

## DEEP LEARNING FOR 3D BUILDING RECONSTRUCTION: A REVIEW

Mehmet Buyukdemircioglu<sup>1,2,4</sup>, Sultan Kocaman<sup>2,3\*</sup>, Martin Kada<sup>4</sup>

<sup>1</sup> Hacettepe University, Graduate School of Science and Engineering, Ankara, Turkey - mbuyukdemircioglu@hacettepe.edu.tr

<sup>2</sup> Hacettepe University, Department of Geomatics Engineering, Ankara, Turkey – sultankocaman@hacettepe.edu.tr

<sup>3</sup> ETH Zurich, Institute of Geodesy and Photogrammetry, 8093 Zurich, Switzerland

<sup>4</sup> Technische Universität Berlin, Institute of Geodesy and Geoinformation Science, Berlin, Germany – martin.kada@tu-berlin.de

### Commission II, WG II/4

**KEY WORDS:** Deep Learning, Machine Learning, 3D City Models, 3D Reconstruction, LoD.

### ABSTRACT:

3D building reconstruction using Earth Observation (EO) data (aerial and satellite imagery, point clouds, etc.) is an important and active research topic in different fields, such as photogrammetry, remote sensing, computer vision and Geographic Information Systems (GIS). Nowadays 3D city models have become an essential part of 3D GIS environments and they can be used in many applications and analyses in urban areas. The conventional 3D building reconstruction methods depend heavily on the data quality and source; and manual efforts are still needed for generating the object models. Several tasks in photogrammetry and remote sensing have been revolutionized by using deep learning (DL) methods, such as image segmentation, classification, and 3D reconstruction. In this study, we provide a review on the state-of-the-art machine learning and in particular the DL methods for 3D building reconstruction for the purpose of city modelling using EO data. This is the first review with a focus on object model generation based on the DL methods and EO data. A brief overview of the recent building reconstruction studies with DL is also given. We have investigated the different DL architectures, such as convolutional neural networks (CNNs), generative adversarial networks (GANs), and the combinations of conventional approaches with DL in this paper and reported their advantages and disadvantages. An outlook on the future developments of 3D building modelling based on DL is also presented.

### 1. INTRODUCTION

Although 3D city models have initially been used for visualization purposes; they have been increasingly utilized in a variety of domains and tasks, such as collaborative urban planning, population density analysis, mobile telecommunication applications, solar potential assessments, disaster management, 3D navigation, and environmental simulations (Biljecki et al., 2015). Researchers have been investigating methods on the automatic 3D reconstruction of buildings from Earth Observation (EO) data and their modelling for three decades (Haala and Kada, 2010). There are several manual steps involved in 3D reconstruction in traditional methods, including image pre-processing, 3D point cloud extraction, data fusion, and texture mapping. Thus, cumulative errors occur in the process and cause inaccurate semantic features in the reconstruction of 3D shapes that seriously affect their quality (Liu et al., 2021). Thanks to the availability of benchmark datasets, such as the airborne images and laser scanner data by ISPRS WGIII/4 (Rottensteiner et al., 2014, Rottensteiner et al., 2012), it has been possible to comparatively evaluate various methods for the segmentation of urban objects and also the 3D reconstruction of buildings.

In the last decades, city models were produced either manually by photogrammetry operators from aerial imagery or by using conventional methods (non-Machine Learning methods). The conventional 3D building reconstruction methods can be categorized as model-driven and data-driven methods. The model-driven methods aim to match the geometry of the roof

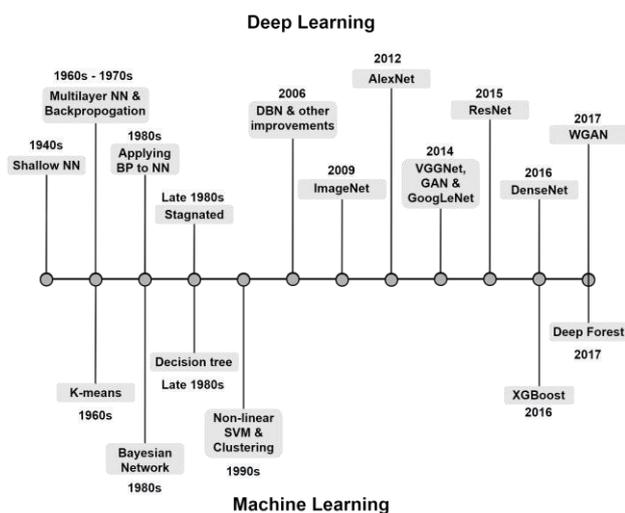
generated from digital surface models (DSMs), e.g. point clouds, with the roof types in a library (Henn et al., 2013). Using this approach, it can be ensured that the reconstructed roof model is topologically correct; but problems may occur if there is no candidate for the roof shape in the library. Moreover, model-driven methods utilize a limited number of pre-defined shapes given in the model libraries, which reduces the production accuracy. In addition, complex roof structures may not be modelled. In data-driven methods, a DSM (often in the form of a point cloud) is utilized as primary data source and the models are generated from this data as a whole without focusing on any particular parameter. In the data-driven approach, the main problem is that the extracted segments may not be intersected successfully leading to topological or geometrical errors. The data-driven methods are usually not robust and highly sensitive to noise in the data. Due to the noise sensitivity of data-driven methods, data pre-processing is an essential step to avoid incorrect results.

