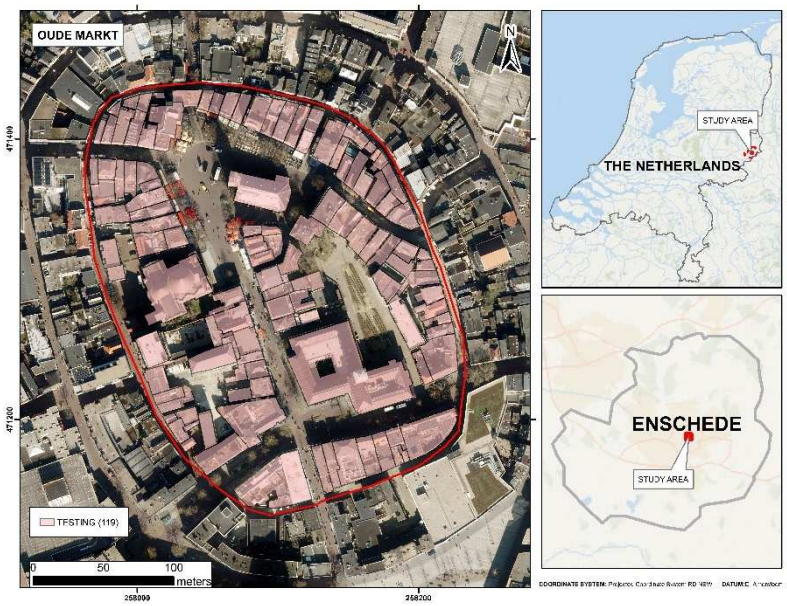

Figure 2. The building distribution for the Oude Markt, Enschede, The Netherlands.

The study area of Lozenets, located in the urban area of Sofia, Bulgaria, covers an area of around 812 hectares. The dataset contains 1440, 90, and 270 building samples for training, validation, and testing, respectively. The extent and distribution of the building's samples are shown in Figure 3.

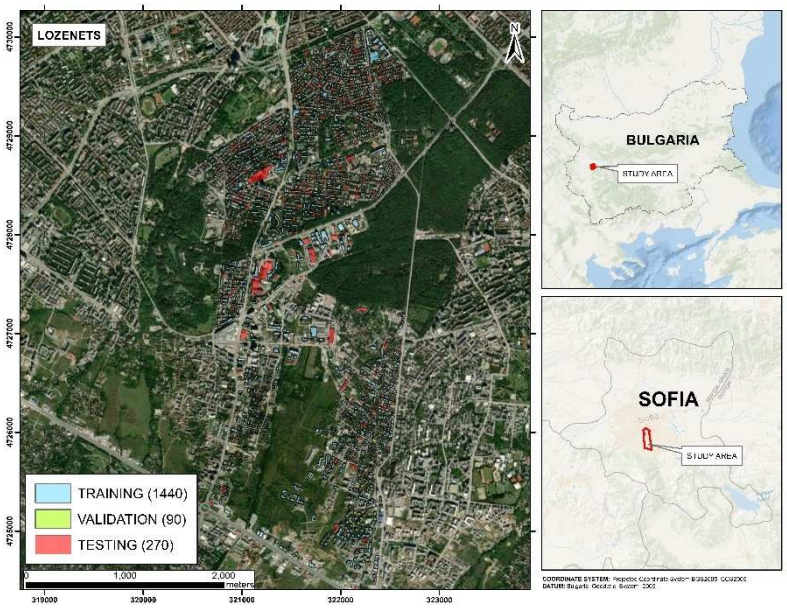

Figure 3. The building distribution for the Lozenets, Sofia, Bulgaria.

2.2. Data

Experiments were performed using the datasets of Enschede, The Netherlands, and Sofia, Bulgaria. The data used for this research is presented in Table 1 and includes: (1) VHR aerial images, (2) The building footprints, (3) The building inner roof planes of the buildings, and (4) the LIDAR point cloud derived for AHN4.

 doi:10.20944/preprints202401.1839.v1

For our experiments, aerial images define our study areas, from which all building samples for the roof plane extraction will be taken. The building footprint will be utilized to conduct a 2m buffer operation around each building to clip each building image sample based on the bounding box of the defined 2m buffer around each building. The inner roof planes of the buildings will be employed to generate planar graph information, serving as a reference to train/validate/test the HEAT model in the different study areas. From the LIDAR point cloud, the Digital Surface Model (DSM) and Digital Terrain Model (DTM) will be derived. Additionally, by subtracting the DTM from the DSM, the Normalized Digital Surface Model (nDSM) will be obtained. For these operations, ArcGIS Pro was used as the GIS software.

Table 1. Description of the dataset used in the presented research.

Area	Data	Source
Stadsveld – 't Zwering, Enschede, The Netherlands	RGB Orthophoto (8 cm)	PDOK [25], from aerial imagery, 2020
	Buildings inner roofs planes, polygon vector format	Digitalized by the author, 2023
	Buildings footprints, polygon vector format	PDOK [25], edited by the author, 2023
Oude Markt Enschede, The Netherlands	RGB Orthophoto (8 cm)	PDOK [25], from aerial imagery, 2020
	Buildings inner roofs planes, polygon vector format	Digitalized by the author, 2023
	Buildings footprints, polygon vector format	PDOK [25], edited by the author, 2023
	LIDAR, point cloud	AHN4 [26] (Point Cloud), 2020
Lozenets, Sofia, Bulgaria	RGB Orthophoto (10 cm)	GATE, from aerial imagery, 2020
	Buildings inner roofs planes, polygon vector format	Digitalized by RMSI, 2023
	Buildings footprints, polygon vector format	Digitalized by RMSI, 2023

*Big Data for Smart Society (GATE) Institute, Sofia, Bulgaria, supported the present work, providing access to high-quality datasets for research purposes. *RMSI is an external consulting company hired to digitize vector information in the study area of Lozenets, Sofia, Bulgaria, based on the aerial image provided by GATE.

2.3. Roof plane extraction

In the first stage, the proposed framework uses as a cornerstone the deep learning approach developed by Cheng, 2022 [21] HEAT (Holistic Edge Attention Transformer) developed for outdoor building reconstruction from satellite images and indoor floorplan reconstruction. This model is applied to the current roof-plane delineation context in aerial images. First, a data preparation process is initiated on the reference data on the study areas based on aerial images and the cadastral data of the buildings in the image. The automation of the data preparation process is to be used in different dataset sizes, and the dataset is split into training, validation, and testing subsets.

Subsequently, the HEAT models are trained using the different subsets of datasets. For the proposed framework, the availability of building footprints is assumed to be necessary for building sample creation and the requisite information for framework utilization.

 doi:10.20944/preprints202401.1839.v1

For the current research, the building footprints datasets were manually edited to match with the aerial imagery. Furthermore, the trained HEAT models were applied to extract the detected roof planes in the selected study areas. Finally, the obtained planar graphs were vectorized using specialized Python libraries to convert the planar graphs representations into a shapefile file. The overall workflow of the roof plane extraction is shown in Figure 4.

