ZANCO Journal of Pure and Applied Sciences

TYPE Original article
DOI http://dx.doi.org/10.21271/ZJPAS.37.2.10

OPEN ACCESS

*Corresponding author

Rojgar Qarani Ismael
rojgar.ismael@su.edu.krd

RECEIVED :22 /11 /2024
ACCEPTED :17/01/ 2025
PUBLISHED :30/ 04/ 2025

KEYWORDS:
LoD2, Deep Learning, DSM, Model-Driven

LoD2 Building Reconstruction from Stereo Satellite Imagery using Deep Learning and Model-Driven Approach

Rojgar Qarani Ismael[1*], Haval Abduljabbar Sadeq[1]

1 Department of Geomatics (Surveying) Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.

ABSTRACT

This study presents a Level of Detail 2 building reconstruction approach for open and occluded areas from stereo-satellite imagery. The approach combines deep learning techniques, and digital surface models with model-driven methodology. The best performance of deep learning algorithms (U-Net, FCN, and Mask R-CNN) for building boundary segmentation was selected and then integrated with model-driven technique for the purpose of accurate geometric building fitting employing digital surface model (DSM) generated by semi global matching. The Reconstructed model was refined by utilizing OpenStreetMap library and graph cut optimization method. The suggested methodology is tested on the GeoEye-1 satellite imagery dataset for Erbil City, which is validated with ground truth data. The proposed algorithm presented promising results, it is shown that the model can predict building heights for ridge and eave to a mean absolute error of 0.70 m, and in the occluded area was approximately 1.0 m. Meanwhile, the computed root mean square error are shown to be within 0.9 m for the ridge and eave, which is essentially small. While for occluded area it was approximately 1.2 m and 0.8 m for ridge and eave heights, respectively. This indicates that the predicted values are close to real values. Furthermore, most of the building's roofs were correctly classified in both open and occluded areas. These findings underline the effectiveness of the model-driven deep learning approach in producing reliable and accurate LoD2 building reconstructions, a precondition for detailed urban analysis and 3D city modeling.

1.Introduction

Level of Detail 2 (LoD2) in building reconstruction provides detailed 3D models containing an accurate building representation with a well-differentiated roof structure and other architectural components. LoD2 proved its capability for various applications such as 3D city modeling (Wysocki et al., 2024); urban planning and environmental monitoring (Peters et al., 2022); and disaster management (Dukai et al., 2019). This LoD2 becomes more important particularly when cities continuously growing and expanding; consequently, the demand for precise and updated 3D models brings importance in various fields of simulation even in the areas of resource management and in the assessment of environmental impacts of newly built up structures in an urban environment (Nys et al., 2020). With the advancement in satellite imagery technology has led to the ability to acquire stereo satellite imageries for any area thus playing an important role in 3D building reconstruction (Stucker and Schindler, 2022). This enables depth perception and can acquire elevation data, which might be used when producing 3-dimensional models (Duan and Lafarge, 2016). Besides, the digital surface models (DSM) that generated from stereo satellite images facilitate a highly detailed representation of the surface of the Earth, such as buildings, vegetation, and other surface structures. More importantly, when stereo satellite imagery is further processed with some sophisticated CNN algorithms, building footprints and DSMs turn into a very powerful tool for reconstructing an urban environment at LoD2 (Bittner et al., 2018a).

Conventional LoD2 building reconstructing approaches are mainly based on manual digitizing from photogrammetry and LiDAR data. Although they are regarded as accurate; it is marked as time-consuming and costly (Zhang et al., 2020). Furthermore, these approaches will also face difficulties in handling occlusions, complex roof geometries, or variations in building materials, which might result in incomplete or inaccurate models (dos Santos et al., 2020).

The deep learning technique is designed to detect and recognize patterns and features from satellite images, which enables the automatic extraction of building geometries. By its incorporation of domain-specific knowledge and predefined rules in a model-driven approach, deep learning by itself was capable of handling many of the weaknesses of conventional methods, especially in complicated and occluded structures (Huang et al., 2020, Ps and Aithal, 2023).

On the other hand, Model-driven approaches rely on libraries of parameterized models to find the most suitable roof models for DSM and point clouds. These methods often incorporate additional data, such as building footprints, to separate the roofs into simpler components. Additionally, building footprints assist in localizing the region of interest in images and point clouds, thus minimizing the search area and concentrating on building-specific regions (Partovi et al., 2019). Parametric models describe roof primitives in libraries, which are created by using a few parameters: coordinate origin (Xθ, Yθ), orientation (θ), slope of roof plane (α), length (L), width (W), heights of eave (Heave) and ridge lines (Hridge)), as shown in Figure 1.

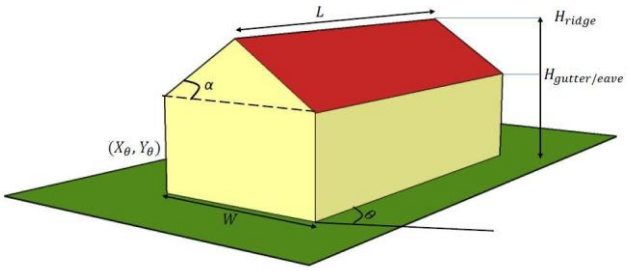


Figure 1: Parametric gable roof model illustrates the related parameters (Partovi et al., 2019).

This paper focuses on the proposing an approach for a model-driven deep learning for LoD2 building reconstruction from stereo satellite images and DSMs. Specifically, it aims to: increase LoD2 building model accuracy by deep learning combined with model-driven approach; solve problems of occlusions in urban environments; and make a scalable solution applicable for large urban areas. Through these objectives, the study shall provide a more efficient method for LoD2 building reconstruction that ensures better accuracy in supporting the field of 3D city modeling and urban planning.

The main contributions of this paper toward 3D building reconstruction from satellite imagery are as follows: First, it presents a unique approach that integrates deep learning and model-driven methods to provide more accurate and reliable LoD2 building models. Second, by incorporating domain-specific knowledge into the deep learning pipeline, this paper overcomes common challenges in urban reconstruction, including occlusions and a variety of types of buildings.

The paper is organized as follows: Section 2 provides a critical review of literature related to advances in LoD2 building reconstruction that have been made using deep learning and model-driven approaches. Section 3 describes the methodology adopted, covering the proposed model architecture and the integration process. Section 4 presents the experimental results. Section 5 presents a discussion of analysis, challenges and limitations, while Section 6 concludes the study, suggesting some future directions.

2. Related Works:

Reconstruction of the 3D building model up to LoD2 from satellite data is considered to be a challenging task, mainly due to its lower spatial resolution compared to aerial images and LiDAR (Weng, 2018). Therefore, detecting and modeling buildings accurately from satellite imagery is a very difficult task, especially in dense areas where buildings are very close to each other and their details are barely distinct (Xu et al., 2020).