The way the geospatial domain operates is changing fundamentally as a result of Artificial Intelligence (AI) (Döllner, 2020). Deep learning (DL) methods, in particular, Convolutional Neural Networks (CNNs) have been the game changers for several tasks related to photogrammetry and remote sensing in recent years. The recently developed DL methods have potential to overcome the limitations of conventional 3D city modelling and building reconstruction methods.

\* Corresponding author

3D building construction using DL is a relatively new research area studied during the last years with few publications on this topic. The DL approaches achieved state-of-the-art results in classification, segmentation, and change detection using EO data when compared to the conventional methods and there are already published review articles on these topics (Ma et al., 2019; Hoerer and Kuenzer, 2020; Heipke and Rottensteiner, 2020). As DL gains popularity in different fields, new areas of application will emerge in the future.

DL based 3D reconstruction has become increasingly feasible with the rapid development of 3D building models and the availability of many different 3D shapes in recent years. DL models can be trained to learn 3D shapes and their features, characteristics and details. Wichmann et al. (2018) presented RoofN3D, a new 3D point cloud training dataset that can be used to train machine learning (ML) and DL models for a variety of tasks in the context of 3D building reconstruction. An overview of the timeline of the development of commonly used machine learning (ML) algorithms and the DL methods is given in Figure 1.



**Figure 1.** Timeline of the development of DL and commonly used ML algorithms (modified after Cao et al., 2018). NN: neural network; BP: backpropagation; DBN: deep belief network; SVM: support vector machines; AE: auto-encoder; VAE: variational AE; GAN: generative adversarial network; XGBoost: Extreme Gradient Boost; WGAN: Wasserstein GAN

In this paper, based on the recent advancements in the field and considering the increasing interest in the community; we aim at providing an overview of methods and state-of-the-art applications for DL-based 3D building reconstruction with a focus on object generation. The remainder of this paper is structured as follows: Section 2 gives a brief overview of conventional (non-DL) 3D building reconstruction methods. Section 3 presents a summary of ML methods used in 3D building reconstruction. The state-of-the-art DL method studies according to method types are given in Section 4. Discussions and conclusions are provided in the final section together with future directions.

## 2. CONVENTIONAL METHODS

3D building reconstruction is still largely based on conventional methods and algorithms (i.e., non-DL-based). Two comprehensive reviews on conventional urban reconstruction methods were presented by Musialski et al. (2013) and Halaa and Kada (2010). Sub-surface growing is an example to conventional methods for 3D building reconstruction (Kada and Wichmann, 2012). Nan and Wonka (2017) proposed a data-driven method, Polyfit, for reconstructing lightweight polygonal surfaces from point clouds. By combining the Random Sample consensus (RANSAC) method (Fischler and Bolles, 1981) with contextual knowledge, Malihi et al. (2018) have developed a novel two-level segmentation scheme for generating 3D building models from point clouds derived from UAV photogrammetry. LoD1 building models can be automatically constructed with the combination of 2D building footprints and digital surface models (DSMs) (Buyukdemircioglu and Kocaman, 2018). Model-driven approaches can also be used for semi-automatic reconstruction of 3D city models in LoD2 using stereo aerial imagery (Buyukdemircioglu et al. (2018) as depicted in Figure 2.



**Figure 2.** LoD2 city model of Cesme, Turkey reconstructed with a semi-automatic method (Buyukdemircioglu et al., 2018)

3D building models can be generated from different EO data types such as aerial imagery, UAV imagery, satellite imagery, or point clouds using conventional approaches such as rule-based methods (Xie et al., 2021), model-driven methods, or data-driven methods. One of the fast and widely used reconstruction method for 3D city models is to extrude building footprints. Another automatic 3D building reconstruction method based on half-spaces in LoD2 was proposed by Bizjak et al. (2021). Their proposed algorithm performed reconstruction on the ISPRS benchmark dataset with RMSE of 1.31m and completeness level of 98.9%, respectively. Drešček et al. (2020) presented an approach for 3D building reconstruction using an unmanned aerial vehicle (UAV) photogrammetric point cloud based on an extract, transform, load (ETL) solution. A data-driven and algorithmic solution to the automatic reconstruction of 3D buildings at LoD2 from UAV point clouds was presented by Murtiyoso et al. (2020).