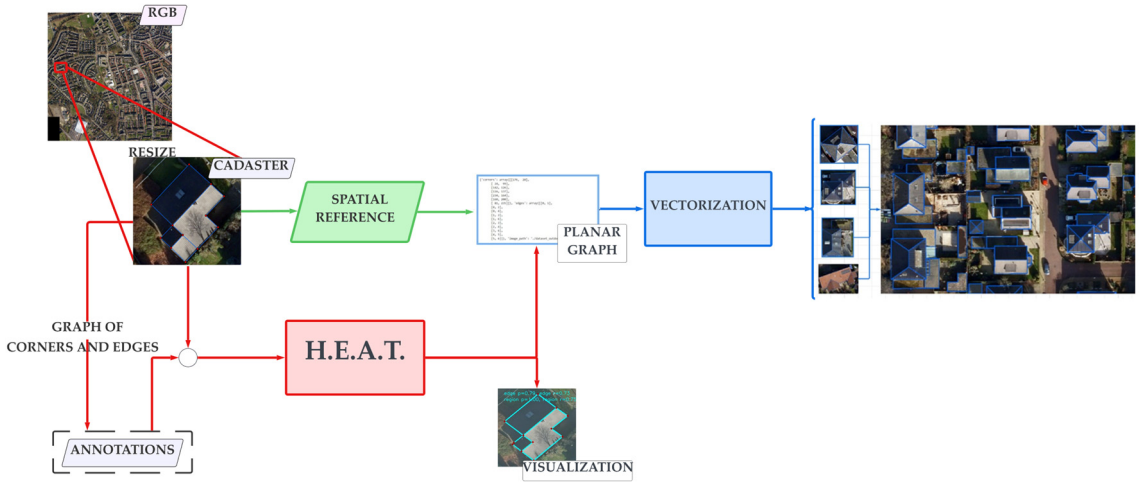

Figure 4. The workflow of the taken HEAT method adapted for rooftop plane building delineation from cadaster information and RGB data.

2.43. D Modelling

The aim of this 3D modellng process it to use the extracted vector roof output from the remotely sensed data and using the building footprints to reconstruct a full 3D building model in LOD2. This process aligns with the 3D city modelling methodology shown in the 3DBasemap extension of the commercial GIS software ArcGIS. The proposed methodology involves a multistep procedure that integrates various GIS tools to combine the inner roof planes, Digital Surface Model (DSM), Digital Terrain Model (DTM), and normalized Digital Surface Model (nDSM), and subsequently enables the creation of a 3D building objects at LOD2.

This phase underwent testing exclusively using the Oude Markt dataset. The overall workflow of the 3D modelling stage is shown in Figure 5. The figure illustrates the sequential steps involved in the GIS software.

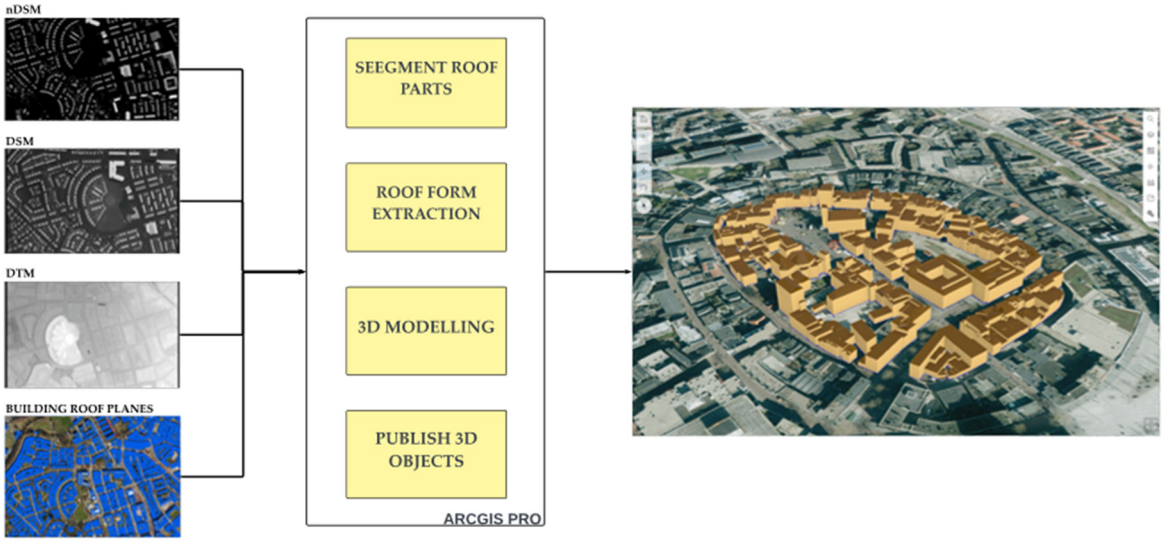

Figure 5. The workflow of 3D modelling using 3DBasemap extension in ArcGIS Pro is from the obtained rooftop planes and different digital elevation models.

 doi:10.20944/preprints202401.1839.v1

The accuracy of the results is evaluated using the metrics proposed in [21] for the extraction stage for detecting corners, edges, and regions. Furthermore, the vectorization stage is analyzed by using intersection over union metric. This additional accuracy metric captures the quality of the extracted vector format polygons, a factor not considered in the standard metrics presented by [21]. The Root Mean Square Error (RMSE) is calculated to measure the difference between the Digital Surface Model (DSM) and the resulting LOD2 3D city model derived from the inner roof planes of buildings. In this calculation, every pixel within the inner roof planes is considered. To maintain consistency with the resolution of the DSM, the LOD2 3D city model is rasterized to a 0.2-meter resolution.

3. Proposed Framework

The proposed framework includes innovative methods for extracting building roof planes from RGB photos, addressing the complexities of the urban environment for 3D modelling. The framework includes the following steps explained in the next subsections.

3.1. Data Preparation Implementation

The application of the HEAT approach is explored for mapping building rooftop structures. The approach infers a planar graph (comprising corners, edges, and regions) from a cropped 256 x 256 image. An automated data preparation process is implemented to apply the approach to an entire area, taking an aerial image and the building footprint and inner roof planes as input. The process involves cropping buildings from the aerial image using a bounding box derived from a 2-meter polygon buffer around each building footprint. Subsequently, the cropped images are resized to the 256x256 pixel format compatible with HEAT.

Furthermore, reference training data is essential to generate the training data for the HEAT model, which includes cropped images of the building samples and their associated planar graph. Building image samples are generated in .jpg format alongside the planar graphs for each building sample, considering the necessary resizing from real-world coordinates to the 256x256 image coordinates. Figure 6 shows the overall workflow of the data preparation process.

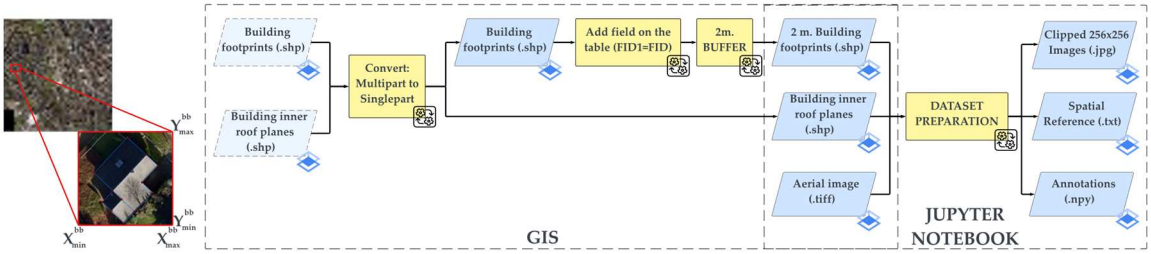

Figure 6. The overall workflow of the data preparation process.

