UNB Off-site Construction Research Centre

CSCE Canadian Society for Civil Engineering

Transforming Construction with Reality Capture Technologies: The Digital Reality of Tomorrow

August 23-25, 2022, Fredericton, New Brunswick, Canada

DEEP NEURAL NETWORKS TO ASSIST IN BIM CREATION USING SCANNED DATA: A REVIEW

Seydgar, M[1*], Motamedi, A[1], and Poirier, É.A[1]

[1] Department of Construction Engineering, École de Technologie Supérieure, Université du Québec, Canada

[*] majid.seydgar.1@ens.etsmtl.ca

Abstract: Deep neural networks (DNNs) have been revolutionizing various 3D processing fields, such as autonomous driving and augmented reality. They also have a great potential to facilitate building information modeling (BIM) based on 3D scanned data, a process known as Scan-to-BIM. Several studies have investigated the potential of DNNs to improve the Scan-to-BIM pipeline. However, most of the existing studies are high-level overviews of DNN applications. In this study, an in-depth investigation of the DNNs' capacities in improving 3D reconstruction, object detection, and object parametrization, which serve as the core components of Scan-to-BIM, is provided. Our investigation is performed by reviewing the most relevant studies of the computer vision and construction literatures to gain a comprehensive view of both the state-of-the-art processing methods of scanned data, as well as the progress of automated BIModeling using Industry Foundation Class (IFC) objects. Based on the reviewed state-of-the-art studies, current challenges and limitations are discussed to identify further avenues for research.

Keywords: Deep Neural Networks (DNN); Industry Foundation Classes (IFC); Building Information Modeling (BIM); Object classification; Scan-to-BIM

1 INTRODUCTION

With the emergence of building information modeling (BIM), there has been a surge of interest in harnessing digital technology in the Architecture, Engineering, and Construction and Facilities Management (AEC/FM) domain. BIM and its representation schemas, such as Industry Foundation Classes (IFC), enable AEC stakeholders to store and share various types of building information in a more efficient and integrated manner. BIM can also be enriched with various data sources, e.g., Internet of Things (IoT) sensors, to be applied in advanced Digital Twin's (DT) applications (Pan and Zhang 2021; Turner et al. 2020; Boje et al. 2020). For roughly two decades, the methods used to create BIM models has been addressed in the literature (Volk, Stengel, and Schultmann 2014). Early methods mostly relied on manual modeling and beyond their requiring labor-intensive procedures, they also require expert knowledge (Tang et al. 2010). In recent years, with the advancement of computer vision-based (CV) methods, various methods have been introduced to automatically create BIM models (M. Huang, Ninić, and Zhang 2021). The CV methods not only help achieve more cost-effective BIModeling solutions, they also greatly support the modeling task (Bloch and Sacks 2018). These methods can be used to construct BIM from scanned data (Bosché et al. 2015). The creation of models using scanned data, e.g., point cloud (PC), has become popular because the BIModeling can be effectively performed using this data source without any prior information (M. Huang, Ninić, and Zhang 2021). The methods that apply scanned data to construct BIM are mostly known as Scan-to-BIM methods. In recent years, several studies have provided comprehensive reviews of the DNN applications in BIM related projects (Czerniawski and Leite 2020; Zabin et al. 2022; 2022). However, most of the reviews related to the application of DNNs for Scan-to-BIM are limited, providing only an overview. Thus, this study aims to provide an in-depth review of the studies related to the use of DNNs for Scan-to-BIM to facilitate the identification of possible research avenues for more efficient BIModeling using DNNs.

2 CATEGORIES OF SCAN-TO-BIM SOLUTIONS

Typically, the procedure of Scan-to-BIM is comprised of four main sub-tasks, namely, data collection, point cloud registration (i.e., 3D reconstruction), object detection, and, BIModeling (i.e., semantic enrichment) (Czerniawski and Leite 2020). Figure 1 provides an illustration of the mentioned sub-tasks for a scene of the ScanNet dataset (Dai et al. 2017). As shown in Figure 1, the scanning task is first performed to collect the 3D information (i.e., points) of the scene. The 3D point sets are then registered in respect to each other to reconstruct a 3D scene from the collected data. The 3D reconstructed scene is then used for the 3D object detection task. Lastly, the BIM is created using the detected objects of the scenes.

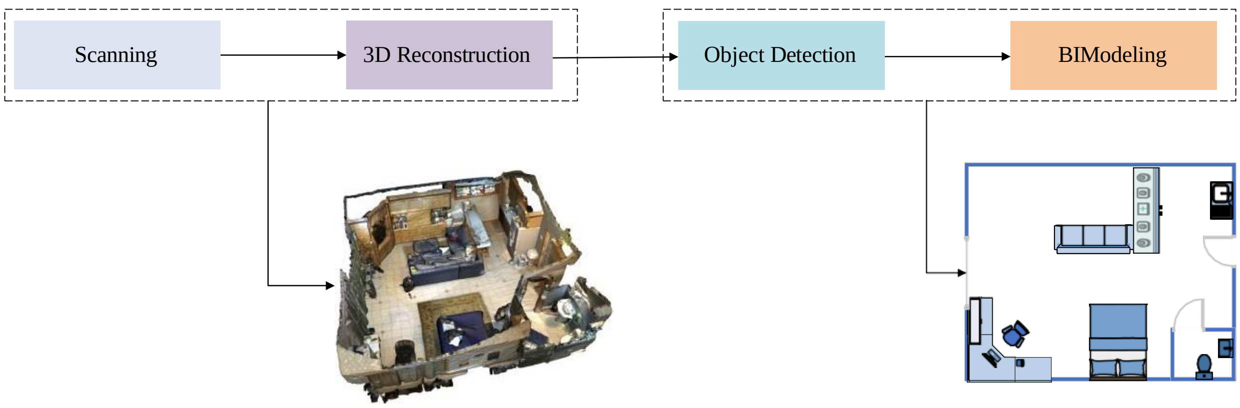

Figure 1: Procedure of Scan-to-BIM for a scene of the ScanNet dataset (Dai et al. 2017).