RANSAC has been a popular method used with 3D point clouds. Li and Wu (2020) generated 3D models of complex buildings automatically using incomplete point clouds with RANSAC and topological-relation constraints. A RANSAC-based multi primitive reconstruction (MPR) method was proposed by Li and Shan (2022) to segment a compound boundary into predefined primitives and determine their parameter values from the point clouds.

Automatic generation of high detailed LoD3 models is still a challenging topic for researchers. These models can be generated manually and combined with automatically generated 3D city models in different LoDs (Buyukdemircioglu and Kocaman, 2020). An automatic workflow for reconstructing 3D building models in LoD1 and LoD2 based on 2D building footprints and LiDAR (Light Detection and Ranging) point cloud was developed by Peters et al. (2021). Their approach was used to reconstruct 10 million buildings in the Netherlands.

### 3. MACHINE LEARNING METHODS

In this Section, an overview of the ML methods other than the DL-based methods (e.g., random forests, SVM, etc.) used in literature for reconstructing 3D building models is given. Dehbi et al. (2016) developed weighted attribute context-free grammar rules for 3D building reconstruction. The weighted context-free grammar was inferred using SVMs from input-output pairs as structured data; and Markov Logic Networks (MLNs) was used to reconstruct the 3D buildings as a statistical relational learning method.

In their study, Biljecki et al. (2017) have shown that building models also can be automatically reconstructed without elevation data by using the random forest method. The proposed method predicts the height of buildings based on the footprints and building attributes and then extrudes the footprints to generate 3D models. As a result, they have reached a mean absolute error of 0.8 m in the inferred heights. Biljecki and Dehbi (2019) demonstrated that it is possible to predict the roof types from lower LoD (i.e., LoD0 and LoD1) datasets and to generate LoD2 models without roof measurements. They achieved an accuracy of 85% of the roof type from sparse data using a multiclass classification and 92% accuracy in predicting whether a roof is flat or not.

Park and Guldmann (2019) reconstructed LoD1+ building models with ML-based point cloud classification methodology that assigns LiDAR points to different classes, extracts the points reflecting a rooftop surface, and uses those points to estimate building heights. The ML methods can also be used to reconstruct multi-temporal (4D) city models. Farella et al. (2021) presented a methodology for reconstructing buildings in 4D using ML algorithms and historical information. Using digitized historical city maps and information about actual city conditions, they were able to reconstruct multi-temporal 3D representations of two urban city centres using different regression algorithms for inferring missing building heights.

### 4. DEEP LEARNING METHODS

Through DL, the computational models consisting of multiple layers of processing can learn the representations of data at multiple abstraction levels (LeCun et al., 2015). According to Liu et al. (2021), there are two main problems with conventional 3D reconstruction methods. First, they involve multiple manual designs that can lead to the accumulation of

errors but can hardly learn semantic features of 3D shapes automatically. Secondly, they are highly dependent on the quality and content of images, as well as a precisely calibrated camera. The DL-based 3D reconstruction methods overcome these bottlenecks by automatically learning 3D shape semantics from images or point clouds using deep networks.

Different DL architectures have been used in the literature for 3D building reconstruction from EO data (aerial imagery, UAV imagery, satellite imagery, point clouds, etc.). In this section, we have investigated convolutional neural networks (CNNs), generative adversarial networks (GANs), and the combinations of DL methods with conventional methods in detail for realizing this task. Here, the DL-based methods are explained in a separate section due to their increasing popularity and the availability of reference datasets.

#### 4.1 Convolutional Neural Networks

CNNs (Simonyan and Zisserman, 2014) allow to learn the characteristics of images at various levels using convolution and pooling operations, which is an extremely useful DL model for image classification and reconstruction. They can also be used for 3D building reconstruction from EO data. A DL approach was proposed by Wang and Frahm (2017) for performing a single-view parametric reconstruction of buildings based on satellite imagery by parametrizing buildings as 3D cuboids. CNNs also can be used for the procedural reconstruction of buildings. Using Neural Procedural Reconstruction (Zeng et al., 2018), 3D points were mapped into CAD (Computer-aided design)-quality models with procedural structures inferred by sequences of shape grammar rules. An interactive tool was developed by Nishida et al., (2018), with which users can generate a grammar automatically from a single image of a building with the help of CNNs for procedural modeling of buildings. Alidoost et al. (2019) developed a DL architecture for detecting buildings from a single aerial image. Using the proposed method, the 3D reconstruction of buildings with a variety of shapes and complexity was achieved with root mean square error (RMSE) values of 3.43 m and 1.13 m for the predicted normalized DSM (nDSM), respectively. To generate block-like city models using depth maps, Agoub et al. (2019) developed a pipeline based on multiple CNNs with an encoder-decoder architecture. A view of the reconstructed buildings of Manhattan area in their study is given in Figure 3.