The entire data preparation process is executed using a combination of GIS software tools and Python libraries comprised in a Jupyter notebook with Python 3.8. The information regarding the coordinate reference system of the used bounding box for each building sample necessary for resizing from real-world coordinates to image coordinates is stored in a text file per image building sample. This information is crucial for mapping the inner roof planar graph delineated on each building image sample to real-world coordinates.

3.2. Training HEAT model

The training process starts with the pre-trained HEAT model for outdoor architectural reconstruction, in which training parameters are outlined in [21]. Using the pre-trained HEAT model instead of training a new one is to leverage the existing understanding of outdoor architectural reconstruction. Three models underwent training based on the designated dataset for each study area: the model trained with the Stadsveld – 't Zwering dataset, the model trained with the Lozenets

 doi:10.20944/preprints202401.1839.v1

dataset, and the combined model was generated by combining the same training and validation datasets considered for the individual models. This approach helps prevent overfitting when testing the combined model on the respective areas.

Based on empirical experiments, an arbitrary number of 646 epochs is set for every training session. The training process is monitored by utilizing the validation accuracy value to find the model with the highest accuracy within the training session. The training was conducted using Python 3.8 and PyTorch 1.12.1. Table 2 shows how the input data for all three models is split on training and validation datasets, including image size, batch size, and maximum number of corners per image.

Table 2. Dataset parameters for training.

Model	Dataset size			Image size	Batch size	Max. number of corners per image
	Training	Validation	Total			
Model trained on the Stadsveld – 't Zwering, Enschede, The Netherlands dataset	1972	123	2095	256	16	150
Model trained on the Lozenets, Sofia, Bulgaria dataset	1440	90	1530			
Model trained on the combined dataset from Stadsveld – 't Zwering and Lozenets dataset	3412	213	3625			

In the following sections of this research, the model trained on the Stadsveld-'t Zwering dataset will be denoted as "Model trained on Enschede dataset", the model trained on the Lozenets dataset will be denoted as "Model trained on Sofia dataset", and the model trained on the combined dataset will be labeled as "Model trained on the combined dataset."

3.3. Building roof plane extraction

After completing the training process, this study focuses on applying and evaluating the performance of the generated trained models in delineating the inner roof planes of buildings within our designated testing datasets. Subsequently, post-processing operations are conducted to convert the obtained planar graph text file of each building sample into vector format datasets. The planar graph output obtained from HEAT is stored in a Python dictionary with three distinct keys, which are described as follows:

- **'corners'**: This key corresponds to a 2D array of integers, where each row represents the x and y coordinates of an identified corner in the building image sample.
- **'edges'**: This key corresponds to a 2D array of integers. Each row represents a pair of corners (indicated by their indices in the 'corners' array) forming an edge.
- **'image_path'**: This key corresponds to a string specifying an image file's path. This image aligns with the deduced corners and edges on the input-image building sample.

The following approach transforms the acquired planar graph text file into a real-world coordinate vector polygon format. The conversion process has a similar structure as in the data preparation process, as every planar graph based on the building file name is mapped to the

 doi:10.20944/preprints202401.1839.v1

corresponding bounding box information file used for clipping the image to its original size saved during the data preparation process. Once every sample was resized and georeferenced to its real-world location, all the generated building inner roof plane planar graphs were merged into one vector file. The overall workflow of the building roof plane extraction is shown in Figure 7.

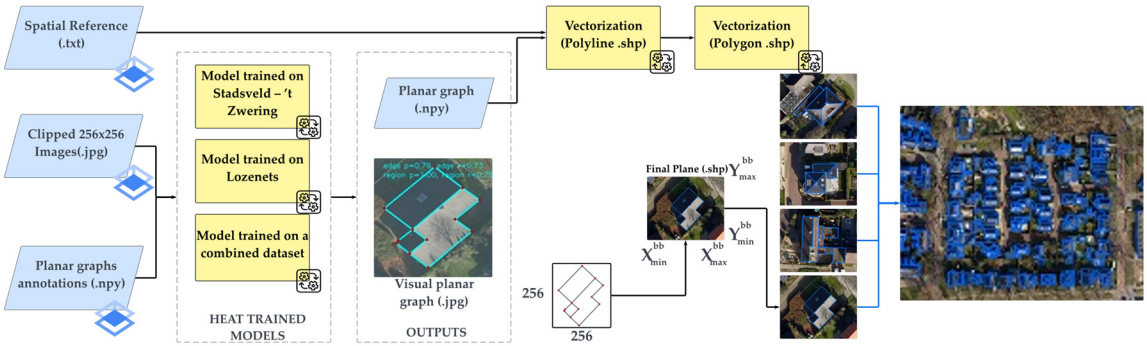

Figure 7. The overall workflow of the building roof plane extraction.

The process described above is performed in a Jupyter Notebook using Python 3.8. The outputs of the Jupyter Notebook are provided in polyline vector format. GIS software is employed to convert the geometry of these outputs to a polygon vector format.

3.43. D Modelling

After acquiring all the roof plane structures, the subsequent stage of the framework concentrates on testing the application of the obtained outputs for 3D modelling. A 3D city model at LOD2 for the Oude Markt area in Enschede, The Netherlands, is generated, utilizing the 3DBasemaps extension of ArcGIS Pro, whose methodology for combining roof form structures with DEMs is detailed in [27]. This approach was applied to generate a LOD2 3D city modelling for the Oude Markt area in Enschede, The Netherlands. This area was employed to test the whole framework without being part of the training stage for the roof structure extraction step. The overall workflow of this stage is shown in Figure 5.

To display the results of the created 3D model at LOD2 of the Oude Markt, a webmap was developed to provide a user-friendly interface for assessing the qualitative outcomes obtained in this stage. Users can access the platform following the resource: https://arcg.is/1raWvS0 [28].

3.5. Evaluation Metrics

The results are evaluated across the different stages: building inner roof plane delineation, vectorization, and 3D modelling.

Building inner roof plane delineation. To evaluate the corectness of the inner roof plane delineation results on every building sample, the trained models' performance are assessed on detecting corners, edges, and regions using the standard formulas for precision, recall, and F1 score.

Corners. A corner is successfully predicted and considered a true positive if a ground-truth corner is located within an Euclidean distance of an 8-pixel radius. In cases where multiple corners are detected around a single ground-truth corner, only the closest corner will be deemed a true positive.

Edges. An edge is successfully predicted and considered a true positive if both end corners are detected and the pair of corners exists on the ground-truth.

Regions. A region is successfully predicted and considered a true positive if the Intersection over Union (IoU) of a region defined by the different connected components of predicted corners and edges and a ground-truth region is greater than or equal to 0.7.