In the field of LoD2 building model reconstruction, three major approaches are commonly applied: data-driven, model-driven, and hybrid approaches. Data-driven approaches extract geometrical features from DSMs or point clouds. Since the quality and resolution of satellite-derived data are usually lower, it is challenging to handle them. These methods are very flexible but less accurate if the quality of the data is not high enough to capture fine details, such as in complex urban environments (Gui et al., 2022, Lai and Yang, 2020). On the other hand, model-driven approaches' employing predefined building models fits to the satellite data. These methods are capable of producing very structured and consistent models, but they typically lack the flexibility to accommodate the irregularities and complexities of real urban settings (Wang et al., 2021). Hybrid approach seeking to combine the strengths of data-driven and model-driven strategies, aims to enhance the accuracy and robustness of the reconstruction model. Hybrid approaches are therefore a more balanced solution to these issues raised during LoD-2 reconstruction, using data-driven methods for better adaptability and model-driven for better consistency of structures (Buyukdemircioglu et al., 2022).

Building detection and segmentation are the initial steps in LoD-2 building model reconstruction. Deep learning techniques, such as U-Net (Ronneberger et al., 2015), FCN (Long et al., 2015), and Mask R-CNN (He et al., 2017), have already been widely used for this purpose. Particularly, Semantic-based algorithms, such as, U-Net and FCN have been explicitly proved to exhibit very good performance in detail preservation during segmentation, which is plays an important role in preservation of building outlines within the final model (Bittner et al., 2018b, Wagner et al., 2020). Mask R-CNN is known to be very powerful in instance segmentation and will present an opportunity to delineate the exact boundaries of individual buildings (Amo-Boateng et al., 2022, Noureldeen and Wahed, 2024, Zhao et al., 2018), even in densely populated urban areas (Han et al., 2022).

After building detection and segmentation, the extraction of the building polygons will be performed. The Douglas–Peucker algorithm (Douglas and Peucker, 1973), and Line Segment Detector (LSD) by (Grompone von Gioi et al., 2010) are commonly utilized for simplification and regularization of obtained polygons, thus ensuring that they do restore the original outline of buildings without additional and excessive complications (Dorninger and Pfeifer, 2008). A grid-based decomposition approach is commonly proposed method for decomposing the building shape into simpler rectangles (Sugihara et al., 2015). Rectangle based approach by (Partovi et al., 2019) allowing the easy fitting of regular building models and increasing the accuracy in the reconstruction process.

The next process involve the 3D model fitting which consistes of basic building models, such as flat, gable, and hip roofs, to the extracted building polygons using DSM data (Gui et al., 2022, Huang et al., 2022, Muftah et al., 2022). This step is important for generate LoD-2 model that is accurate enough to present the actual architectural styles found in urban environments (Peters et al., 2022). In post-refining processes, consistency among the reconstructed models should be enforced so that similar building types become uniformly treated (Alidoost et al., 2019, Gao et al., 2024). Additional refinement is required when models of individual buildings must be combined into larger, more complex structures (Dukai et al., 2021).

3. Methodology

The implemented workflow in this study involves three main stages: data collection and preprocessing, segmentation of building boundaries, and the reconstruction of LoD2 buildings. The following steps provide illustration related to each step individually in details as shown in Figure 2.

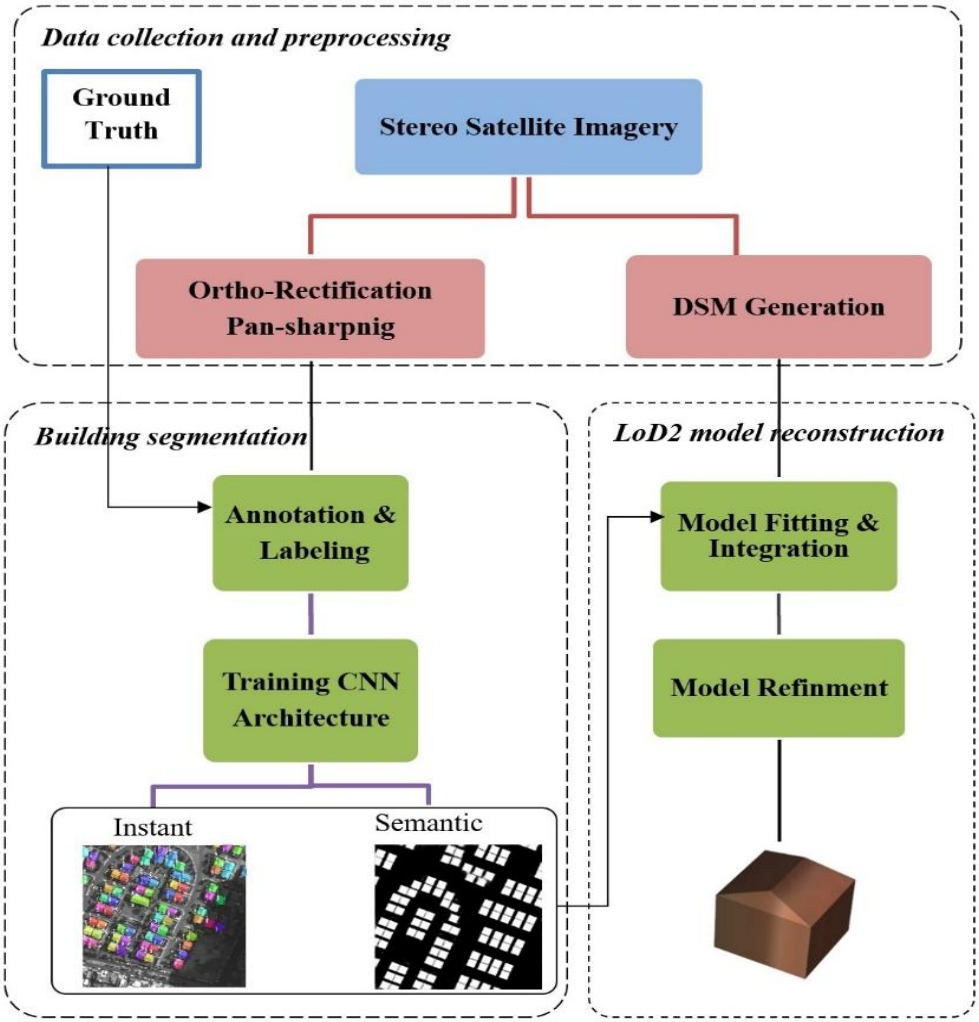
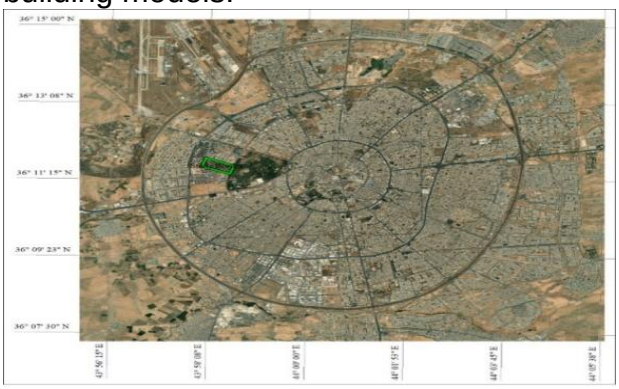


Figure 2: Workflow of the study

3.1. Data Collection and Preprocessing

The stereo satellite images used in the research were dated November 8th, 2020, by the GeoEye-1 satellite in a north-western part of Erbil city, which is the capital of Kurdistan, Figure 3. GeoEye-1 is among the first satellites to take very high-resolution images of Earth imaging. Which provides panchromatic images with a Ground Sampling Distance (GSD) of 0.5 meters and multispectral images with a GSD of 2.0 meters. For instance, high spatial resolution

images is essential for detailed urban analysis; particularly in tasks like building reconstruction, where fine details of the structure are required to be captured accurately (Proulx-Bourque et al., 2019). The panchromatic images deliver detailed grayscale information that aids in edge detection and the delineation of fine structures, while multispectral images provide relevant spectral information that enables differentiation between different materials and surfaces. A combination of both kinds of data produces pan-sharpening, enhancing the robustness of the reconstruction process by allowing more accurate identification and modeling of characteristics referring to urban features (Amro et al., 2011). For the deep learning training, aground truth data for the study was prepared by VOSSING Company in 2012. The important point while choosing this type of dataset, is there has been no change in this area after its establishment. This dataset provides reliable reference information for the validation of building boundary extraction and subsequent 3D building models.