**Figure 3.** Automatically reconstructed building models of Manhattan area using CNN (Agoub et al., 2019)

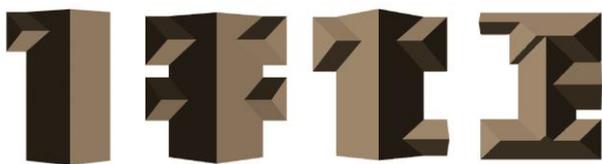
Based on a Y-shaped CNN (Y-Net), a modern DL-based framework was proposed by Alidoost et al. (2020) for automatic detecting, localizing, and estimating building heights simultaneously from a single aerial image for LoD2 building reconstruction. A multi-task, multi-feature learning framework

was presented by Mahmud et al. (2020) for modeling a building in 3D from a single overhead image. Using this approach, the authors generate 2D building outline proposals, a pixel-by-pixel heightmap, a modified signed distance function (BPSH), and pixel-by-pixel semantic labels; and then produce 3D models of each building. Another example of CNN applied to a specific grid structure was presented by Knyaz et al. (2020), where CNN was used for automatic semantic segmentation of wire structures and overcoming the limitation of photogrammetric processing applied to the 3D reconstruction of complex grid structures. The roof structure lines can be used for reconstructing 3D building models. Muftah et al. (2021) used a CNN-based method for classifying and segmenting roofs based on aerial imagery for 3D building reconstruction in LoD2. The Deep Roof Definer network proposed by Qian et al., (2022) uses satellite imagery to generate roof structure lines using a detail-oriented DL network.

#### 4.2 Generative Adversarial Networks

GANs were proposed by Goodfellow et al. (2014). A generator and a discriminator are the main parts of a GAN. The generators are mostly utilized to learn the distribution of real images, resulting in more realistic-looking images and fooling the discriminator. Discriminator involves judging the generated images either as real or fake. As part of the generation modeling process, GAN-based methods introduce the adversarial discriminator, which implicitly learns the similarities and differences between 3D shapes and can therefore identify occluded or missing portions (Liu et al., 2021).

By applying a Conditional GAN (cGAN), Bittner et al. (2018a) presented an automatic processing method for better-quality LiDAR-like DSMs with refined 3D building shape extraction from noisy DSMs. They used stereo half-meter resolution satellite imagery to create three-dimensional surfaces models and refine building shapes in LoD2 (Bittner et al., 2018b). In their next study, Bittner et al (2019), produced good-quality DSMs that show a full, accurate level of detail that is similar to LoD2-like building forms, as well as assign an additional object class label to every pixel. GANs can also be used for automatically reconstructing building models in LoD1 (Beer, 2019). FrankenGAN (Kelly et al., 2018) proposed a network for modeling realistic geometric and texture information on large-scale mass models of coarse buildings with examples as guides, users can add realistic details to large-scale models. Qian et al. (2021) presented Roof-GAN, a network that generates structured geometry of residential roof structures as a combination of roof primitives. A sample view of generated roof models by Roof-GAN is given in Figure 4.



**Figure 4.** An overview of GAN-based reconstruction of 3D roof models using Roof-GAN with 2,3,4 and 5 primitives (left to right) (Qian et al., 2021)

#### 4.3 Combination of DL-based and Conventional Methods

Rectified linear unit neural network (ReLU-NN) is another DL network that was used for the classification and reconstruction of building from airborne laser scanning point cloud data (Zhang et al., 2018). Using 3D CNNs, a deep Q-network and a residual recurrent neural network (RNN), Zhang and Zhang (2018) developed a deep reinforcement learning framework for parsing the semantics of large-scale 3D point clouds and reconstructing 3D building models. In another study, an end-to-end system for reconstructing urban 3D buildings from WorldView-3 multiview satellite imagery using DL was demonstrated by Leotta et al. (2019) by segmenting buildings and bridges and reconstructing low polygon 3D textured mesh models. Satellite imagery-derived point clouds were used by Xu et al. (2020) to build an automated DL-guided 3D reconstruction framework that distinguishes the shape of building roofs in complex and noisy scenes.

The study by Yu et al. (2021) introduced a new fully automatic 3D building reconstruction pipeline based on DL that can automatically construct building models at LoD1 from multiview aerial images without any assistance from the other data sources. In a recent study, Gui and Qin (2021) presented a DL-based approach for reconstructing the LoD-2 models using data derived from very-high-resolution multi-view satellite stereo images. Several steps were involved in their proposed method including instance-level building segment detection, initial building polygon extraction, building polygon decomposition and refinement, basic model fitting, and merging. In their study, Kapoor et al. (2019) proposed a four-step approach for generating 3D city models from historical images using DL. Another automatic 3D building model reconstruction workflow was proposed by Partovi et al. (2019). The workflow was composed of several steps, including DL-based building boundary extraction, decomposition, classification of roof types based on images, and computation of initial roof parameters for 3D model fitting.