Vectorization. The accuracy of the final outputs is evaluated by the number of detected closed planes that resulted from edges. Since some edges do not converge into planes as they were delineated as unclosed structures. The IoU metric is employed. This metric compares the obtained

 doi:10.20944/preprints202401.1839.v1

polygon planes with the ground-truth vector planes, providing a measure of accuracy for the final results.

3D modelling. In the 3D modelling process, the Root Mean Square Error (RMSE) was calculated to assess the discrepancies between the generated 3D city model at LOD2 and the DSM. This involves a pixel-pixel computation, with the 3D city model at LOD2 rasterized to a 0.2-meter resolution.

4. Results

The proposed framework was developed to automatically extract building inner roof planes on a study area and generate a 3D model at LOD2. This section compares the results obtained on the test dataset for the different study areas. The processing outputs from the building roof plane extraction and the 3D modelling stage are presented in the following subsections.

4.1. Quantitative results

4.1.1. Building roof plane extraction

Table 3 presents the quantitative evaluations for delineating and extracting building inner roof planes. The models customized for Stadsveld -'t Zwering and Lozenets areas showed the best results throughout the entire workflow. Upon testing the trained model on the combined dataset within the Oude Markt study area, the model trained on this integrated dataset demonstrates superior performance, showcasing a noteworthy advantage, particularly during the final vectorization stage.

Our experiments were conducted using image samples with a resolution of 256 x 256. Our hypothesis suggests that the model trained on the combined dataset exhibits more advanced holistic geometric reasoning [21] than the other two models. Consequently, when tested in a different environment without prior context, the model trained on the combined dataset emerges as the superior performer due to its training on a more diverse dataset.

Table 3. Quantitative evaluations on building roof plane extraction. The values in bold mark the top results on our experiments.

Area	Models	Corners			Edges			Regions			Vectorization
		Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	IoU
Stadsveld -'t Zwering, Enschede, The Netherlands	Model trained on Enschede dataset	**0.85**	0.68	**0.76**	**0.61**	0.50	0.55	0.72	**0.64**	**0.68**	**0.82**
	Model trained on Sofia dataset	0.52	**0.72**	0.60	0.34	0.48	0.40	0.41	0.56	0.47	-
	Model trained on combined dataset	**0.85**	0.68	**0.76**	**0.61**	**0.51**	**0.56**	**0.73**	**0.64**	**0.68**	0.80
Oude Markt, Enschede, The Netherlands	Model trained on Enschede dataset	**0.69**	0.46	0.55	**0.38**	0.24	0.29	**0.49**	0.30	0.37	0.66
	Model trained	0.43	**0.64**	0.51	0.22	**0.34**	0.27	0.27	0.40	0.32	-

 doi:10.20944/preprints202401.1839.v1

	on Sofia dataset Model trained on combined dataset	0.60	0.55	**0.57**	0.31	0.29	**0.30**	0.44	**0.43**	**0.43**	**0.82**
Lozenets, Sofia, Bulgaria	Model trained on Enschede dataset	**0.84**	0.27	0.41	0.39	0.12	0.19	0.45	0.13	0.21	-
	Model trained on Sofia dataset	0.80	**0.53**	**0.63**	**0.44**	**0.31**	**0.37**	**0.47**	**0.37**	**0.41**	**0.71**
	Model trained on combined dataset (enschede + sofia)	0.81	0.50	0.62	**0.44**	0.30	0.36	**0.47**	0.35	**0.41**	0.70

It is observed that the IoU test for the vectorization stage was only conducted on the top two models from the delineation stage.

4.1.13. D City Modelling

Table 4 presents the quantitative evaluations for the 3D city modeling stage, relying on the extracted inner roof planes of buildings obtained in the preceding stage, utilizing the model trained on the combined dataset specific to the Oude Markt area. The RMSE values indicate that approximately 95% of the buildings within the study area exhibit discrepancies ranging from 0 to 10 meters between the generated LOD2 3D model and the DSM for this specific geographical region.

Table 4. Quantitative evaluations on 3D modelling stage.

Area		**RSME**					**Total**
		(0-5) m.	**(5-10) m.**	**(10-15) m.**	**(15-20) m.**	**(25-30) m.**	
Oude Markt, Enschede, The Netherlands	No. of buildings planes	473	164	25	8	2	672
	%	70.39	24.40	3.72	1.19	0.30	100.00

The 3D city modelling process utilized a pre-established method dictated by the software employed. Consequently, this stage offers room for improvement by exploring alternative options to enhance the 3D city model.

4.2. Qualitative results

4.2.1. Building roof Plane extraction

 doi:10.20944/preprints202401.1839.v1

Figure 8 provides a qualitative comparison between the top 2 trained models that outperform in Stadsveld – 't Zwering, Enschede, The Netherlands. As illustrated, the reconstruction quality is similar between the two models and close to the ground-truth. However, the reconstruction ability of the model trained on the combined dataset is notable for detecting even more building inner roof planes than the ones detected from the model trained on the Enschede dataset only, which certainly impacts the qualitative results.