Scan-to-BIM solutions have been widely developed using various methods. The methods can be broadly categorized as rule-based (Jung et al. 2018) or machine learning (ML)-based methods (B. Wang et al. 2022). The rule-based methods can be considered as conventional approaches where the BIModel is created by detecting a set of primitives from the scanned data in the first stage. The primitives are analyzed by a set of rules to join or disjoin them and to construct the elements. Despite reasonable performances, rule-based methods lack high adaptability in different data structures and require extensive user interventions (Bloch and Sacks 2018). ML methods provide more efficient BIModeling solutions by transforming the conventional procedure of the Scan-to-BIM process into an automated and seamless one. The ML methods mainly rely on DNNs, such as feed-forward or recurrent neural networks, which can learn discriminative features from data, to automatically perform various tasks, such as 3D object detection, and PC registration. However, the DNNs come with their own challenges, such as relying on large scale labeled datasets for the supervision (Beyer and Dai 2022), and a high computational burden (Emunds et al. 2022). In this study, we mainly focus on reviewing the advancements and challenges of DNN methods for 3D data interpretation and understanding, which can be applied for the Scan-to-BIM task. To achieve this, we reviewed the related studies for each of the Scan-to-BIM sub-tasks in the construction and CV literature, as most of the DNN-based Scan-to-BIM methods are inspired by the related CV studies.

2.1 Datasets

One of the major concerns regarding the utilization of DNNs is the requirement of annotated data. It is vital for DNNs to achieve promising performance in 3D reconstruction tasks (Dai et al. 2017). In recent years, various datasets have been introduced in the literature to facilitate the application of 3D understanding. The datasets can be broadly categorized as synthetic-based or scanned-based.

Synthetic datasets are mostly constructed by manual modeling (Zhao and Vela 2019; Koo, Jung, and Yu 2021), or artificial intelligence (AI)-simulation engines (Puig et al. 2018; Szot et al. 2021). For example, the manual modeling scheme is used in (Koo, Jung, and Yu 2021) to collect 824 IFC-based objects of walls and doors and to annotate them based on their sub-types (e.g., generic wall, wall with openings). More comprehensive non-IFC datasets, such as Structured3D (Zheng et al. 2020) provide synthetic images of 3,500 scenes from multiple rooms with various objects. InteriorNet (W. Li et al. 2018) introduces a CAD- based synthetic dataset that contains millions of interior layouts and furniture models. House3D (Y. Wu et al. 2018) is a volumetric synthetic dataset with 45,000 annotated scenes with various types of objects. Moreover, several studies address 3D dataset creation for industrial scenes using the manual modeling scheme. For example, a bridge component modeling is performed in (Zhao and Vela 2019) to construct the annotated training dataset for 3D object classification. Despite the considerable progress in the creation of synthetic datasets, their capacity to represent real world cluttered scenes is questionable.

Scanned-based datasets can be created using mobile and terrestrial scanning devices. In the construction literature, most of the introduced datasets are created using terrestrial laser scanners (Yin et al. 2021; Agapaki and Brilakis 2021) due to their geometrical accuracy for BIModeling (Esfahani et al. 2021; Xiong et al. 2013). However, most of the developed datasets are not publicly accessible. In the CV literature, a few datasets are publicly available for the tasks of point cloud registration and 3D object classification. For instance, ScanNet (Dai et al. 2017) uses RGB-D images to represent 1,500 scenes in a mesh format with camera pose information, CAD models, and label annotations. The RGB-D sensors are used in several other datasets, such as Depth V2 (Silberman et al. 2012) and Matterport3D (Chang et al. 2017), for 3D mesh-based representations of indoor environments. Panoramic RGB images are also applied in the Stanford 2D-3D-S dataset (Armeni et al. 2017) to represent the 3D mesh information of six large indoor scenes. Table 1 compares the prominent 3D datasets' creation methods and their parametric attributes.

Table 1: Comparison of well-known 3D indoor datasets.

Dataset	Type	CAD models	Data Type	# of Scenes
ScanNet (Dai et al. 2017)	Scanned	296	Mesh	1506
Matterport3D (Chang et al. 2017)	Scanned	N/A	Mesh	2056
SUN3D (Xiao, Owens, and Torralba 2013)	Scanned	N/A	PC	254
Scan2CAD (Avetisyan et al. 2019)	Scanned	3049	Mesh	1506
OpenRooms (Z. Li et al. 2020)	Scanned	2500	Mesh	1068
Depth v2 (Silberman et al. 2012)	Scanned	N/A	RGB-D	N/A
Hypersim (Roberts et al. 2021)	Synthetic	N/A	RGB-D	N/A
3D-FRONT (Fu et al. 2021)	Synthetic	13151	Mesh	18968
SceneNet (Handa et al. 2016)	Synthetic	57	Mesh	57
InteriorNet (W. Li et al. 2018)	Synthetic	N/A	RGB-D	N/A
House-3D (Y. Wu et al. 2018)	Synthetic	N/A	RGB-D	45K
Structured3D (Zheng et al. 2020)	Synthetic	N/A	3D layout	3500

2.2 Point Cloud Registration

Matching 3D geometry has been a long-standing work in progress and it has numerous applications, such as 3D reconstruction, object retrieval, and object pose estimation (Zeng et al. 2017). In this study, we mainly focus on geometric registration of RGB-D data because after the prevalence of using handheld 3D scanning sensors with low-density depth sensors, more studies have focused on extracting robust semantic and geometric feature descriptors. The reason for this is that the visual information provided by RGB sensors, provides the opportunity to extract semantic features along with geometric features and perform more effective point-set registration. Figure 2 shows the general pipeline of the geometric matching task for most of the methods that use RGB-D sensors in the registration task. Considering two sets of point clouds with the RGB data, denoted as Input and Reference, the main assumption for geometric registration and matching is that the Input is transformed from the Reference by a spatial (e.g., rigid) transformation. Thus, the aim of geometric registration is to calculate the transformation such that the mean-squared error of the two sets' alignment converges to zero (Pomerleau, Colas, and Siegwart 2015). More specifically, the 3D- registration approaches mostly use a feature extractor to extract meaningful semantic and geometric descriptors, which enables the methods to improve the transformation estimation task. This is because estimating rigid transformation can be challenging in various scenarios, such as in low overlaps and different densities scenarios (Qin et al. 2022; X. Huang, Mei, and Zhang 2020).