Teo (2019) proposed a DL approach (i.e. Fully Convolutional Network- FCN) to detect initial building regions from LiDAR data and automatically reconstruct 3D prismatic building models from 3D LiDAR data. With the integration of 3D BAG CityJSON and floor plan images, Kippers et al. (2021) proposed a new automatic DL-based method for constructing building models. An automatic 3D building reconstruction in LoD1 that consists of three main parts, DSM generation, Deep learning-based 2D building footprint generation, and 3D building reconstruction proposed by Yu et al. (2020). Another study by Li et al. (2021) demonstrated a novel method for reconstructing 3D building models with accurate roofs, facades, footprints, and height from monocular remote sensing images. In their study, Zhao et al. (2021) developed a novel 3D reconstruction framework based on an off-nadir satellite image. Their approach consists of three parts: Scale-ONet for model reconstruction, Optim-Net for model scale optimization, and Model-Image match for restoring reconstructed scenes. The holistic primitive fitting method (Zhang et al., 2021) was also used along with PointNet++ (Qi et al., 2017) for 3D building reconstruction from point clouds. Another three-step 3D building reconstruction approach using deep implicit fields and point clouds was proposed by Chen et al. (2021). The DL methods also can be combined with GIS for reconstructing 3D city models from high-resolution satellite imagery (Pepe et al., 2021).

## 5. CONCLUSIONS

3D city models and digital twins are being produced and used more popularly by all over the world. Producing these models requires a great deal of data, processing, and expertise. Conventional methods for 3D building reconstruction have certain limitation, such as efficient reconstruction of large numbers of buildings at city scale. Typically, conventional methods cannot produce fully automatically; and several steps must be performed manually by users. In addition, data pre-processing is essential. Conventional methods could generate incorrect model geometry with noisy data since they are often not robust and sensitive to noise.

The DL methods have been more successful than conventional methods in many fields. Additionally, more and more city models and EO data are becoming publicly available. With the advancements in computer hardware, especially GPU technology, the DL is becoming more popular in various fields. With the growing popularity of 3D DL libraries such as PyTorch3D (Ravi et al., 2020), Tensorflow3D (Google Research, 2021) and Nvidia Kaolin (Nvidia, 2021), DL studies using 3D data will become more popular after their success in 2D image studies. 3D DL libraries can directly learn the 3D object models to reconstruct the model. As a result, 3D city models and buildings can be automatically generated without having to rely on any specific roof libraries.

Using the DL methods, different tasks such as building detection, building segmentation, footprint extraction and 3D reconstruction can be performed consecutively on different data sets and a full automation may be possible. Many countries and municipalities are offering 3D city models in different LoDs through open data portals. The data of different 3D city models or DSMs can be used to train the DL models and to automate the production. Consequently, global-applicable models can be produced with a higher level of accuracy. Furthermore, deep learning can also be used to automatically extract roof segment lines from aerial or satellite imagery, which can then be used to generate a 3D model of the building.

With the availability of EO data, the city models and point clouds, city-scale automatic 3D reconstruction with DL methods will be a research topic that will be actively studied in the coming years, especially semantic 3D city models. It may even be possible to produce global models that are not specific to a particular area if millions of building data are used to train a DL model. In the next few years, it is also expected that open-source 3D DL libraries such as PyTorch 3D, Tensorflow3D, and Nvidia Kaolin will facilitate the 3D reconstruction and enable more research.

## ACKNOWLEDGEMENTS

Mehmet Buyukdemircioglu acknowledges the support of the Scientific and Technological Research Council of Turkey (TUBITAK) 2214-A International Doctoral Research Fellowship Programme, 2224-A Grant Program for Participation in Scientific Meetings Abroad and Higher Education Council of Turkey (YÖK) for 100/2000 Ph.D. research scholarship. The paper is part of the Ph.D. thesis research of Mehmet Buyukdemircioglu.