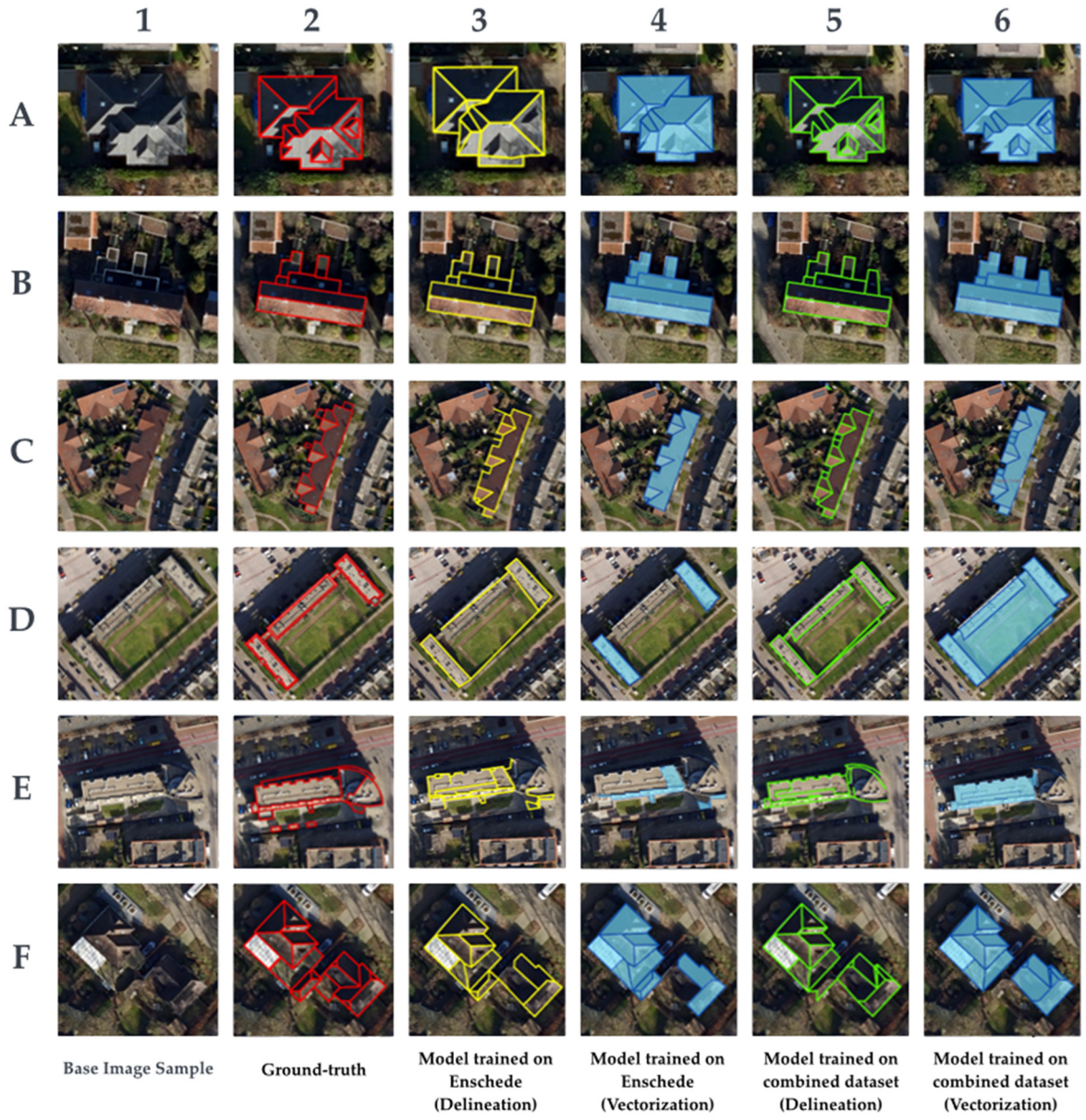

Figure 8. Qualitative evaluations on building roof plane extraction in the Stadsveld – 't Zwering, Enschede, The Netherlands dataset.

Figure 9 compares the top 2 trained models outperforming Oude Markt, Enschede, The Netherlands. As observed, the reconstruction quality is similar between the two models with the gound-truth facing the same challenges in the reconstruction of structures that are covered by shadows and in reconstructing structures with circular shapes (row E). The model trained on the combined dataset performs better than the model trained on the Stadsveld – 't Zwering dataset.

 doi:10.20944/preprints202401.1839.v1

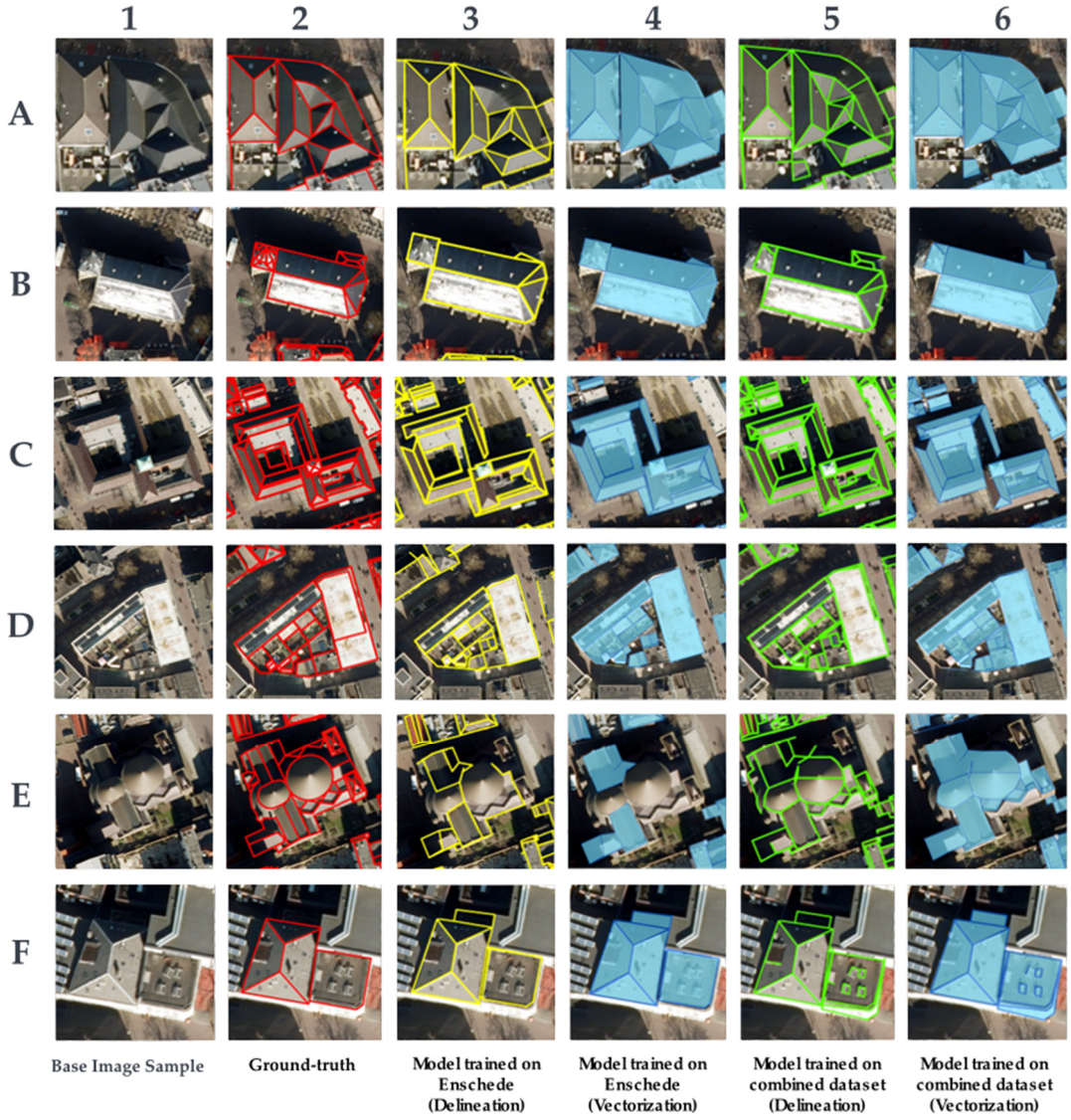

Figure 9. Qualitative evaluations on building roof plane extraction in the Oude Markt, Enschede, The Netherlands dataset.