(a)

(b)

Figure 3: (a) Erbil city, (b) study area that has been identified in the research.

3.1.1. DSM Generation

DSM is a major input of LoD2 reconstruction, providing the elevations of the earth's surface and man-made structures above the ground. In this study, a digital surface model is generated using the Semi-Global Matching (SGM) algorithm, which is one of the common methods for generating high-quality DSMs from stereo imagery (Zhang et al., 2017). SGM works by matching pixels between the stereo pair using known sensor geometry which is reperesnted by Rational Polynomial Coefficients (RPC) to estimate a 3D representation of the surface. The RPC model is accompanied by satellite images, which consider the orbit parameters and the characteristics of its sensors, enabling accurate geometric matching of stereo pairs (Akiki et al., 2021). The DSM was produced in Catalyst Earth, a software commonly used due to its high functionality in remote sensing and photogrammetry, as shown in Figure 4. It has advanced tools for DSM generation, which controls the SGM parameters as well as refine the DSM through the availability of post-processing options (Lv et al., 2022). Noise reduction techniques such as Gaussian smoothing, and median filtering have been applied to improve the quality of the Digital Surface Model and the 3D reconstruction. Good noise reduction has to be performed to retain the integrity of building boundaries, and other fine architectural details are required for LoD2 reconstruction.

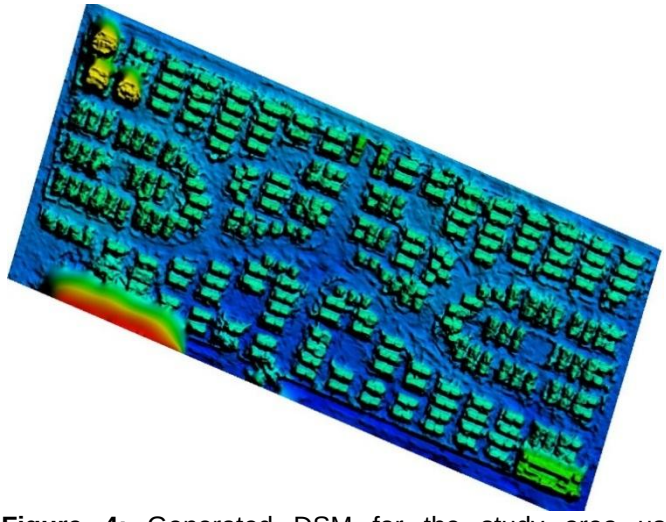


Figure 4: Generated DSM for the study area using Catalyst Earth

3.1.2. Ortho-rectification

Ortho-rectification is one of the preprocessing steps that is crucial to correct the geometric distortions in satellite images, which systematically aligns them to geo-referenced coordinates (Li et al., 2020). In this study, ortho-rectification was performed using a provided RPC model with GeoEye-1 imagery utilized Catalyst Earth (PCI Geomatica) software. This model takes into account the sensor geometry of the satellite while correcting the distortions to provide high-accuracy alignment of the acquired imagery to a common map projection.

3.2. Segmentation of Building Boundaries

For the training purposes, the image dataset was resized to 512 × 512 pixels for consistency during training. Precise annotation or labeling of the image data during the dataset preparation process is essential and it is very important for identifying the boundaries of buildings properly. These annotated process is considered the base for training and validation of the FCN, U-Net, and Mask-RCNN models, which also can be regarded as a form for the core of the segmentation tasks. The annotation was done through the Roboflow framework (Alin et al., 2023) and the VGG Image Annotator (Dutta and Zisserman, 2019), both software are well known for their efficiency in getting an accurate annotation. On the other hand, the geometric transformations and color adjustments which are part of data augmentation techniques were also applied to the dataset to artificially increase the size and variety of training data. The data augmentation that was applied in this research was performed using Python's "imgaug" library (Amarù et al., 2023), which is a comprehensive suite of tools for the systematic variation of a dataset.

3.2.1. Deep Learning Model Architectures

In this study, the used deep learning models for the detection and segmentation of buildings from satellite images were consisted of U-Net, Fully Convolutional Networks (FCN), and Mask R-CNN. The proposed architectures for building boundary segmentation were implemented from scratch using the PyCharm IDE. The models were created to handle two types of input images: 512×512×3 (RGB images) and 512x512x1 (panchromatic images).

3.2.1.1 Semantic and Instance Segmentation

One of the reasons for selecting the U-Net in semantic segmentation is due to the process of the encoder-decoder architecture (Minaee et al., 2022, Ronneberger et al., 2015). This architecture enables it to model both high-level context and detailed spatial information. In LoD2 reconstruction, detailed outlining of buildings requires the preservation of fine details in the segmentation output, and in this regard, U-Net has been shown to achieve excellent performance (Alsabhan et al., 2022, Muftah et al., 2022). FCN is another commonly implemented deep neural network architecture for semantic segmentation tasks. Unlike U-Net, FCN where the fully connected layers are replaced with convolutional layers, making the network able to accept images of any size. This feature makes FCN in handling a wide range of building sizes and shapes which is found in urban settings. Mask R-CNN is an extension of Faster R-CNN that adds a branch to predict a segmentation mask for each Region of interest (RoI), which helps in performing instance segmentation. This model works very well in identifying individual buildings in densely populated zones, whereby buildings can overlap or be close to each other.

3.2.1.2 Model Training

All three CNN models of U-Net, FCN, and Mask R-CNN were trained individually using a batch size of 1. The learning rate during training was set to 0.001 and optimized by the Adam optimizer. Binary cross-entropy loss function was used to deal with the binary classification problem. Their architecture's ability to model and accurately segment complex building structures was fundamental in LoD2 building reconstruction, as it came through with very good performance of the models while being trained for 1000 epochs in the case of the U-Net and FCN, and 100 epochs for Mask R-CNN. Sigmoid activation along with a threshold of 0.3 ensured separating buildings from the background and hence guaranteed reliable outputs for further model-driven refinement. ResNet 101 backbone architecture for the Mask R-CNN model was used. Implementations of all models were done

using Python v.3.7 with TensorFlow v.1.15.1, taking advantage of the computational power provided by a Core i9 CPU, 64 GB RAM, and NVIDIA GTX1070 GPU.