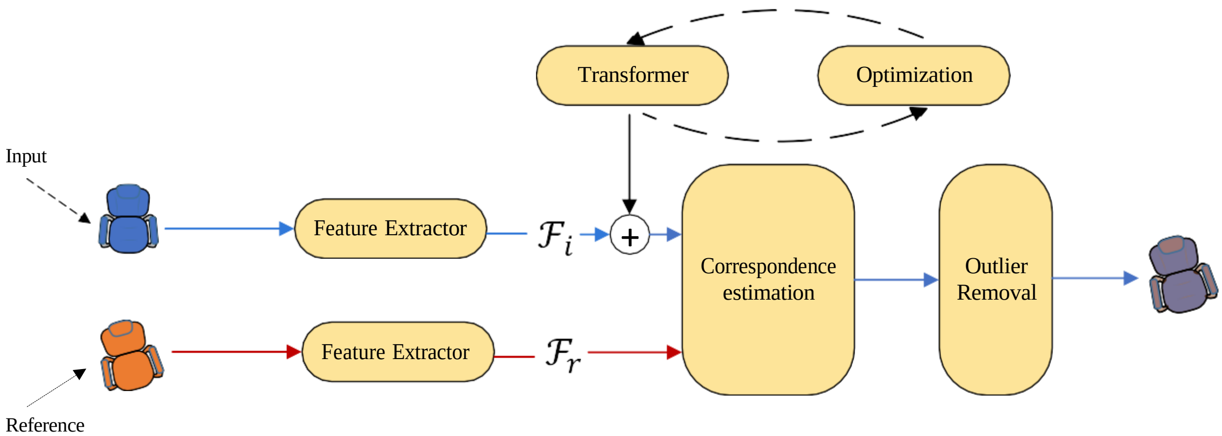

Figure 2: The general overview of the geometric matching pipeline.

Early approaches extract simple feature descriptors to find visual correspondences among RGB data, such as using corners for stereo matching (Moravec 1981), geometric primitives (Rabbani et al. 2007) and local 3D geometric descriptors as geometric features (Johnson and Hebert 1999; Rusu, Blodow, and Beetz 2009). However, considering the point cloud registration challenges, such as low density and partial overlap problems, obtaining best alignment between point sets is difficult using these conventional methods. In recent years, plenty of DNNs-based methods have attempted to extract robust feature descriptors using visual features (Dusmanu et al. 2019; DeTone, Malisiewicz, and Rabinovich 2018; El Banani, Gao, and Johnson 2021; Yi et al. 2016) and geometrical features (Choy et al. 2016; Deng, Birdal, and Ilic 2018; Bai et al. 2020). The visual descriptors, extracted by the DNNs, typically contain abstract semantic features from the visual information of the input RGB pixels, which can be used to estimate reliable correspondence points. For instance, Choy et al. (2016) applied a supervised recurrent neural network (RNN) to learn 3D geometric shapes from a collection of synthetic images (Choy et al. 2016). In (Deng, Birdal, and Ilic 2018) an autoencoder-based method is proposed to learn 3D local descriptors in an unsupervised manner. Additionally, a 3D feature detector and descriptor is proposed in (Bai et al. 2020) which uses a fully convolutional neural network (FCNN). Regarding the learning of invariant visual features using geometric transformation, CNN models are applied for both local feature detectors and descriptors (Yi et al. 2016; Dusmanu et al. 2019). (DeTone, Malisiewicz, and Rabinovich 2018) addressed the problem of correspondence estimation in multiple-view images by proposing a self-supervised point detector and descriptor. In addition, a number of studies applied RGB-D data for extracting visual descriptors (Zeng et al. 2017; Xiao, Owens, and Torralba 2013). Although using the DNNs showed promising performance, a number of challenges still exist in the geometric registration tasks, such as high dimensional parameter space (W. Chen et al. 2022), reliance of the state-of-the-art methods on labeled data (El Banani, Gao, and Johnson 2021), poor performance on low-textured spaces (Z. Chen et al. 2022) , and complexities of transformation of the scans with partial or limited overlap (Qin et al. 2022; Sun et al. 2022).

2.3 3D-Object Detection

3D-object detection is a preliminary stage for various downstream tasks, such as semantic enrichment (Werbrouck et al. 2020) and CAD-based element fitting (Kuo et al. 2020). The aim of 3D object detection is to determine the type and location of various objects based on scanned data (Charles R. Qi et al. 2019). Figure 3 illustrates the results of object detection on a 3D scene from the ScanNet dataset (Dai et al. 2017).

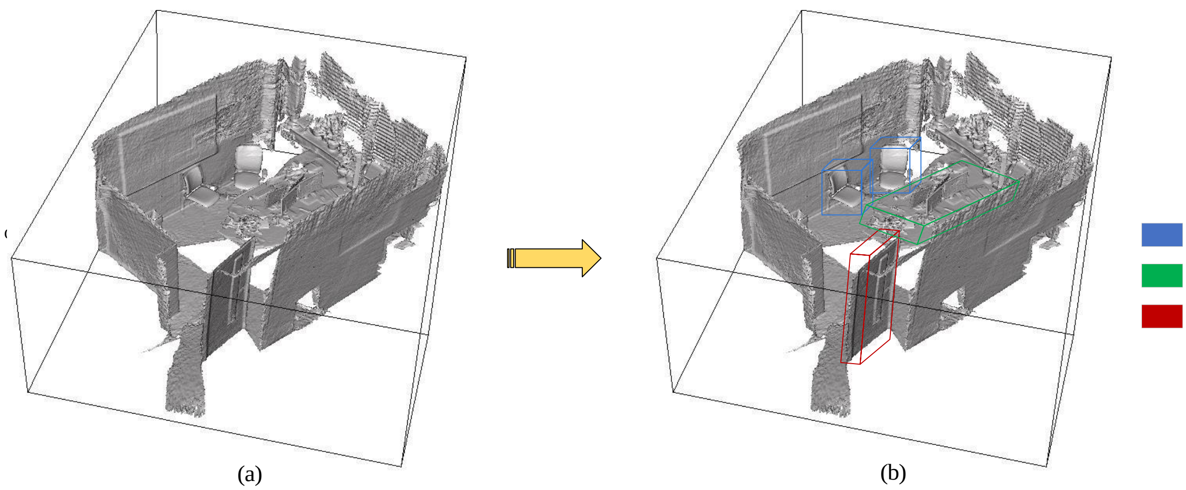

Figure 3: Example of 3D-object detection using a ScanNet dataset scene (Dai et al. 2017) a) Input point cloud b) Object detection result.