Figure 10 compares the top 2 trained models outperforming in the Lozenets, Sofia, Bulgaria area. The reconstruction quality is similar between the two models with the gound-truth. The same traits were observed in the previous experiments, in which the model trained on the combined dataset could detect even more planes than those presented on the ground-truth. However, the qualitative evaluations show the ability of the model trained on the combined dataset to detect the finest details, including the inner planes of buildings that are occluded by the canopy of trees (row E).

 doi:10.20944/preprints202401.1839.v1

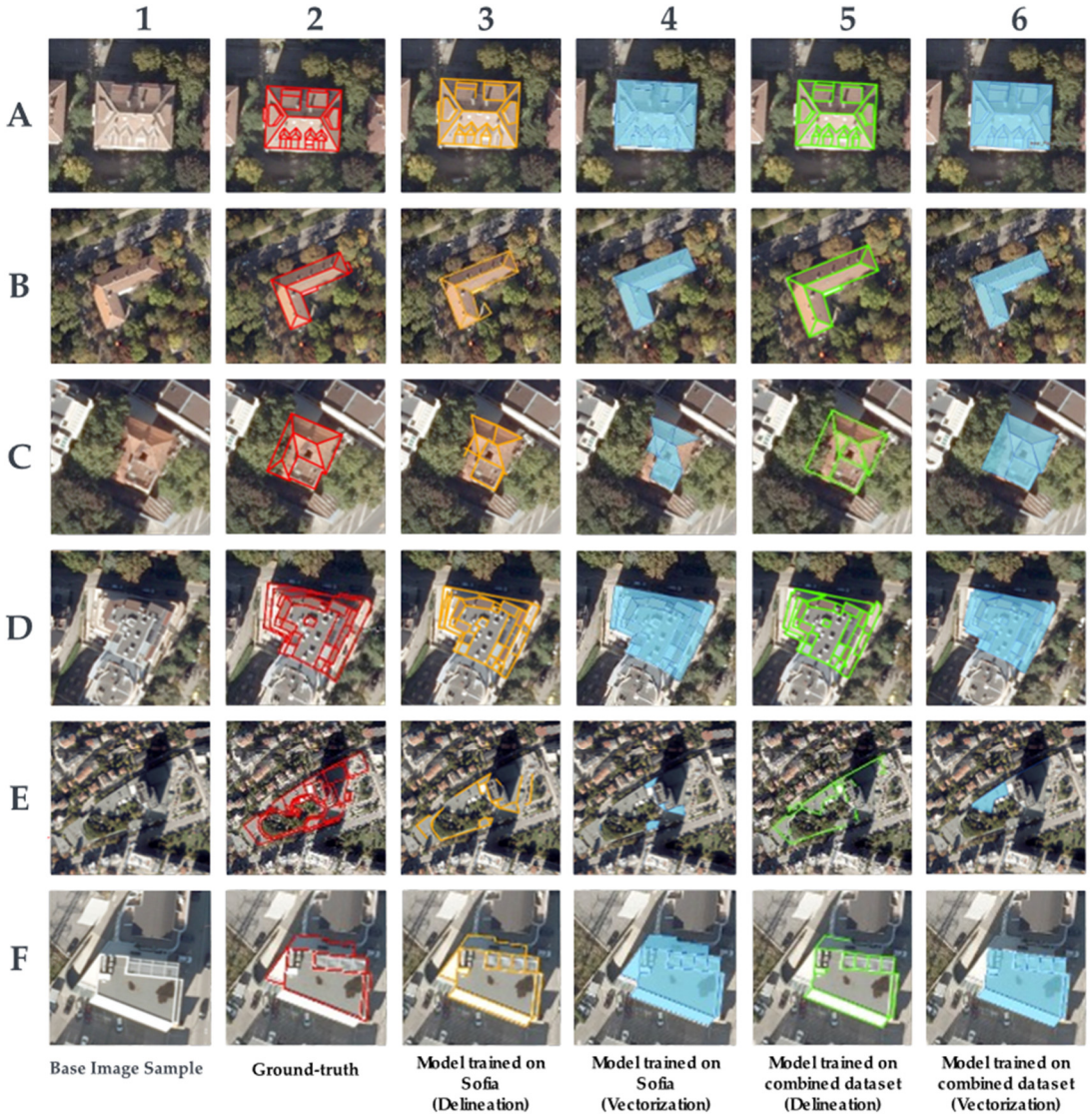

Figure 10. Qualitative evaluations on building roof plane extraction in the Lozenets, Sofia, Bulgaria dataset.

4.2.23. D Modelling

Figure 11 shows the generated 3D model at LOD2 for the Oude Markt area. A comparative analysis of selected buildings considered representative within the area is conducted, Contrasting the generated 3D structures with their counterparts from Google Earth Pro 3D buildings. Noteworthy are the discernible disparities in our model that require some rectification. Specific structures that were not recognized in the previous stage exhibit susceptibility to distortion in the 3D modelling phase.

 doi:10.20944/preprints202401.1839.v1

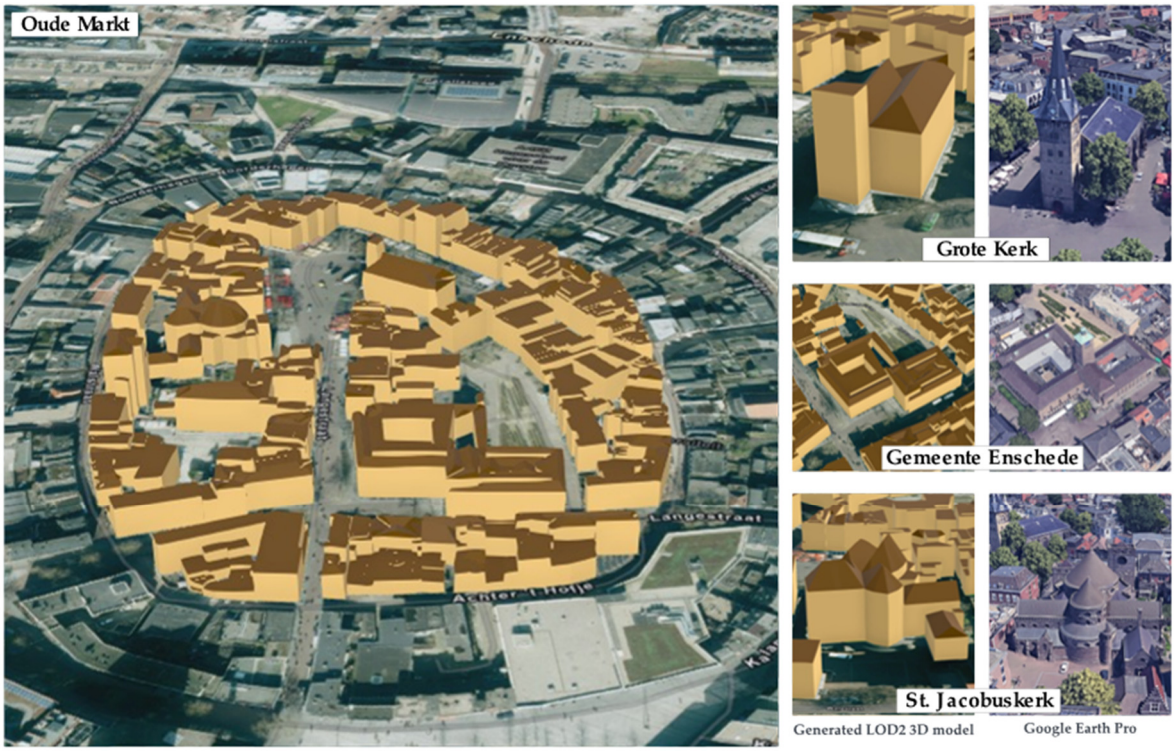

Figure 11. Qualitative evaluations of the generated LOD2 3D model of the Oude Markt. The assessment involves a comparative analysis of some representative 3D structures modeled against the 3D building structures from Google Earth Pro.

The generated LOD2 3D model of the Oude Markt at LOD2 and all the generated outputs on this research are published on a webmap ArcGIS online application [28].