3.3. LoD2 Building Reconstruction

3.3.1. Model Fitting and Integration

In this study, the integration of deep learning with model-driven approach is used for LoD2 building reconstruction. It commences by integration of best building segmentation results performed by deep learning models (section 3.2) with the DSM, then predefined models of buildings will be examined to identify initial building model. Figure 5 describes the predefined building models. The initial building models are generated through several geometrical parameters proposed by (Partovi et al., 2019) as shown in equation (1):

$$\psi \in \Psi;\ \Psi = \{P, C, S\} \quad --(1)$$

where Ψ defines the position parameters P = {x_o, y_o, orientation}, the contour parameters C = {length, width}, and S which is the shape parameters of the building model including Zridge, Zeave, hipl1, hipl2, hipw1 and hipw2, respectively, as shown in Figure 6. Both *Z*ridge and Zeave was calculated based on *Building height* parameter, which is computed as the average elevation of the DSM, while other roof components, such as vertices, edges and facets, and their relationships are determined from the geometrical parameters.

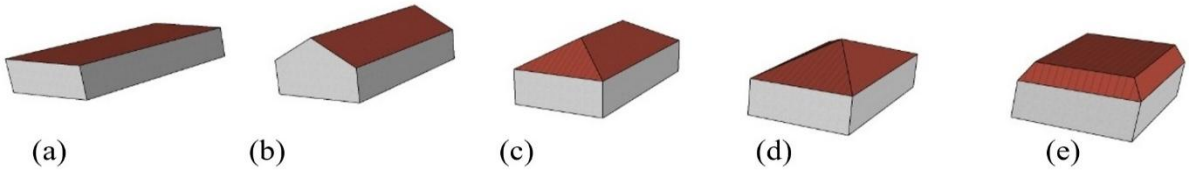


Figure 5: Types of roof model library which classified based on the type of the roof (a) flat, (b) gable, (c) hip, (d) pyramid, and (e) mansard roof

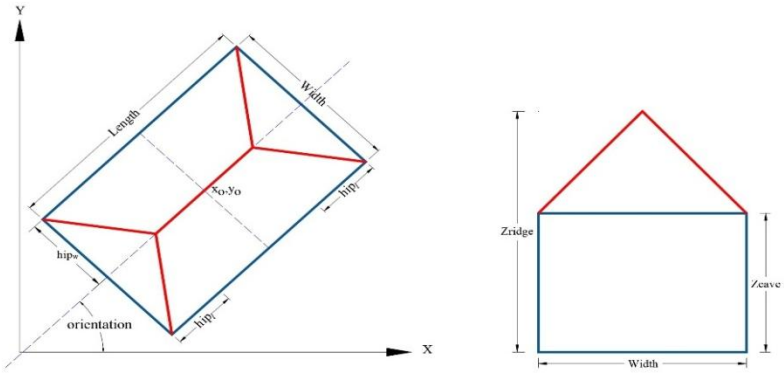


Figure 6: Roof model geometrical parameters

Model fitting initiated through orientation of the basic shapes of buildings according to the contours of buildings that segmented from best performance of CNN architectures obtained in section 3.2.1.2. In order to align these predefined models as accurately as possible, geometric transformations of translation, rotation, and scaling are applied to the segmented outlines. These geometric principles, as part of the model-driven approach, bring down probable gaps between the detected outlines and predefined models, hence giving correct and constant LoD2 reconstructions. In this study, a number of preliminary roof height parameters were utilized for model fitting. *Flat roofs* have no ridge or eave projection. The heights above ground of both the ridge and the eave are the building height decreased by 0.5 meters. *Gable roofs* have a ridge at half of the building width, starting at a height of full building height, with an eave height of the height decreased by 0.5 meters. The ridge for the *hip roofs* should be one-quarter through the length of the building and half through the width of the building. The height at the ridge should be full building height, while at the eave it

should be classed as height with 0.5 meters reduced from it.

3.3.2. Model Refinement

The last step in reconstruction is post-refinement: integration each of semantic and instance segmentation individually with the OpenStreetMap (OSM), which allows the inclusion of external geographic data for refinment process, and checking the fitted models for consistency and accuracy through graph cut optimization method (GC). This may involve the adjustment of model parameters to obtain a better fit to the observed data. In order to ensure the reconstructed models, to be remained in reasonable architectural forms, some geometric constraints of position, orientation, height, and hip ditances are utilized: maintaining the parallelism and orthogonality of the edges of buildings. These constraints can be very usefully applied in urban environments where buildings normally contain some structural patterns. According to (Partovi et al., 2019), the fact that these constraints are taken into consideration in the model adjustment process increases the accuracy of the whole process and gives a guarantee that the resulting 3D models are an exact manifestation of the actual architectural styles expressed in the urban landscape.

4. Experimental Results

The trained models are applied on the study area. The performance of U-Net, FCN, and Mask-RCNN architectures was evaluated using metrics (precision, recall, F1-score, and IoU) based on the rich dataset for building boundary segmentation from satellite images. More specifically, this dataset consists of two diverse subsets, concerning both image size and type, which are presented in sections 3.1 and 3.2. The study area of 0.3 square kilometers included 405 buildings in total. The dataset contains 201 images, split into the following parts: training 80%, validation 10%, and test 10%. Segmentation results are shown in (Figure 7, 8, and 9).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

$$IoU = \frac{A \cap B}{A \cup B} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Where |A∩B| is the area of overlap between the predicted segmentation (A) and the ground truth (B), and |A∪B| is the area of their union, TP is the number of true positives, FP is the number of false positives, FN is the number of false negatives, TN is the number of true negative.

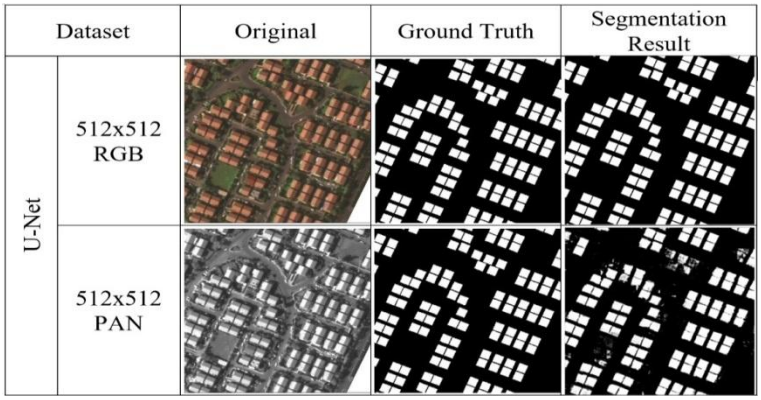


Figure 7: Building extraction using U-Net semantic segmentation

Figure 8: Building extraction using FCN semantic segmentation

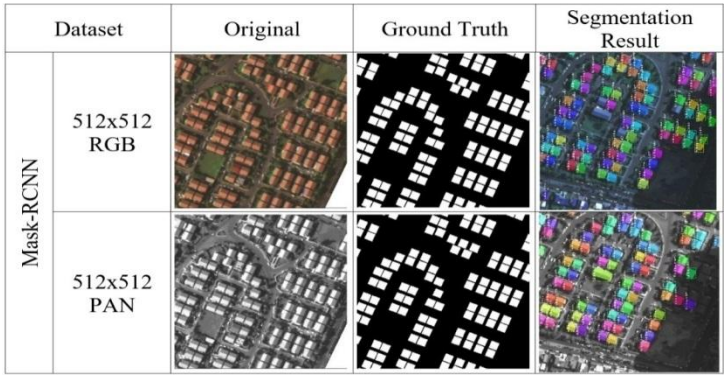


Figure 9: Building extraction using Mask-RCNN Instant segmentation

The results shown that U-Net performed best among other models. On RGB images, very close to perfect precision of 0.999, F1-score of 0.970, and IoU of 0.990. This model performed very well on the PAN images, with a Precision of 0.905, F1-Score of 0.880, and an IoU of 0.953. FCN, performed a mid-level of building extraction, in which F1-Score and accuracy result did very well using RGB images at 0.930 and 0.962, respectively. Mask-RCNN was recorded best results on PAN images with Precision at 0.763, Recall at 0.793, and Accuracy of 0.965. Evaluation metrics shown in Table 1.