Object detection from scanned data differs from object detection using other data sources (e.g., RGB images) due to the disorderedness (i.e., irregular) of the point clouds. Considerable efforts have been made for automatic 3D-object detection in outdoor and indoor scenes using DNNs (H. Wang et al. 2022; Liang et al. 2019). In this study, the indoor-based 3D-object detection and classification methods were focused. These methods can be broadly categorized as either projection-based or raw PC-based analysis methods. Projection-based methods mostly convert the scanned data into a regular data structure, such as images (B. Wang et al. 2022) and voxels (Song and Xiao 2016). Several studies have investigated the projection- based methods for IFC object classification. For instance, multi-view convolutional neural networks (MVCNN) (Su et al. 2015), which performs a bird's eye view projection on PC, is applied in (Koo, Jung, and Yu 2021; Park and Ergan 2022; Park, Ergan, and Feng 2021) to recognize and classify IFC objects in indoor environments. These projection-based methods enable DNN architectures, such as convolutional neural networks (CNNs) to directly process the scene information and achieve reasonable results. For example, the state-of-the-art performance for IFC MVCNN object classification was reported in (Emunds et al. 2021). However, a loss of information, such as dimensions and geometric structures, is inevitable in these projection methods.

The raw PC-based approach includes methods that directly perform object detection using the raw point sets collected from the scene (Charles R. Qi et al. 2018; 2019; Shi, Wang, and Li 2019). Early studies in this area were inspired by Standford University's PointNet network (Charles R Qi et al. 2017) and its variations (Charles Ruizhongtai Qi et al. 2017; Jiang et al. 2019) to estimate an object's bounding box using a hierarchical DNN architecture for raw points analysis. For instance, Point R-CNN (Shi, Wang, and Li 2019) applies a region pooling strategy to extract features from a group of points and then estimate the 3D bounding boxes. Fast Point R-CNN (Y. Chen et al. 2019) further enhanced the Point R-CNN pipeline by using the voxel-based representation in the framework's warm-up stage (i.e., pre-training). More recent studies use probabilistic estimation (Meyer et al. 2019), voting strategy (Charles R. Qi et al. 2019), and attention mechanism (Liu et al. 2021) to extract features from a group of points. LaserNet (Meyer et al.

2019) estimates a probabilistic distribution for each point using a fully CNN architecture, and then fuses the distributions to construct 3D bounding boxes of the objects. VoteNet (Charles R. Qi et al. 2019) and ImVotenet (Charles R. Qi et al. 2020) use Hough transform (Ballard 1981) as a voting strategy to detect 3D objects based on the DNN predictions for the point sets. Group-Free (Liu et al. 2021) deploys an attention mechanism to group point sets and then detect objects within the groups. The aforementioned studies show remarkable efforts in the improvement of point grouping strategies, which is a core component of the 3D- object detection. However, various concerns remain, such as sensitivity to the foreground and irrelevant point sets and high memory requirements for large point set processing, which are discussed in these very recent studies (H. Wang et al. 2022; Agapaki and Brilakis 2021; Xie et al. 2021; Zhang et al. 2022). The reconstruction of BIM from scanned data requires both plane and complex object detection. DNN- based object detection is a major step toward the creation of a rich BIModel from scanned data of a cluttered scene containing complex objects. However, plane objects (e.g., ceiling, floor, and walls) must still be detected for accurate geometric modeling. Plane object detection has been used for semi-automated Scan- to-BIM solutions for many years. Random sample consensus (RANSAC) (Fischler and Bolles 1981) and region growing-based methods (Adams and Bischof 1994; Vieira and Shimada 2005) are among the commonly used approaches for plane object detection. RANSAC is a lightweight method that uses randomly selected points from the point sets to perform primitive fitting. RANSAC's procedure of primitive fitting is straightforward. However, the accuracy of the fitting outcome is not guaranteed and can be incomplete due to various noise effects. To tackle this issue, advanced methods attempt to control the fitting procedure of primitives by using different strategies. For instance, a B-splines geometric representation algorithm is used in (Rausch and Haas 2021) to fit the primitives to the point sets more effectively. A quantization strategy is used in (Shao et al. 2021) to smooth the point sets and improve the fitting outcome. A primitive fitting strategy based on the vector normal is applied in (Schnabel, Wahl, and Klein 2007) to improve the fitting outcome. These methods considerably succeeded in improving the primitive fittings. However, extensive hyper-parameter tuning/selection is a typical requirement of these methods. Region growing is another well-known approach to detect plane objects by segmenting the point set (Ning et al. 2009; Deschaud and Goulette 2010; Vo et al. 2015). Region growing-based methods also achieved reasonable results for plane object detection. However, high computational burden and sensitivity to noise are notable drawbacks of these approach (Teboul et al. 2010; Boulaassal et al. 2007). To address these limitations, one approach is to use graph-based DNNs (Landrieu and Boussaha 2019; Landrieu and Simonovsky 2018). Graph-based DNNs are able to consider geometrical structure and object relations in 3D-object detection, which is critical for accurate plane object detection (Kim and Kim 2021).

2.4 Semantic Enrichment and BIModeling

Despite showing increasingly improved results for DNN-based methods for 3D reconstruction and 3D-object detection, the semantic enrichment of the reconstructed scene remains a fundamental limitation for DNN- based Scan-to-BIM methods. One way to address this is to retrieve a set of parametric BIM-based objects from a stored dataset and fit them to the detected objects of the scene. If the parametric objects successfully match with the objects of the scene, a full parametric BIM can be replaced. However, the successful alignment and fitting of parametric objects is a challenge. A great number of rule-based methods have been proposed in the literature to address the problem of fitting parametric elements to scanned data (Romero- Jarén and Arranz 2021; Yang et al. 2019; Jung et al. 2018; Bassier and Vergauwen 2020; K. Wu, Shi, and Ahmed 2020; Thomson and Boehm 2015). Typically, these methods rely on handcrafted geometric features for the fitting task, e.g., features based on point pairs (Drost and Ilic 2012). However, these handcrafted features severely limit the feasible application of rule-based methods for various reconstructed scenes with different shapes and levels of complexity (Bloch and Sacks 2018).

More recently, DNNs have been applied for the task of parametric object fitting. Unlike the rule-based methods, DNNs are capable of explicitly learning geometric features from the scanned data (Avetisyan et al. 2019). A number of studies attempted to parametrize 3D scenes using pre-stored CAD databases (Gümeli, Dai, and Nießner 2022; Siddiqui et al. 2021; Manni et al. 2021). For instance, (Siddiqui et al. 2021) proposed a joint embedding framework, in which the objects of the scanned data and CAD database were both mapped to a common embedding space. Similar CAD-based objects were then substituted to the objects of the scanned data to form a complete parametric model. The joint embedding approach has also been used in the backbone of other DNN-based object parameterization methods (Gümeli, Dai, and Nießner 2022; Dahnert et al. 2019). Moreover, the possibility of harnessing the joint embedding approach to a parametrized scanned scene using IFC elements is discussed in (Emunds et al. 2022). To provide a better description of the joint embedding approach which parametrizes the detected objects in the scanned data, Figure 4 provides a generic illustration of this process using ScanNet (Dai et al. 2017) and IFCNet (Emunds et al. 2021) dataset samples. Overall, the DNNs show a great potential to robustly perform semantic enrichment and parametrize the objects in the scanned data. However, the DNN-based methods remain at their infancy due to the lack of large-scale parametric datasets and challenges of processing low- texture geometric objects (Beyer and Dai 2022).