5. Discussion

The present study introduced a novel framework to delineate building roof planes applicable across entire areas, utilizing aerial imagery and building footprint information using as a bases and upgrading the work developed by Cheng in 2022 [21]. In contrast to conventional segmentation methods for outdoor building reconstruction, which face challenges detecting straight edges [29], our proposed framework is specifically trained to identify corners, edges, and establish geometric relationships between these corners. Notably, the framework achieves remarkable results, demonstrated by its ability to detect building inner roof planes, even in complex scenarios where vegetation obscures edges. This capability is facilitated by accurately identifying the end corners of the building's inner roof planes and then, based on its holistic geometry reasoning, drawing the edges between them as demonstrated in sample C5 and C6 (Figure 10) of the model trained on the combined dataset for Lozenets, Sofia, Bulgaria. This achievement surpasses traditional methods employed in such cases.

Despite the robust performance exhibited by our approach in delineating inner roof plane structures, specific limitations become evident in certain scenarios. A constraint arises when image samples encompass more than one building in the same image with a ground surface between, as illustrated in samples D in Stadsveld – 't Zwering. Samples C in the Oude Markt (Figure 9), and samples E in the Lozenets (Figure 10). The trained models could predict the ground as a planar structure in such situations. This misclassification is attributed to the model's challenge in discerning between various types of flat surfaces, particularly distinguishing between the ground, vegetation, and the roofs of buildings. This difficulty intensifies when the ground is a substantial portion of the image and assumes a shape similar to that of the building. Misinterpreting the ground as part of the building roof structure could introduce inaccuracies into the final 3D model, as exemplified in image

 doi:10.20944/preprints202401.1839.v1

sample C in Oude Markt (Figure 9), representing "Gemeente Enschede," where the 3D model appears different from the representation in Google Earth's 3D model. Furthermore, these misinterpretations may result in significant errors that must be considered when assessing the model's performance.

An additional complication arises when the model endeavors to infer inner planes on image samples featuring intricate roof graphs characterized by many corners intended to generate circular or highly complex roof structures with many planes, as demonstrated in image building sample E in Stadsveld – 't Zwering and Oude Markt (Figure 9). The model encounters challenges in accurately interpreting and reconstructing the geometrical attributes inherent in these complex roof structures. Specifically, it frequently encounters difficulties in correctly identifying and processing planes characterized by numerous corners, resulting in the inaccurate generation of circular roof planes. As the complexity and number of corners within the roof graph increase, the model's performance tends to diminish, indicating a potential vulnerability in handling geometric intricacies.

In instances where the image samples comprise large, densely packed, and complex buildings, as illustrated in samples E in Stadsveld – 't Zwering and Lozenets (Figure 10), all trained models exhibited suboptimal performance in detecting planes. This limitation can be ascribed to the model's inherent difficulty in discerning between diverse structural elements within extensive, densely packed, and intricate building configurations within a limited pixelated image sample.

As shown by the observations in samples F6 in Oude Markt and row F in Lozenets (Figure 10), the output quality depends on the input data quality. In these cases, the trained model misinterpreted the building facades as part of the interior roof planes. This misconception is attributed to the building samples' tilt of the aerial picture, which allowed some facade planes to be visible in the image samples.

Quantitative evaluations reveal susceptibility to bias in some instances, such as samples A, B, D, E in Stadsveld – 't Zwering, and C Oude Markt (Figure 9), where the model trained on the combined dataset detects more planes than those presented in the ground-truth, resulting in false positives that impact the quantitative assessment. However, when evaluating qualitative metrics, this discrepancy demonstrates that the trained model on the combined dataset performance is superior, detecting finer details, thus providing an advantage for the modeling stage.

Despite the advancements demonstrated in this study, certain limitations persist within the proposed framework. While this research highlights the feasibility of generating a Level of Detail 2 (LOD2) 3D model, some enhancements may involve integrating additional factors. For instance, incorporating normalized Digital Surface Models (nDSM) as a four-band component in the building image sample is proposed to improve surface discrimination. Furthermore, future studies could benefit from experimenting with building image samples featuring a resolution of 512x512 pixels.

Finally, the current 3D city modelling methodology relies on pre-established tools within GIS software, and future investigations can extend and refine this presented framework by constructing a comprehensive end-to-end open-source framework.

6. Conclusions

This study introduces a multi-phase framework based on HEAT, a transformer-based deep learning model, to automatically extract building inner roof planes and then use them to generate 3D city models at LOD2. The proposed approach extracts inner and outer rooflines applicable to different roof topologies. The inner roof planes are obtained in vector format, without additional post-processing, addressing challenges in image segmentation methods.

Further experiments in different areas demonstrated sensitivity to bias, affecting model performance. Quantitative assessments reveal that models tailored to specific study areas successfully extracted inner roof plane structures with performance similar to a model trained on a combined dataset. However, a model trained on a combined dataset from both study areas demonstrated superior performance when tested on a study that was not included in the training process, such as the Oude Markt, Enschede study area. Nevertheless, limitations include topological inconsistencies requiring GIS post-processing. Qualitative assessments indicate superior performance of the model trained on the combined dataset, excelling in detecting roof boundaries

even when vegetation obscures them. This achievement is significant because it allows for the prediction of invisible boundaries using the HEAT capability for structural reasoning in an integrated manner. Furthermore, the model's ability to generate straight edges in the outputs represents a noteworthy success. Considering these accomplishments is crucial, as they address common challenges that classical deep learning methods based on image segmentation often struggle with.

The developed framework successfully extracts inner roof planes, however, there are still inconsistencies including incomplete corner prediction in complex roof structures, requiring additional post-processing. The study demonstrates the feasibility of creating LOD2 3D city models by integrating the generated inner roof planes with DSM, DTM, and nDSM. While improvements can be added to the framework, the approach confirms the viability of combining remote sensing, GIS, and deep learning for urban mapping and 3D city modelling, making it a cornerstone for future research and growth.

Author Contributions: Conceptualization, C.C., M.K., C.P., D.P-A; methodology, C.C., M.K., C.P., D.P-A, K.M., W.J.; formal analysis, C.C., M.K., C.P., D.P-A, K.M., W.J.; writing—original draft preparation C.C., M.K., C.P.,, final version review C.C., M.K., C.P., D.P-A, K.M., W.J. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this research are openly available through the references [30][31][32].