Table1: evaluation metrics of deep learning architectures

	U-NET		FCN		Mask-RCNN	
	PAN	RGB	PAN	RGB	PAN	RGB
Recall	0.798	**0.940**	0.725	**0.882**	**0.793**	0.697
Precision	0.905	**0.999**	0.700	**0.903**	**0.763**	0.650
Accuracy	0.953	**0.990**	0.900	**0.962**	**0.965**	0.960
F1- Score	0.880	**0.970**	0.820	**0.930**	**0.777**	0.673
IoU	0.953	**0.990**	0.900	**0.962**		

4.1 Accuracy Assessment of LoD2

In order to perform an accuracy assessment of the reconstructed LoD2 buildings from deep learning model-driven approach, two areas were selected: Area 1 and Area 2, as shown in Figure 10. the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are commonly used evaluation metrics were utilized. Area 1 which is considered as open area, 30 buildings out of 70 were nominated for assessment. While in Area 2, the assessment was mostly concentrated on the occluded (5 building) and its surrounding buildings (9 building), as shown in Figure (11 and 12).

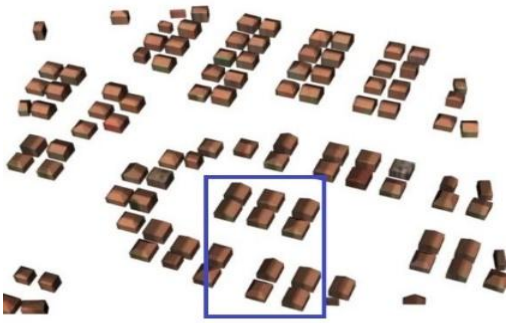
$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - y_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - y_i)^2} \quad (7)$$

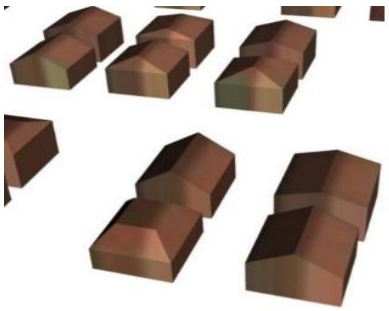
Where n is the number of predictions, Y_i is the predicted value, y_i is the actual value.

Figure 10: Accuracy assessment areas of reconstructed LoD2 building

(a)



(b)



(c)

Figure 11: (a) Ortho images of Area 1, (b) Reconstructed LoD2 building, (c) Enlarged building (blue rectangle)

(a)

(b)

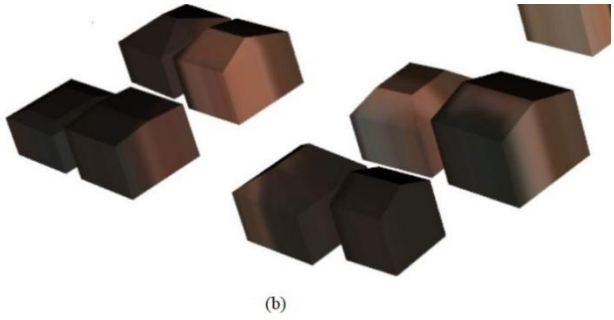


Figure 12: (a) Ortho images of Area 2, (b) Reconstructed LoD2 building of occluded area (blue rectangle)

The MAE of ridge and eave heights in Area 1 were 0.61 m and 0.68 m, respectively; 0.75 m and 0.55m in Area 2 including occluded buildings. The obtained MAE of ridge and eave heights for occluded area were 0.97 m and 0.61 m. The RMSE of ridge and eave heights in Area 1 was 0.74 m and 0.82 m, respectively; 0.89 m and 0.61 m in Area 2; and 1.19 m and 0.78 m in occluded area. Table 2, 3, and 4 shown the assessment results of LoD2 of Area 1, Area 2, and occluded area respectively.

Table 2: Accuracy assessment of the reconstructed LoD2 buildings of Area 1

Building No	Ridge Line			Eave Line		
	Predicted	Reference	ΔH	Predicted	Reference	ΔH
1	9.53	9.2	0.33	6.2	7.0	0.8
2	10.32	9.2	1.12	8.56	7.0	1.56
3	9.54	9.2	0.34	5.96	7.0	1.04
4	9.18	9.2	0.02	6.48	7.0	0.52
5	8.98	9.2	0.22	6.00	7.0	1
6	9.63	9.2	0.43	7.07	7.0	0.07
7	9.68	9.2	0.48	6.55	7.0	0.45
8	9.29	9.2	0.09	7.18	7.0	0.18
9	10.28	9.2	1.08	5.32	7.0	1.68
10	9.24	9.2	0.04	8.18	7.0	1.18
11	9.64	9.2	0.44	7.71	7.0	0.71
12	9.09	9.2	0.11	6.83	7.0	0.17
13	9.55	9.2	0.35	6.98	7.0	0.02
14	9.90	9.2	0.7	7.42	7.0	0.42
15	9.40	9.2	0.2	7.25	7.0	0.25
16	9.82	9.2	0.62	7.78	7.0	0.78
17	9.77	9.2	0.57	6.53	7.0	0.47
18	9.31	9.2	0.11	6.01	7.0	0.99
19	9.68	9.2	0.48	6.41	7.0	0.59
20	8.97	9.2	0.23	6.60	7.0	0.4
21	7.71	9.2	1.49	6.52	7.0	0.48
22	8.21	9.2	0.99	6.42	7.0	0.58
23	8.60	9.2	0.6	6.68	7.0	0.32
24	8.53	9.2	0.67	6.66	7.0	0.34
25	10.13	9.2	0.93	7.28	7.0	0.28
26	10.09	9.2	0.89	7.78	7.0	0.78

27	8.14	9.2	1.06	6.34	7.0	0.66
28	7.89	9.2	1.31	5.72	7.0	1.28
29	8.00	9.2	1.2	5.14	7.0	1.86
30	7.91	9.2	1.29	6.48	7.0	0.52
RMSE (m)			0.74			0.82
MAE (m)			0.61			0.68