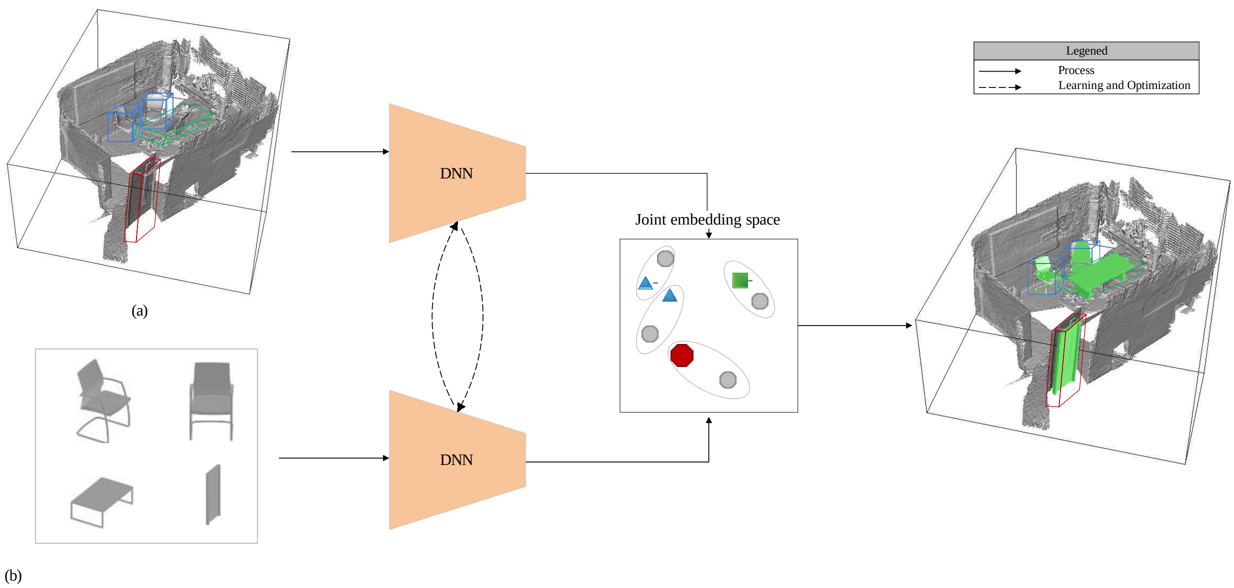

Figure 4: A generic framework, adapted from (Avetisyan et al. 2019; Emunds et al. 2022), for semantic enrichment of a scene using ScanNet dataset (Dai et al. 2017). a) the scanned data b) the pre-stored parametric dataset.

3 CONCLUSIONS

In this paper, we investigated the application of deep neural networks (DNNs) for BIModeling from scanned data, i.e., Scan-to-BIM. We focused on three sub-tasks of the Scan-to-BIM, namely point cloud registration and 3D reconstruction, object and plane detections, and BIM object fitting. Our paper reviewed the most relevant studies in both computer vision and construction literature to provide a better overview of the state- of-the-art research in the 3D interpretation of scanned data. Moreover, the current challenges of the state- of-the-art studies were discussed in detail for each sub-task of Scan-to-BIM.

According to the reviewed studies, many efforts have been made to achieve robust registration of point sets with low density and noise problems, which is generally more prevalent in handheld sensors compared to terrestrial laser scanners (TLSs). Performing robust registration using DNNs makes it possible to use handheld sensors instead of the TLS for accurate 3D reconstruction, which is the backbone of BIModeling from scanned data. The performance of DNNs is also promising for object detection and semantic enrichment tasks. Although a handful of methods took advantage of the DNNs to enhance their Scan-to- BIM pipelines, the construction community still heavily relies on conventional methods. As described in detail at the end of each section in this paper, future work will help to broaden access to the opportunities of DNN usage for more effective BIModeling from scanned data.

4 REFERENCES

Adams, Rolf, and Leanne Bischof. 1994. “Seeded Region Growing.” IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (6): 641–47.

Agapaki, Eva, and Ioannis Brilakis. 2021. “Instance Segmentation of Industrial Point Cloud Data.” Journal of Computing in Civil Engineering 35 (6): 04021022.

Armeni, Iro, Sasha Sax, Amir R Zamir, and Silvio Savarese. 2017. “Joint 2d-3d-Semantic Data for Indoor Scene Understanding.” ArXiv Preprint ArXiv:1702.01105.

Avetisyan, Armen, Manuel Dahnert, Angela Dai, Manolis Savva, Angel X. Chang, and Matthias Nießner. 2019. “Scan2cad: Learning Cad Model Alignment in Rgb-d Scans.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2614–23.

Bai, Xuyang, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. 2020. “D3feat: Joint Learning of Dense Detection and Description of 3d Local Features.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6359–67.

Ballard, Dana H. 1981. “Generalizing the Hough Transform to Detect Arbitrary Shapes.” Pattern Recognition 13 (2): 111–22.

Bassier, Maarten, and Maarten Vergauwen. 2020. “Unsupervised Reconstruction of Building Information Modeling Wall Objects from Point Cloud Data.” Automation in Construction 120: 103338.

Beyer, Tim, and Angela Dai. 2022. “Weakly-Supervised End-to-End CAD Retrieval to Scan Objects.” ArXiv Preprint ArXiv:2203.12873.

Bloch, Tanya, and Rafael Sacks. 2018. “Comparing Machine Learning and Rule-Based Inferencing for Semantic Enrichment of BIM Models.” Automation in Construction 91: 256–72.

Boje, Calin, Annie Guerriero, Sylvain Kubicki, and Yacine Rezgui. 2020. “Towards a Semantic Construction Digital Twin: Directions for Future Research.” Automation in Construction 114: 103179.