## REFERENCES

- Agoub, A., Schmidt, V., & Kada, M., 2019. Generating 3D City Models Based on the Semantic Segmentation of Lidar Data Using Convolutional Neural Networks. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, 4, 3-10. <https://doi.org/10.5194/isprs-annals-IV-4-W8-3-2019>
- 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(19), 2219. <https://doi.org/10.3390/rs11192219>
- Alidoost, F., Arefi, H., & Hahn, M., 2020. Y-Shaped Convolutional Neural Network for 3d Roof Elements Extraction to Reconstruct Building Models from A Single Aerial Image. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(2). <https://doi.org/10.5194/isprs-annals-V-2-2020-321-2020>
- Beer, L., 2019. Automatic generation of lod1 city models and building segmentation from single aerial orthographic images using conditional generative adversarial networks. *GI\_Forum 2019*, 7, 119-133. [https://doi.org/10.1553/giscience2019\\_01\\_s119](https://doi.org/10.1553/giscience2019_01_s119)
- Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S. and Çöltekin, A., 2015. Applications of 3D city models: State of the art review. *ISPRS International Journal of Geo-Information*, 4(4), pp.2842-2889. <https://doi.org/10.3390/ijgi4042842>
- Biljecki, F., Ledoux, H., & Stoter, J., 2017. Generating 3D city models without elevation data. *Computers, Environment and Urban Systems*, 64, 1-18. <https://doi.org/10.1016/j.compenvurbsys.2017.01.001>
- Biljecki, F., & Dehbi, Y., 2019. Raise the Roof: Towards Generating Lod2 Models Without Aerial Surveys Using Machine Learning. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4. <https://doi.org/10.5194/isprs-annals-IV-4-W8-27-2019>
- Bittner, K., Körner, M., Fraundorfer, F., & Reinartz, P., 2019. Multi-task cGAN for simultaneous spaceborne DSM refinement and roof-type classification. *Remote Sensing*, 11(11), 1262. <https://doi.org/10.3390/rs11111262>
- Bittner, K., d'Angelo, P., Körner, M., & Reinartz, P., 2018a. Automatic Large-Scale 3D Building Shape Refinement Using Conditional Generative Adversarial Networks. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 422, 103-108. <https://doi.org/10.5194/isprs-archives-XLII-2-103-2018>
- Bittner, K., d'Angelo, P., Körner, M., & Reinartz, P., 2018b. Dsm-to-lod2: Spaceborne stereo digital surface model refinement. *Remote Sensing*, 10(12), 1926. <https://doi.org/10.3390/rs10121926>
- Bizjak M, Žalik B, Lukač N., 2021. Parameter-Free Half-Spaces Based 3D Building Reconstruction Using Ground and Segmented Building Points from Airborne LiDAR Data with 2D Outlines. *Remote Sensing*; 13(21):4430. <https://doi.org/10.3390/rs13214430>

- Buyukdemircioglu, M., Kocaman, S., & Isikdag, U., 2018. Semi-automatic 3D city model generation from large-format aerial images. *ISPRS International Journal of Geo-Information*, 7(9), 339. <https://doi.org/10.3390/ijgi7090339>
- Buyukdemircioglu, M. and Kocaman, S., 2018. A 3D campus application based on city models and WebGL, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-5, 161–165, <https://doi.org/10.5194/isprs-archives-XLII-5-161-2018>
- Buyukdemircioglu, M. and Kocaman, S., 2020. Reconstruction and efficient visualization of heterogeneous 3D city models. *Remote Sensing*, 12(13), 2128. <https://doi.org/10.3390/rs12132128>
- Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., ... & Xie, Z., 2018. Deep learning and its applications in biomedicine. *Genomics, proteomics & bioinformatics*, 16(1), 17-32. <https://doi.org/10.1016/j.gpb.2017.07.003>
- Chen, Z., Khademi, S., Ledoux, H., & Nan, L., 2021. Reconstructing Compact Building Models from Point Clouds Using Deep Implicit Fields. *arXiv preprint*, <https://doi.org/10.48550/arXiv.2112.13142>
- Dehbi, Y., Hadiji, F., Gröger, G., Kersting, K., & Plümer, L., 2017. Statistical relational learning of grammar rules for 3D building reconstruction. *Transactions in GIS*, 21(1), 134-150. <https://doi.org/10.1111/tgis.12200>
- Döllner, J., 2020. Geospatial artificial intelligence: Potentials of machine learning for 3d point clouds and geospatial digital twins. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1), 15-24. <https://doi.org/10.1007/s41064-020-00102-3>
- Drešček, U., Kosmatin Fras, M., Tekavec, J., & Liseč, A., 2020. Spatial ETL for 3D building modelling based on unmanned aerial vehicle data in semi-urban areas. *Remote Sensing*, 12(12), 1972. <https://doi.org/10.3390/rs12121972>
- Farella, E. M., Özdemir, E., & Remondino, F., 2021. 4D Building Reconstruction with Machine Learning and Historical Maps. *Applied Sciences*, 11(4), 1445. <https://doi.org/10.3390/app11041445>
- Fischler, M. A., & Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395. <https://doi.org/10.1145/358669.358692>
- Google Research, 3D Scene Understanding with TensorFlow 3D, 2021. <https://github.com/google-research/google-research/tree/master/tf3d> (accessed on 01.04.2022).
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27. <https://arxiv.org/abs/1406.2661v1>
- Gui, S., & Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 1-19. <https://doi.org/10.1016/j.isprsjprs.2021.08.025>
- Haala, N., & Kada, M., 2010. An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 570-580. <https://doi.org/10.1016/j.isprsjprs.2010.09.006>
- Heipke, C., & Rottensteiner, F., 2020. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geo-spatial Information Science*, 23(1), 10-19. <https://doi.org/10.1080/10095020.2020.1718003>
- Henn, A., Gröger, G., Stroh, V., & Plümer, L., 2013. Model driven reconstruction of roofs from sparse LIDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 76, 17-29. <https://doi.org/10.1016/j.isprsjprs.2012.11.004>
- Hoeser, T., & Kuenzer, C., 2020. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. *Remote Sensing*, 12(10), 1667. <https://doi.org/10.3390/rs12101667>
- Kada, M. and Wichmann, A., 2012. Sub-Surface Growing and Boundary Generalization for 3D Building Reconstruction, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, I-3, 233–238, <https://doi.org/10.5194/isprsannals-I-3-233-2012>
- Kaolin: A Pytorch Library for Accelerating 3D Deep Learning Research, 2021. <https://github.com/NVIDIAGameWorks/kaolin> (accessed on 01.04.2022).
- Kapoor, A., Larco, H., & Kiveris, R., 2019. Nostalgin: Extracting 3d city models from historical image data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2565-2575). <https://arxiv.org/abs/1905.01772>
- Kelly, T., Guerrero, P., Steed, A., Wonka, P., & Mitra, N. J., 2018. FrankenGAN: guided detail synthesis for building mass models using style-synchronized GANs. *ACM Transactions on Graphics (TOG)*, 37(6), 1-14. <https://doi.org/10.1145/3272127.3275065>
- Kippers, R. G., Koeva, M., van Keulen, M., & Elberink, S. O., 2021. Automatic 3D building model generation using deep learning methods based on CityJSON and 2D floor plans. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46, 4. <https://doi.org/10.5194/isprs-archives-XLVI-4-W4-2021-49-2021>
- Knyaz, V. A., Kniaz, V. V., Remondino, F., Zheltov, S. Y., & Gruen, A., 2020. 3D reconstruction of a complex grid structure combining UAS images and Deep Learning. *Remote Sensing*, 12(19), 3128. <https://doi.org/10.3390/rs12193128>
- LeCun, Y., Bengio, Y., & Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Leotta, M. J., Long, C., Jacquet, B., Zins, M., Lipsa, D., Shan, J., ... & Dana, K., 2019. Urban semantic 3D reconstruction from multiview satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0). <https://doi.org/10.1109/CVPRW.2019.00186>