Table 3: Accuracy assessment of the reconstructed LoD2 buildings of Area 2

Building No	Ridge Line			Eave Line		
	Predicted (m)	Reference (m)	ΔH	Predicted (m)	Reference (m)	ΔH
1	10.21	9.2	1.01	7.47	7.0	0.47
2	10.02	9.2	0.82	7.78	7.0	0.78
3	10.02	9.2	0.82	7.38	7.0	0.38
4	9.90	9.2	0.7	6.64	7.0	0.36
5	9.57	9.2	0.37	7.83	7.0	0.83
6	8.24	9.2	0.96	6.39	7.0	0.61
7	9.26	9.2	0.06	6.00	7.0	1
8	8.83	9.2	0.37	6.74	7.0	0.26
9	7.51	9.2	1.69	6.00	7.0	1
10	7.43	9.2	1.77	6.98	7.0	0.02
11	9.31	9.2	0.11	6.64	7.0	0.36
12	9.84	9.2	0.64	7.44	7.0	0.44
13	9.71	9.2	0.51	6.55	7.0	0.45
14	9.81	9.2	0.61	7.70	7.0	0.7
RMSE			0.89			0.61
MAE			0.75			0.55

Table 4: Evaluation of LOD2 of occluded area

Building No	Ridge Line			Eave Line		
	Predicted (m)	Reference (m)	ΔH	Predicted (m)	Reference (m)	ΔH
5	9.57	9.2	0.37	7.83	7	0.83
6	8.24	9.2	0.96	6.39	7	0.61
7	9.26	9.2	0.06	6.00	7	1
9	7.51	9.2	1.69	6.00	7	1
11	7.43	9.2	1.77	6.98	7	0.02
RMSE			1.19			0.78
MAE			0.97			0.69

5. Discussion

In this study, three deep learning models, U-Net, FCN, and Mask-RCNN, are used for the segmentation of building boundaries on panchromatic and RGB images as shown in Figure 13. Across all metrics for both PAN and RGB images, results showed U-Net to be the best among the considered models. This model performed very well on the RGB and PAN images. On the other hand, FCN based building extraction demonstrated mid-level performance; it shows better result with RGB images. Nevertheless, FCN had a very poor with respect to the PAN images in comparison to U-Net. It is

observed that the performance of Mask R-CNN was the lower among the three models on RGB images. While PAN images had better performance especially on the occluded areas, but still way below the other models with the best Accuracy of about 0.965. Hence, the semantic segmentation U-Net was used in open areas, and instant segmentation Mask R-CNN was used for LoD2 building reconstruction in occluded areas.

In order to perform the segmentation over the open and occluded areas simultaneously through a unique deep learning model, its recommended to make integration between semantic and instant segmentation architectures, and reduce the input patch size into 256 x 256 pixel.

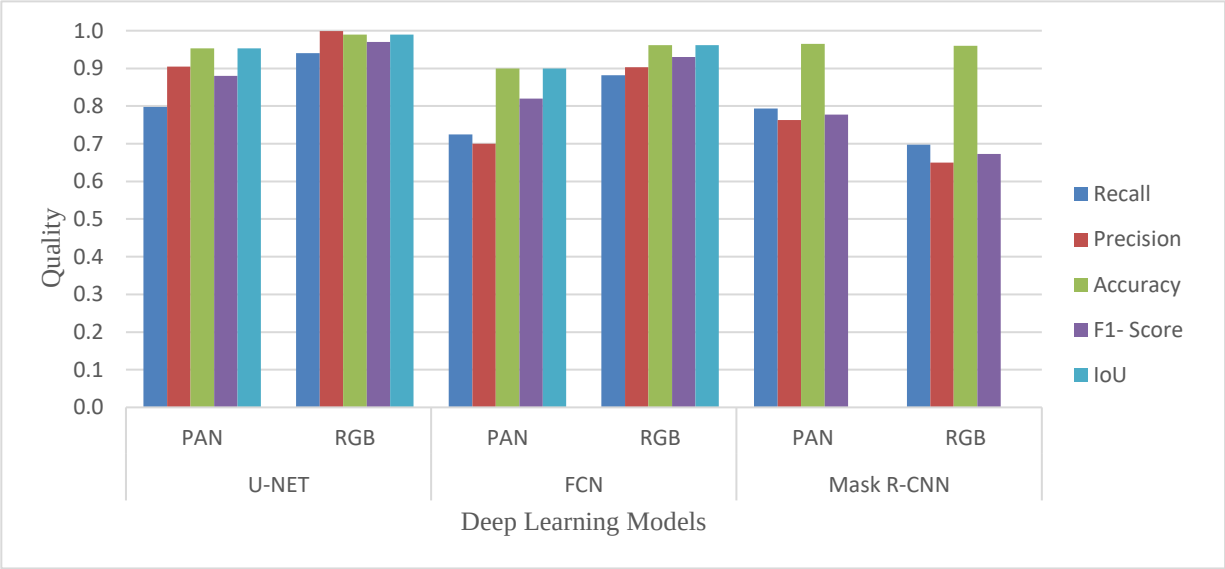


Figure 13: Segmentation results

In the comparison of the predicted and actual heights of buildings in Area 1, Area 2 including occluded buildings, Figure 14 following metrics were computed to check the accuracy of the prediction: The Mean absolute error (MAE) of ridge and eave heights in Area 1 Area 2 was comparable of range (0.6 - 0.75) m. The obtained MAE of ridge and eave heights for occluded area falls within sub-meter. This indicates that on average, predicted building heights is close and show a small deviation. The Root mean square error (RMSE) of ridge and eave heights in Area 1 and Area 2 was close to limits (0.61 - 0.89) m; and reaches to sub-meter in occluded area. The RMSE gives greater weight to provide a more conservative estimate for prediction accuracy. Most of the predictions are located within a reasonable range around the actual height. LoD2 of 5 buildings out of 8 was perfectly constructed in an occluded area. A few overestimations and underestimations measurements were found, but most predictions are quite close to real heights. Some of the roof types seems to be suffering from misclassification; this is due to the quality of the DSM. In terms of occluded area, the roof type of 5 buildings out of 8 was correctly classified.

Additionally, obtained RMSE of Area 1 in this study has been compared to the works performed by (Partovi et al., 2019), and (Wang et al., 2021) as shown in Figure 15. Its evident the achieved RMSE of ridge in this study matches (Wang et al., 2021) and is better than (Partovi et al., 2019). While gained RMSE of eave is lowest among studies, showing significant improvement over them.