Bosché, Frédéric, Mahmoud Ahmed, Yelda Turkan, Carl T Haas, and Ralph Haas. 2015. “The Value of Integrating Scan-to-BIM and Scan-vs-BIM Techniques for Construction Monitoring Using Laser Scanning and BIM: The Case of Cylindrical MEP Components.” Automation in Construction 49: 201–13.

Boulaassal, Hakim, Tania Landes, Pierre Grussenmeyer, and Fayez Tarsha-Kurdi. 2007. “Automatic Segmentation of Building Facades Using Terrestrial Laser Data.” In ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, 65–70.

Chang, Angel, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2017. “Matterport3d: Learning from Rgb-d Data in Indoor Environments.” ArXiv Preprint ArXiv:1709.06158.

Chen, Wen, Haoang Li, Qiang Nie, and Yun-Hui Liu. 2022. “Deterministic Point Cloud Registration via Novel Transformation Decomposition.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6348–56.

Chen, Yilun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. 2019. “Fast Point R-Cnn.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 9775–84.

Chen, Zhi, Kun Sun, Fan Yang, and Wenbing Tao. 2022. “SC2-PCR: A Second Order Spatial Compatibility for Efficient and Robust Point Cloud Registration.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13221–31.

Choy, Christopher B, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 2016. “3d-R2n2: A Unified Approach for Single and Multi-View 3d Object Reconstruction.” In European Conference on Computer Vision, 628–44. Springer.

Czerniawski, Thomas, and Fernanda Leite. 2020. “Automated Digital Modeling of Existing Buildings: A Review of Visual Object Recognition Methods.” Automation in Construction 113: 103131.

Dahnert, Manuel, Angela Dai, Leonidas J. Guibas, and Matthias Nießner. 2019. “Joint Embedding of 3d Scan and Cad Objects.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 8749–58.

Dai, Angela, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. “Scannet: Richly-Annotated 3d Reconstructions of Indoor Scenes.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5828–39.

Deng, Haowen, Tolga Birdal, and Slobodan Ilic. 2018. “Ppf-Foldnet: Unsupervised Learning of Rotation Invariant 3d Local Descriptors.” In Proceedings of the European Conference on Computer Vision (ECCV), 602–18.

Deschaud, Jean-Emmanuel, and François Goulette. 2010. “A Fast and Accurate Plane Detection Algorithm for Large Noisy Point Clouds Using Filtered Normals and Voxel Growing.” In 3DPVT. Hal Archives-Ouvertes Paris, France.
DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich. 2018. “Superpoint: Self-Supervised Interest Point Detection and Description.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 224–36.
Drost, Bertram, and Slobodan Ilic. 2012. “3d Object Detection and Localization Using Multimodal Point Pair Features.” In 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, 9–16. IEEE.
Dusmanu, Mihai, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, and Torsten Sattler. 2019. “D2-Net: A Trainable Cnn for Joint Detection and Description of Local Features.” ArXiv Preprint ArXiv:1905.03561.
El Banani, Mohamed, Luya Gao, and Justin Johnson. 2021. “Unsupervisedr&r: Unsupervised Point Cloud Registration via Differentiable Rendering.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7129–39.
Emunds, Christoph, Nicolas Pauen, Veronika Richter, Jérôme Frisch, and Christoph van Treeck. 2021. “IFCNet: A Benchmark Dataset for IFC Entity Classification.” In EG-ICE 2021 Workshop on Intelligent Computing in Engineering, 166. Universitätsverlag der TU Berlin.
———. 2022. “SpaRSE-BIM: Classification of IFC-Based Geometry via Sparse Convolutional Neural Networks.” Advanced Engineering Informatics 53: 101641.
Esfahani, Mansour Esnaashary, Christopher Rausch, Mohammad Mahdi Sharif, Qian Chen, Carl Haas, and Bryan T Adey. 2021. “Quantitative Investigation on the Accuracy and Precision of Scan-to- BIM under Different Modelling Scenarios.” Automation in Construction 126: 103686.
Fischler, Martin A, and Robert C Bolles. 1981. “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.” Communications of the ACM 24 (6): 381–95.
Fu, Huan, Bowen Cai, Lin Gao, Ling-Xiao Zhang, Jiaming Wang, Cao Li, Qixun Zeng, Chengyue Sun, Rongfei Jia, and Binqiang Zhao. 2021. “3d-Front: 3d Furnished Rooms with Layouts and Semantics.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 10933–42.
Gümeli, Can, Angela Dai, and Matthias Nießner. 2022. “ROCA: Robust CAD Model Retrieval and Alignment from a Single Image.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4022–31.
Handa, Ankur, Viorica Patraucean, Vijay Badrinarayanan, Simon Stent, and Roberto Cipolla. 2016. “Understanding Real World Indoor Scenes with Synthetic Data.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4077–85.
Huang, MQ, J Ninić, and QB Zhang. 2021. “BIM, Machine Learning and Computer Vision Techniques in Underground Construction: Current Status and Future Perspectives.” Tunnelling and Underground Space Technology 108: 103677.
Huang, Xiaoshui, Guofeng Mei, and Jian Zhang. 2020. “Feature-Metric Registration: A Fast Semi-Supervised Approach for Robust Point Cloud Registration without Correspondences.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11366– 74.
Jiang, Li, Hengshuang Zhao, Shu Liu, Xiaoyong Shen, Chi-Wing Fu, and Jiaya Jia. 2019. “Hierarchical Point-Edge Interaction Network for Point Cloud Semantic Segmentation.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 10433–41.
Johnson, Andrew E., and Martial Hebert. 1999. “Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes.” IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (5): 433–49.
Jung, Jaehoon, Cyrill Stachniss, Sungha Ju, and Joon Heo. 2018. “Automated 3D Volumetric Reconstruction of Multiple-Room Building Interiors for as-Built BIM.” Advanced Engineering Informatics 38: 811–25.
Kim, Hyunsoo, and Changwan Kim. 2021. “3D As-Built Modeling from Incomplete Point Clouds Using Connectivity Relations.” Automation in Construction 130: 103855.
Koo, Bonsang, Raekyu Jung, and Youngsu Yu. 2021. “Automatic Classification of Wall and Door BIM Element Subtypes Using 3D Geometric Deep Neural Networks.” Advanced Engineering Informatics 47: 101200.

Kuo, Weicheng, Anelia Angelova, Tsung-Yi Lin, and Angela Dai. 2020. “Mask2CAD: 3D Shape Prediction by Learning to Segment and Retrieve.” In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, 260–77. Springer.