Acknowledgments: The dataset was provided by the GATE, The Big Data for Smart Society Institute, in Sofia, Bulgaria, and RMSI, an external consulting company hired to digitize vector information in the study area of Lozenets, Sofia, Bulgaria, with funding provided by ITC.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shojaei, D.; Olfat, H.; Rajabifard, A.; Briffa, M. Design and Development of a 3D Digital Cadastre Visualization Prototype. *ISPRS Int. J. Geo-Information* **2018**, *7*, doi:10.3390/ijgi7100384.
2. Van Oosterom, P. Research and Development in 3D Cadastres. *Comput. Environ. Urban Syst.* 2013, *40*, 1–6.
3. Rau, J.Y.; Cheng, C.K. A Cost-Effective Strategy for Multi-Scale Photo-Realistic Building Modeling and Web-Based 3-D GIS Applications in Real Estate. *Comput. Environ. Urban Syst.* **2013**, *38*, 35–44, doi:10.1016/j.compenvurbsys.2012.10.006.
4. Hajji, R.; Yaagoubi, R.; Meliana, I.; Laafou, I.; Gholabzouri, A. El Development of an Integrated BIM-3D GIS Approach for 3D Cadastre in Morocco. *ISPRS Int. J. Geo-Information* **2021**, *10*, doi:10.3390/ijgi10050351.
5. Dimitrov, H.; Petrova-Antonova, D. 3D City model as a First Step towards Digital Twin of Sofia City. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **2021**, *43*, 23–30, doi:10.5194/isprs-archives-XLIII-B4-2021-23-2021.
6. Peters, R.; Dukai, B.; Vitalis, S.; van Liempt, J.; Stoter, J. Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands. *Photogramm. Eng. Remote Sensing* **2022**, *88*, 165–170, doi:10.14358/PERS.21-00032R2.
7. Kolbe, T.H.; Gröger, G.; Plümer, L. CityGML: Interoperable Access to 3D City Models. *Proc. Int. Symp. Geo-information Disaster Manag.* **2005**, 883–899, doi:10.1007/3-540-27468-5_63.
8. Macay Moreia, J.M.; Nex, F.; Agugiaro, G.; Remondino, F.; Lim, N.J. From Dsm To 3D Building Models: A Quantitative Evaluation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-1/W1*, 213–219, doi:10.5194/isprsarchives-xl-1-w1-213-2013.
9. Biljecki, F.; Ledoux, H.; Stoter, J. An Improved LOD Specification for 3D Building Models. *Comput. Environ. Urban Syst.* **2016**, *59*, 25–37, doi:10.1016/j.compenvurbsys.2016.04.005.

10. Santhanavanich, J. Open-Source CityGML 3D Semantical Building Models: A Complete List of Open-Source 3D City Models Available online: https://towardsdatascience.com/open-source-3d-semantical-building-models-in-2020-f47c91f6cd97 (accessed on 18 December 2023).
11. Lee, J.; Zlatanova, S.; Gartner, G.; Meng, L.; Peterson, M.P. 3D Geo-Information Sciences. In *Lecture Notes in Geoinformation and Cartography*; Lee, J., Zlatanova, S., Eds.; Springer-Verlag Berlin Heidelberg 2009: Delft, 2009; pp. 79–96 ISBN 9783540873945.
12. Soilán, M.; Truong-Hong, L.; Riveiro, B.; Laefer, D. Automatic Extraction of Road Features in Urban Environments Using Dense ALS Data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 226–236, doi:10.1016/j.jag.2017.09.010.
13. Sun, X. Deep Learning-Based Building Extraction Using Aerial Images and Digital Surface Models, University of Twente, 2021.
14. Huang, J.; Stoter, J.; Peters, R.; Nan, L. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. *Remote Sens.* **2022**, *14*, 1–18, doi:10.3390/rs14092254.
15. Zhao, W.; Persello, C.; Stein, A. Extracting Planar Roof Structures from Very High Resolution Images Using Graph Neural Networks. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 34–45, doi:10.1016/j.isprsjprs.2022.02.022.
16. Ok, A.O. Automated Detection of Buildings from Single VHR Multispectral Images Using Shadow Information and Graph Cuts. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 21–40, doi:10.1016/j.isprsjprs.2013.09.004.
17. Qin, Y.; Wu, Y.; Li, B.; Gao, S.; Liu, M.; Zhan, Y. Semantic Segmentation of Building Roof in Dense Urban Environment with Deep Convolutional Neural Network: A Case Study Using GF2 VHR Imagery in China. *Sensors (Switzerland)* **2019**, *19*, 1–12, doi:10.3390/s19051164.
18. Liu, K.; Ma, H.; Ma, H.; Cai, Z.; Zhang, L. Building Extraction from Airborne LiDAR Data Based on Min-Cut and Improved Post-Processing. *Remote Sens.* **2020**, *12*, 2849, doi:10.3390/rs12172849.
19. Rezaei, Z.; Vahidnia, M.H.; Aghamohammadi, H.; Azizi, Z.; Behzadi, S. Digital Twins and 3D Information Modeling in a Smart City for Traffic Controlling : A Review. *J. Geogr. Cartogr.* **2023**, *6*, 1–27, doi:10.24294/jgc.v6i1.1865.
20. Zhao, W.; Persello, C.; Stein, A. Building Outline Delineation: From Aerial Images to Polygons with an Improved End-to-End Learning Framework. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 119–131, doi:10.1016/j.isprsjprs.2021.02.014.
21. Chen, J.; Qian, Y.; Furukawa, Y. HEAT: Holistic Edge Attention Transformer for Structured Reconstruction. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2022**, *2022-June*, 3856–3865, doi:10.1109/CVPR52688.2022.00384.
22. Golnia, M. Building Outline Delineation and Roofline Extraction : A Deep Learning Approach, University of Twente, 2021.
23. Tarabalka, N.G. and Y. END-TO-END LEARNING OF POLYGONS FOR REMOTE SENSING IMAGE CLASSIFICATION Nicolas Girard and Yuliya Tarabalka Universit' e Cˆ Ote d' Azur , Inria , TITANE Team , France Email : Nicolas.Girard@inria.Fr. **2018**, 2087–2090.
24. Marcos, D.; Tuia, D.; Kellenberger, B.; Zhang, L.; Bai, M.; Liao, R.; Urtasun, R. Learning Deep Structured Active Contours End-to-End. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; IEEE, June 2018; pp. 8877–8885.
25. PDOK (the Public Services On the Map). Available online: https://www.pdok.nl/ (accessed on 5 December 2022).
26. AHN Viewer Available online: https://ahn.arcgisonline.nl/ahnviewer/ (accessed on 29 April 2023).
27. Use 3D Basemaps Available online: https://doc.arcgis.com/en/arcgis-solutions/10.9.1/reference/use-3d-basemaps.htm (accessed on 22 November 2023).
28. AUTOMATIC BUILDING ROOF PLANE STRUCTURE EXTRACTION FROM REMOTE SENSING DATA Available online: https://www.arcgis.com/home/webscene/viewer.html?webscene=b09bd9fcb9ec4d39a85f9d672776b06e&viewpoint=cam:6.89874543,52.21450794,543.283;349.682,52.253 (accessed on 27 November 2023).
29. Hossain, M.D.; Chen, D. A Hybrid Image Segmentation Method for Building Extraction from High-Resolution RGB Images. *ISPRS J. Photogramm. Remote Sens.* **2022**, *192*, 299–314, doi:10.1016/j.isprsjprs.2022.08.024.

 doi:10.20944/preprints202401.1839.v1

30. Campoverde, C. Carecamp93/Automatic_Roof_Plane_Extraction Available online: https://github.com/carecamp93/Automatic_Roof_Plane_Extraction.
31. Campoverde, C. 1.- ROOF EXTRACTION Available online: https://drive.google.com/drive/folders/1ZDmQDv58faQrKPdYRurABFxSAF1o398h?usp=sharing.
32. Campoverde, C. 2.- 3D MODELLING Available online: https://drive.google.com/drive/folders/1C0qwlgx6gXsfIcFQd_x9gPT2yih6e-jf?usp=sharing.