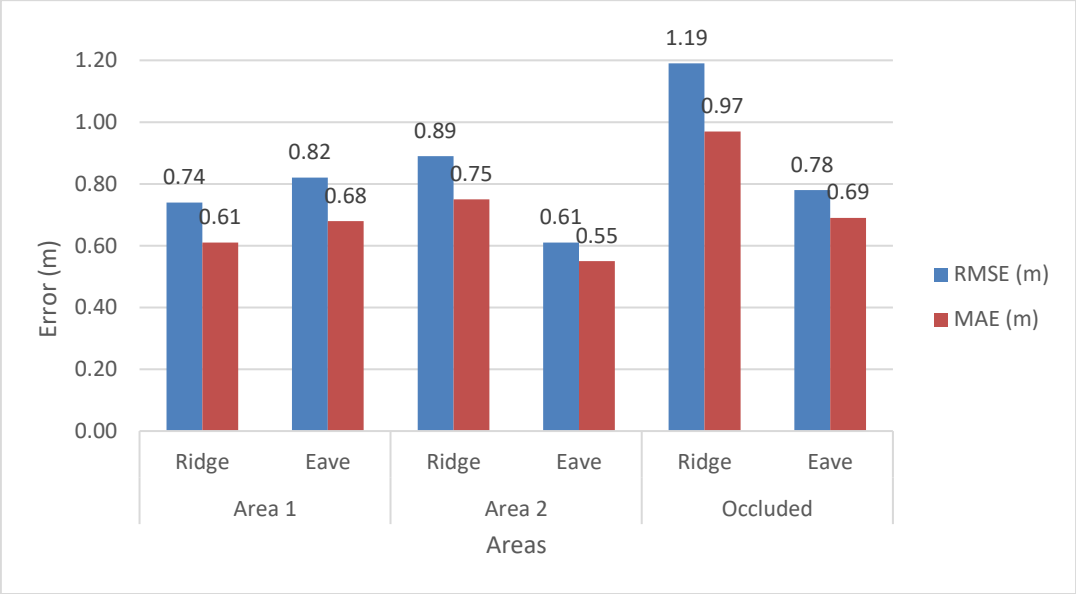


Figure 14: Evaluation of LoD2 building reconstruction

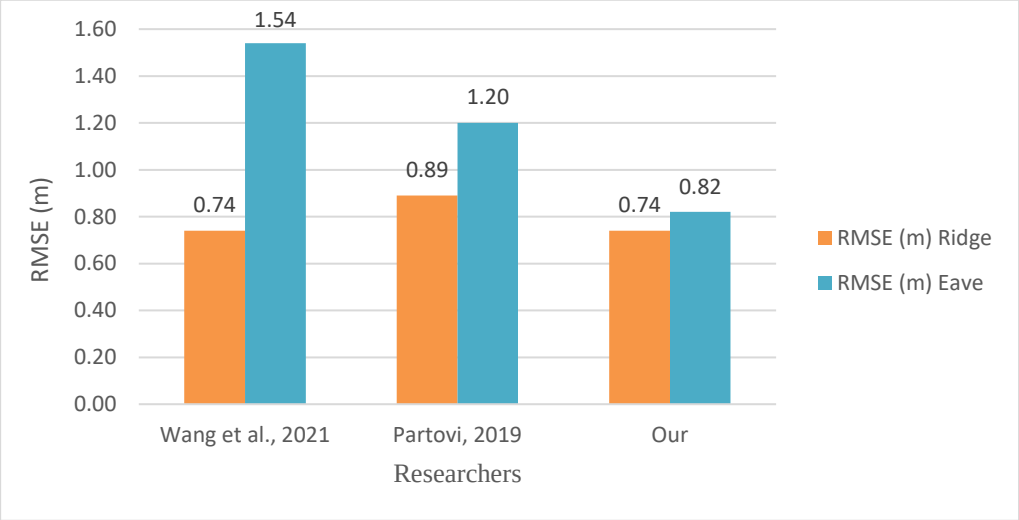


Figure 15: Comparison between LoD2 results obtained by ours and researchers

5. Conclusion

This study has proposed a deep learning model-driven approach for LoD2 building reconstruction using stereo satellite imagery. The metric parameters for the segmentation of building boundaries using U-Net, FCN, and Mask R-CNN for PAN and RGB images. U-Net was performed better with RGB images, where it turns out that almost perfect precision and accuracy, Table 1. This clearly exhibits the excellent ability of U-Net in capturing buildings boundary. FCN demonstration is modest in results, working out better in the case when RGB images are used with high F1-score and accuracy, Table 1. Mask R-CNN has the lowest performance among the three models. This has far lower recall and precision, which reduces the F1-scores and segmentation accuracy. It performed better in the PAN image but still behind the U-Net and FCN. Although, U-Net architecture and RGB images have effectively better segmentation results across other architectures, Mask R-CNN with PAN images has had a better performance for instance segmentation in occluded areas, as shown in Figure 13.

The height and roof type assessment of buildings in Area 1 and Area 2 including occluded buildings indicates the prediction performance was generally correct but deviated slightly from the real height. For most buildings, the height is

predicted accurately to be around 9.2 m and 7.0 m for ridge and eave, respectively, which also means that this approach for reconstruction works quite well, Table 2-4. Obtained MAE for ridge and eave heights in Area 1 and Area 2 indicates that the average prediction keeps within 0.7 meters of actual building heights. While MAE for ridge and eave heights in occluded area was approximately 1.0 m, Figure 14. The RMSE for ridge and eave heights in Area 1 and 2 indicates that most buildings' height errors are within 0.9 m which is can be acceptable. Achieved RMSE for the occluded area was approximately within 1.0 m for ridge and eave heights, Figure 14. The reconstruction error mean for building heights is reasonable, which might be considered to be accurate enough for such a task. Based on these results, this model could reconstruct the height of buildings reliable within a close range from the real value. Although, the proposed approach has shown the ability to expand to larger areas having occlusions, its highly recommend to use multi-view stereo images, decreasing patch size, and increasing the size of dataset to work the model robustly. Obtained accuracy indicates the methodology used in this study is effective for improving accuracy, further refinement still required to trim down the error margin for better roof type classification. Future works will focus on the improvement of roof type classification through utilizing conditional generative adversarial network (cGAN).

Acknowledgment: Authors would like to thank Salahaddin University-Erbil for their support throughout the research.

Financial support: No financial support.

Potential conflicts of interest: Authors declare no conflicts of interest relevant to this article.

References

Akiki, R., Mari, R., De Franchis, C., Morel, J.-M. & Facciolo, G. Robust Rational Polynomial Camera Modelling for SAR and Pushbroom Imaging. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021/7// 2021 Brussels, Belgium. IEEE, 7908-7911.

Alidoost, F., Arefi, H. & Tombari, F. 2019. 2D image-to-3D model: Knowledge-based 3D building reconstruction (3DBR) using single aerial images and convolutional neural networks (CNNs). *Remote Sensing,* 11.

Alin, A. Y., Kusrini, K. & Yuana, K. A. 2023. The Effect of Data Augmentation in Deep Learning with Drone Object Detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems),* 17**,** 237-248.

Alsabhan, W., Alotaiby, T. & Dudin, B. 2022. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Computational Intelligence and Neuroscience,* 2022.

Amarù, S., Marelli, D., Ciocca, G. & Schettini, R. 2023. DALib: A Curated Repository of Libraries for Data Augmentation in Computer Vision. *J Imaging,* 9.

Amo-Boateng, M., Ekow Nkwa Sey, N., Ampah Amproche, A. & Kyereh Domfeh, M. 2022. Instance segmentation scheme for roofs in rural areas based on Mask R-CNN. *Egyptian Journal of Remote Sensing and Space Science,* 25**,** 569-577.

Amro, I., Mateos, J., Vega, M., Molina, R. & Katsaggelos, A. K. 2011. A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP Journal on Advances in Signal Processing,* 2011.

Bittner, K., Adam, F., Cui, S., Körner, M. & Reinartz, P. 2018a. Building Footprint Extraction From VHR Remote Sensing Images Combined With Normalized DSMs Using Fused Fully Convolutional Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 11**,** 2615-2629.

Bittner, K., D'angelo, P., Körner, M. & Reinartz, P. 2018b. DSM-to-LoD2: Spaceborne stereo digital surface model refinement. *Remote Sensing,* 10.