Landrieu, Loic, and Mohamed Boussaha. 2019. “Point Cloud Oversegmentation with Graph-Structured Deep Metric Learning.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7440–49.

Landrieu, Loic, and Martin Simonovsky. 2018. “Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4558–67.

Li, Wenbin, Sajad Saeedi, John McCormac, Ronald Clark, Dimos Tzoumanikas, Qing Ye, Yuzhong Huang, Rui Tang, and Stefan Leutenegger. 2018. “Interiornet: Mega-Scale Multi-Sensor Photo-Realistic Indoor Scenes Dataset.” ArXiv Preprint ArXiv:1809.00716.

Li, Zhengqin, Ting-Wei Yu, Shen Sang, Sarah Wang, Meng Song, Yuhan Liu, Yu-Ying Yeh, Rui Zhu, Nitesh Gundavarapu, and Jia Shi. 2020. “Openrooms: An End-to-End Open Framework for Photorealistic Indoor Scene Datasets.” ArXiv Preprint ArXiv:2007.12868.

Liang, Ming, Bin Yang, Yun Chen, Rui Hu, and Raquel Urtasun. 2019. “Multi-Task Multi-Sensor Fusion for 3d Object Detection.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7345–53.

Liu, Ze, Zheng Zhang, Yue Cao, Han Hu, and Xin Tong. 2021. “Group-Free 3d Object Detection via Transformers.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2949–58.

Manni, Alessandro, Damiano Oriti, Andrea Sanna, Francesco De Pace, and Federico Manuri. 2021. “Snap2cad: 3D Indoor Environment Reconstruction for AR/VR Applications Using a Smartphone Device.” Computers & Graphics 100: 116–24.

Meyer, Gregory P., Ankit Laddha, Eric Kee, Carlos Vallespi-Gonzalez, and Carl K. Wellington. 2019. “Lasernet: An Efficient Probabilistic 3d Object Detector for Autonomous Driving.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12677–86.

Moravec, Hans P. 1981. “Rover Visual Obstacle Avoidance.” In IJCAI, 81:785–90.

Ning, Xiaojuan, Xiaopeng Zhang, Yinghui Wang, and Marc Jaeger. 2009. “Segmentation of Architecture Shape Information from 3D Point Cloud.” In Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry, 127–32.

Pan, Yue, and Limao Zhang. 2021. “A BIM-Data Mining Integrated Digital Twin Framework for Advanced Project Management.” Automation in Construction 124: 103564.

Park, Keundeok, and Semiha Ergan. 2022. “Toward Intelligent Agents to Detect Work Pieces and Processes in Modular Construction: An Approach to Generate Synthetic Training Data.” In Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics, CRC 2022, 802–11. American Society of Civil Engineers (ASCE).

Park, Keundeok, Semiha Ergan, and Chen Feng. 2021. “Towards Intelligent Agents to Assist in Modular Construction: Evaluation of Datasets Generated in Virtual Environments for AI Training.” In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 38:327–33. IAARC Publications.

Pomerleau, François, Francis Colas, and Roland Siegwart. 2015. “A Review of Point Cloud Registration Algorithms for Mobile Robotics.” Foundations and Trends® in Robotics 4 (1): 1–104.

Puig, Xavier, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba.
2018. “Virtualhome: Simulating Household Activities via Programs.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 8494–8502.

Qi, Charles R., Xinlei Chen, Or Litany, and Leonidas J. Guibas. 2020. “Imvotenet: Boosting 3d Object Detection in Point Clouds with Image Votes.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4404–13.

Qi, Charles R., Or Litany, Kaiming He, and Leonidas J. Guibas. 2019. “Deep Hough Voting for 3d Object Detection in Point Clouds.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 9277–86.

Qi, Charles R., Wei Liu, Chenxia Wu, Hao Su, and Leonidas J. Guibas. 2018. “Frustum Pointnets for 3d Object Detection from Rgb-d Data.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 918–27.

Qi, Charles R, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. “Pointnet: Deep Learning on Point Sets for 3d Classification and Segmentation.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 652–60.
Qi, Charles Ruizhongtai, Li Yi, Hao Su, and Leonidas J Guibas. 2017. “Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.” Advances in Neural Information Processing Systems 30.
Qin, Zheng, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. 2022. “Geometric Transformer for Fast and Robust Point Cloud Registration.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11143–52.
Rabbani, Tahir, Sander Dijkman, Frank van den Heuvel, and George Vosselman. 2007. “An Integrated Approach for Modelling and Global Registration of Point Clouds.” ISPRS Journal of Photogrammetry and Remote Sensing 61 (6): 355–70.
Rausch, Christopher, and Carl Haas. 2021. “Automated Shape and Pose Updating of Building Information Model Elements from 3D Point Clouds.” Automation in Construction 124: 103561.
Roberts, Mike, Jason Ramapuram, Anurag Ranjan, Atulit Kumar, Miguel Angel Bautista, Nathan Paczan, Russ Webb, and Joshua M. Susskind. 2021. “Hypersim: A Photorealistic Synthetic Dataset for Holistic Indoor Scene Understanding.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 10912–22.
Romero-Jarén, R, and JJ Arranz. 2021. “Automatic Segmentation and Classification of BIM Elements from Point Clouds.” Automation in Construction 124: 103576.
Rusu, Radu Bogdan, Nico Blodow, and Michael Beetz. 2009. “Fast Point Feature Histograms (FPFH) for 3D Registration.” In 2009 IEEE International Conference on Robotics and Automation, 3212–17. IEEE.
Schnabel, R., R. Wahl, and R. Klein. 2007. “Efficient RANSAC for Point-Cloud Shape Detection.”
Computer Graphics Forum 26 (2): 214–26. https://doi.org/10.1111/j.1467-8659.2007.01016.x. Shao, Jie, Wuming Zhang, Aojie Shen, Nicolas Mellado, Shangshu Cai, Lei Luo, Nan Wang, Guangjian
Yan, and Guoqing Zhou. 2021. “Seed Point Set-Based Building Roof Extraction from Airborne LiDAR Point Clouds Using a Top-down Strategy.” Automation in Construction 126: 103660.
Shi, Shaoshuai, Xiaogang Wang, and Hongsheng Li. 2019. “Pointrcnn: 3d Object Proposal Generation and Detection from Point Cloud.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 770–79.
Siddiqui, Yawar, Justus Thies, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. 2021. “Retrievalfuse: Neural 3d Scene Reconstruction with a Database.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 12568–77.
Silberman, Nathan, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. “Indoor Segmentation and Support Inference from Rgbd Images.” In European Conference on Computer Vision, 746–60. Springer.
Song, Shuran, and Jianxiong Xiao. 2016. “Deep Sliding Shapes for Amodal 3d Object Detection in Rgb-d Images.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 808–16.
Su, Hang, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. 2015. “Multi-View Convolutional Neural Networks for 3d Shape Recognition.” In Proceedings of the IEEE International Conference on Computer Vision, 945–53.
Sun, Che, Yunde Jia, Yi Guo, and Yuwei Wu. 2022. “Global-Aware Registration of Less-Overlap RGB-D Scans.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6357–66.
Szot, Andrew, Alexander Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Singh Chaplot, and Oleksandr Maksymets. 2021. “Habitat 2.0: Training Home Assistants to Rearrange Their Habitat.” Advances in Neural Information Processing Systems 34.
Tang, Pingbo, Daniel Huber, Burcu Akinci, Robert Lipman, and Alan Lytle. 2010. “Automatic Reconstruction of As-Built Building Information Models from Laser-Scanned Point Clouds: A Review of Related Techniques.” Automation in Construction 19 (7): 829–43.
Teboul, Olivier, Loic Simon, Panagiotis Koutsourakis, and Nikos Paragios. 2010. “Segmentation of Building Facades Using Procedural Shape Priors.” In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3105–12. IEEE.
Thomson, Charles, and Jan Boehm. 2015. “Automatic Geometry Generation from Point Clouds for BIM.” Remote Sensing 7 (9): 11753–75.