- Li, W., Meng, L., Wang, J., He, C., Xia, G. S., & Lin, D., 2021. 3D Building Reconstruction from Monocular Remote Sensing Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 12548-12557). <https://doi.org/10.1109/ICCV48922.2021.01232>
- Li, Y. and Wu, B., 2020. Automatic 3D Reconstruction of Complex Buildings from Incomplete Point Clouds with Topological-Relation Constraints, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, V-5-2020, 85–92, <https://doi.org/10.5194/isprs-annals-V-5-2020-85-2020>
- Li, Z., & Shan, J., 2022. RANSAC-based multi primitive building reconstruction from 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 185, <https://doi.org/10.1016/j.isprsjprs.2021.12.012>
- Liu, C., Kong, D., Wang, S., Wang, Z., Li, J., & Yin, B., 2021. Deep3D reconstruction: methods, data, and challenges. *Frontiers of Information Technology & Electronic Engineering*, 22(5), 652-672. <https://doi.org/10.1631/FITEE.2000068>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152, 166-177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Mahmud, J., Price, T., Bapat, A., & Frahm, J. M., 2020. Boundary-aware 3D building reconstruction from a single overhead image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 441-451). <https://doi.org/10.1109/CVPR42600.2020.00052>
- Malihi, S., Valadan Zoej, M. J., & Hahn, M., 2018. Large-scale accurate reconstruction of buildings employing point clouds generated from UAV imagery. *Remote Sensing*, 10(7), 1148. <https://doi.org/10.3390/rs10071148>
- Muftah, H., Rowan, T. S. L., & Butler, A. P., 2021. Towards open-source LOD2 modelling using Convolutional Neural Networks. *Modeling Earth Systems and Environment*, 1-17. <https://doi.org/10.1007/s40808-021-01159-8>
- Murtiyoso, A., Veriandi, M., Suwardhi, D., Soeksmantono, B., & Harto, A. B., 2020. Automatic Workflow for Roof Extraction and Generation of 3D CityGML Models from Low-Cost UAV Image-Derived Point Clouds. *ISPRS International Journal of Geo-Information*, 9(12), 743. <https://doi.org/10.3390/ijgi9120743>
- Musialski, P., Wonka, P., Aliaga, D. G., Wimmer, M., Van Gool, L., & Purgathofer, W., 2013. A survey of urban reconstruction. In *Computer graphics forum* (Vol. 32, No. 6, pp. 146-177). <https://doi.org/10.1111/cgf.12077>
- Nan, L., & Wonka, P., 2017. Polyfit: Polygonal surface reconstruction from point clouds. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2353-2361). <https://doi.org/10.1109/ICCV.2017.258>
- Nishida, G., Bousseau, A., & Aliaga, D. G., 2018. Procedural modeling of a building from a single image. In *Computer graphics forum* (Vol. 37, No. 2, pp. 415-429). <https://doi.org/10.1111/cgf.13372>
- Park, Y., & Guldman, J. M., 2019. Creating 3D city models with building footprints and LIDAR point cloud classification: A machine learning approach. *Computers, environment and urban systems*, 75, 76-89. <https://doi.org/10.1016/j.compenvurbsys.2019.01.004>
- 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(14), 1660. <https://doi.org/10.3390/rs11141660>
- Pepe, M., Costantino, D., Alfio, V. S., Voza, G., & Cartellino, E., 2021. A Novel Method Based on Deep Learning, GIS and Geomatics Software for Building a 3D City Model from VHR Satellite Stereo Imagery. *ISPRS International Journal of Geo-Information*, 10(10), 697. <https://doi.org/10.3390/ijgi10100697>
- Peters, R., Dukai, B., Vitalis, S., van Liempt, J., & Stoter, J., 2021. Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. *arXiv preprint* <https://doi.org/10.48550/arXiv.2201.01191>
- Qi, C. R., Yi, L., Su, H., & Guibas, L. J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *NeurIPS 2017*, <https://arxiv.org/abs/1706.02413>
- Qian, Y., Zhang, H., & Furukawa, Y., 2021. Roof-GAN: learning to generate roof geometry and relations for residential houses. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2796-2805). <https://arxiv.org/abs/2012.09340>
- Qian, Z., Chen, M., Zhong, T., Zhang, F., Zhu, R., Zhang, Z., ... & Lü, G., 2022. Deep Roof Refiner: A detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 107, 102680. <https://doi.org/10.1016/j.jag.2022.102680>
- Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W. Y., Johnson, J., & Gkioxari, G., 2020. Accelerating 3D Deep Learning with PyTorch3D. *arXiv preprint* <https://arxiv.org/abs/2007.08501>
- Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U., & Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS journal of photogrammetry and remote sensing*, 93, 256-271. <https://doi.org/10.1016/j.isprsjprs.2013.10.004>
- Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., and Breitkopf, U., 2012. The ISPRS Benchmark On Urban Object Classification and 3D Building Reconstruction, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, I-3, 293–298, <https://doi.org/10.5194/isprsannals-I-3-293-2012>
- Simonyan, K., & Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://arxiv.org/abs/1409.1556v6>
- Teo, T.A., 2019. Deep-Learning for LoD1 Building Reconstruction from Airborne Lidar Data. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing*