Buyukdemircioglu, M., Kocaman, S. & Kada, M. DEEP LEARNING FOR 3D BUILDING RECONSTRUCTION: A REVIEW. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2, 2022. 359-366.

Dorninger, P. & Pfeifer, N. 2008. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. *Sensors,* 8**,** 7323-7343.

Dos Santos, R. C., Galo, M. & Habib, A. F. 2020. Regularization of building roof boundaries from airborne LiDAR data using an iterative CD-spline. *Remote Sensing,* 12.

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,* 10**,** 112-122.

Duan, L. & Lafarge, F. Towards large-scale city reconstruction from satellites. European Conference on Computer Vision (ECCV), 2016/9// 2016. Computer Vision – ECCV 2016, 89-104.

Dukai, B., Ledoux, H. & Stoter, J. E. A multi-height LOD1 model of all buildings in the Netherlands. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2019. Copernicus GmbH, 51-57.

Dukai, B., Peters, R., Vitalis, S., Liempt, J. V. & Stoter, J. QUALITY ASSESSMENT of A NATIONWIDE DATA SET CONTAINING AUTOMATICALLY RECONSTRUCTED 3D BUILDING MODELS. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2021/9// 2021. International Society for Photogrammetry and Remote Sensing, 17-24.

Dutta, A. & Zisserman, A. The VIA Annotation Software for Images, Audio and Video. Proceedings of the 27th ACM International Conference on Multimedia, 2019/10// 2019 New York, NY, USA. 2276-2279.

Gao, W., Peters, R., Ledoux, H. & Stoter, J. Filling holes in LoD2 building models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2024/6// 2024. Copernicus Publications, 171-177.

Grompone Von Gioi, R., Jakubowicz, J., Morel, J.-M. & Randall, G. 2010. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 32**,** 722-732.

Gui, S., Qin, R. & Tang, Y. 2022. Sat2Lod2: a Software for Automated Lod-2 Building Reconstruction From Satellite-Derived Orthophoto and Digital Surface Model. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives,* 43**,** 379-386.

Han, Q., Yin, Q., Zheng, X. & Chen, Z. 2022. Remote sensing image building detection method based on Mask R-CNN. *Complex and Intelligent Systems,* 8**,** 1847-1855.

He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2017/10// 2017. IEEE, 2980-2988.

Huang, H., Michelini, M., Schmitz, M., Roth, L. & Mayer, H. LOD3 BUILDING RECONSTRUCTION from MULTI-SOURCE IMAGES. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2020/9// 2020. International Society for Photogrammetry and Remote Sensing, 427-434.

Huang, J., Stoter, J., Peters, R. & Nan, L. 2022. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. *Remote Sensing,* 14.

Lai, F. & Yang, X. 2020. Integrating spectral and non-spectral data to improve urban settlement mapping in a large Latin-American city. *GIScience & Remote Sensing,* 57**,** 830-844.

Li, T., Jiang, C., Bian, Z., Wang, M. & Niu, X. 2020. A Review of True Orthophoto Rectification Algorithms. *Materials Science and Engineering,* 780**,** 22-35.

Long, J., Shelhamer, E. & Darrell, T. Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. IEEE, 3431-3440.

Lv, B., Liu, J., Wang, P. & Yasir, M. 2022. DSM Generation from Multi-View High-Resolution Satellite Images Based on the Photometric Mesh Refinement Method. *Remote Sensing,* 14**,** 6259-6259.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N. & Terzopoulos, D. 2022. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 44**,** 3523-3542.

Muftah, H., Rowan, T. S. L. & Butler, A. P. 2022. Towards open-source LOD2 modelling using convolutional neural networks. *Modeling Earth Systems and Environment,* 8**,** 1693-1709.

Noureldeen, A. & Wahed, M. E. 2024. Enhanced building footprint extraction from satellite imagery using Mask R-CNN and PointRend. *Bulletin of Electrical Engineering and Informatics,* 13**,** 3601-3608.

Nys, G. A., Poux, F. & Billen, R. 2020. City json building generation from airborne LiDAR 3D point clouds. *ISPRS International Journal of Geo-Information,* 9.

Partovi, T., Fraundorfer, F., Bahmanyar, R., Huang, H. & Reinartz, P. 2019. Automatic 3-D building model reconstruction from very high resolution stereo satellite imagery. *Remote Sensing,* 11.

Peters, R., Dukai, B., Vitalis, S., Van Liempt, J. & Stoter, J. 2022. Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands. *Photogrammetric Engineering and Remote Sensing,* 88**,** 165-170.

Proulx-Bourque, J.-S., Mathieu, L., Papasodoro, C., Pilon, D., Sabo, N. & Pelchat, M. T. Experiment on the Impact of Spatial Resolution on Building Extraction Accuracy. 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019/8// 2019 Yokohama, Japan. IEEE.

Ps, P. & Aithal, B. H. 2023. Building footprint extraction from very high-resolution satellite images using deep learning. *Journal of Spatial Science,* 68**,** 487-503.

Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. ISBI challenge for segmentation of neuronal structures in electron microscopic stacks, 2015/9// 2015.

Stucker, C. & Schindler, K. 2022. ResDepth: A deep residual prior for 3D reconstruction from high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing,* 183**,** 560-580.

Sugihara, K., Murase, T. & Zhou, X. Automatic generation of 3D building models from building polygons on digital maps. 2015 International Conference on 3D Imaging (IC3D), 2015/12// 2015 Liege, Belgium. IEEE.

Wagner, F. H., Dalagnol, R., Tarabalka, Y., Segantine, T. Y. F., Thomé, R. & Hirye, M. C. M. 2020. U-net-id, an instance segmentation model for building extraction from satellite images-Case study in the Joanopolis City, Brazil. *Remote Sensing,* 12.

Wang, Y., Zorzi, S. & Bittner, K. Machine-learned 3D Building Vectorization from Satellite Imagery. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021/6// 2021. IEEE, 1072-1081.

Weng, Q. Essential Urban Variables from Satellite Observations: An Introduction. IEEE, 2018/9// 2018. 2018 IEEE International Geoscience and Remote Sensing Symposium.

Wysocki, O., Hoegner, L. & Stilla, U. MLS2LoD3: Refining low LoDs building models with MLS point clouds to reconstruct semantic LoD3 building models. *In:* THOMAS H. KOLBE, A. D., CHRISTOF BEIL, ed. 18th 3D GeoInfo Conference, 2024/9// 2024. Springer, 367-380.

Xu, B., Zhang, X., Li, Z., Leotta, M., Chang, S.-F. & Shan, J. 2020. Deep Learning Guided Building Reconstruction from Satellite Imagery-derived Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Zhang, S., Han, F. & Bogus, S. M. Building Footprint and Height Information Extraction from Airborne LiDAR and Aerial Imagery. Construction Research Congress 2020, 2020/11// 2020 Reston, VA. American Society of Civil Engineers, 326-335.

Zhang, Y., Zhang‹, Y., Mo‹, D., Zhang, Y. & Li, X. 2017. Direct Digital Surface Model generation by semi-global vertical line locus matching. *Remote Sensing,* 9.

Zhao, K., Kang, J., Jung, J. & Sohn, G. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018, pp. 247-251, 2018. 247-251.