Turner, Christopher J, John Oyekan, Lampros Stergioulas, and David Griffin. 2020. “Utilizing Industry 4.0 on the Construction Site: Challenges and Opportunities.” IEEE Transactions on Industrial Informatics 17 (2): 746–56.
Vieira, Miguel, and Kenji Shimada. 2005. “Surface Mesh Segmentation and Smooth Surface Extraction through Region Growing.” Computer Aided Geometric Design 22 (8): 771–92.
Vo, Anh-Vu, Linh Truong-Hong, Debra F. Laefer, and Michela Bertolotto. 2015. “Octree-Based Region Growing for Point Cloud Segmentation.” ISPRS Journal of Photogrammetry and Remote Sensing 104 (June): 88–100. https://doi.org/10.1016/j.isprsjprs.2015.01.011.
Volk, Rebekka, Julian Stengel, and Frank Schultmann. 2014. “Building Information Modeling (BIM) for Existing Buildings—Literature Review and Future Needs.” Automation in Construction 38: 109– 27.
Wang, Boyu, Qian Wang, Jack CP Cheng, Changhao Song, and Chao Yin. 2022. “Vision-Assisted BIM Reconstruction from 3D LiDAR Point Clouds for MEP Scenes.” Automation in Construction 133: 103997.
Wang, Haiyang, Shaoshuai Shi, Ze Yang, Rongyao Fang, Qi Qian, Hongsheng Li, Bernt Schiele, and Liwei Wang. 2022. “RBGNet: Ray-Based Grouping for 3D Object Detection.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1110–19.
Werbrouck, Jeroen, Pieter Pauwels, Mathias Bonduel, Jakob Beetz, and Willem Bekers. 2020. “Scan-to-Graph: Semantic Enrichment of Existing Building Geometry.” Automation in Construction 119: 103286.
Wu, Ke, Wenzhong Shi, and Wael Ahmed. 2020. “Structural Elements Detection and Reconstruction (SEDR): A Hybrid Approach for Modeling Complex Indoor Structures.” ISPRS International Journal of Geo-Information 9 (12): 760.
Wu, Yi, Yuxin Wu, Georgia Gkioxari, and Yuandong Tian. 2018. “Building Generalizable Agents with a Realistic and Rich 3d Environment.” ArXiv Preprint ArXiv:1801.02209.
Xiao, Jianxiong, Andrew Owens, and Antonio Torralba. 2013. “Sun3d: A Database of Big Spaces Reconstructed Using Sfm and Object Labels.” In Proceedings of the IEEE International Conference on Computer Vision, 1625–32.
Xie, Qian, Yu-Kun Lai, Jing Wu, Zhoutao Wang, Dening Lu, Mingqiang Wei, and Jun Wang. 2021. “Venet: Voting Enhancement Network for 3d Object Detection.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 3712–21.
Xiong, Xuehan, Antonio Adan, Burcu Akinci, and Daniel Huber. 2013. “Automatic Creation of Semantically Rich 3D Building Models from Laser Scanner Data.” Automation in Construction 31: 325–37.
Yang, Fan, Lin Li, Fei Su, Dalin Li, Haihong Zhu, Shen Ying, Xinkai Zuo, and Lei Tang. 2019. “Semantic Decomposition and Recognition of Indoor Spaces with Structural Constraints for 3D Indoor Modelling.” Automation in Construction 106: 102913.
Yi, Kwang Moo, Eduard Trulls, Vincent Lepetit, and Pascal Fua. 2016. “Lift: Learned Invariant Feature Transform.” In European Conference on Computer Vision, 467–83. Springer.
Yin, Chao, Boyu Wang, Vincent JL Gan, Mingzhu Wang, and Jack CP Cheng. 2021. “Automated Semantic Segmentation of Industrial Point Clouds Using ResPointNet++.” Automation in Construction 130: 103874.
Zabin, Asem, Vicente A. González, Yang Zou, and Robert Amor. 2022. “Applications of Machine Learning to BIM: A Systematic Literature Review.” Advanced Engineering Informatics 51: 101474.
Zeng, Andy, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 2017. “3dmatch: Learning Local Geometric Descriptors from Rgb-d Reconstructions.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1802–11.
Zhang, Yifan, Qingyong Hu, Guoquan Xu, Yanxin Ma, Jianwei Wan, and Yulan Guo. 2022. “Not All Points Are Equal: Learning Highly Efficient Point-Based Detectors for 3d Lidar Point Clouds.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18953– 62.
Zhao, Yi-Pu, and Patricio A. Vela. 2019. “Scan2BrIM: IFC Model Generation of Concrete Bridges from Point Clouds.” In Computing in Civil Engineering 2019, 455–63. Atlanta, Georgia: American Society of Civil Engineers. https://doi.org/10.1061/9780784482421.058.
Zheng, Jia, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. 2020. “Structured3d: A Large Photo-Realistic Dataset for Structured 3d Modeling.” In European Conference on Computer Vision, 519–35. Springer.