*Symposium* (pp. 86-89). *IEEE*.  
<https://doi.org/10.1109/IGARSS.2019.8897810>

Wang, K., & Frahm, J. M., 2017. Single view parametric building reconstruction from satellite imagery. In *2017 International Conference on 3D Vision (3DV)* (pp. 603-611). *IEEE*. <https://doi.org/10.1109/3DV.2017.00074>

Wichmann, A., Agoub, A., & Kada, M., 2018. RoofN3D: Deep Learning Training Data for 3D Building Reconstruction. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2).  
<https://doi.org/10.5194/isprs-archives-XLII-2-1191-2018>

Xie, L., Hu, H., Zhu, Q., Li, X., Tang, S., Li, Y., ... & Wang, W., 2021. Combined rule-based and hypothesis-based method for building model reconstruction from photogrammetric point clouds. *Remote Sensing*, 13(6), 1107.  
<https://doi.org/10.3390/rs13061107>

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. *arXiv preprint*  
<https://doi.org/10.48550/arXiv.2005.09223>

Yu, D., Wei, S., Liu, J., & Ji, S., 2020. Advanced Approach for Automatic Reconstruction of 3D Buildings from Aerial Images. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 43.  
<https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-541-2020>

Yu, D., Ji, S., Liu, J., & Wei, S., 2021. Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171, 155-170.  
<https://doi.org/10.1016/j.isprsjprs.2020.11.011>

Zeng, H., Wu, J., & Furukawa, Y., 2018. Neural procedural reconstruction for residential buildings. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 737-753). [https://doi.org/10.1007/978-3-030-01219-9\\_45](https://doi.org/10.1007/978-3-030-01219-9_45)

Zhang, L., & Zhang, L., 2018. Deep learning-based classification and reconstruction of residential scenes from large-scale point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4), 1887-1897.  
<https://doi.org/10.1109/TGRS.2017.2769120>

Zhang, L., Li, Z., Li, A., & Liu, F., 2018. Large-scale urban point cloud labeling and reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138, 86-100.  
<https://doi.org/10.1016/j.isprsjprs.2018.02.008>

Zhang, W., Li, Z., & Shan, J., 2021. Optimal Model Fitting for Building Reconstruction from Point Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9636-9650.  
<https://doi.org/10.1109/JSTARS.2021.3110429>

Zhao, C., Zhang, C., Yan, Y., & Su, N., 2021. A 3D Reconstruction Framework of Buildings Using Single Off-Nadir Satellite Image. *Remote Sensing*, 13(21), 4434.  
<https://doi.org/10.3390/rs13214434>