Cloud Computing and Data Science
http://ojs.wiserpub.com/index.php/CCDS/

Review

3D Point Cloud for Objects and Scenes Classification, Recognition, Segmentation, and Reconstruction: A Review

Omar Elharrouss[1*], Kawther Hassine[2], Ayman Zayyan[1], Zakariyae Chatri[3], Somaya Al-Maadeed[1], Khalid Abualsaud[1]

[1]Department of Computer Science and Engineering, Qatar University, Doha, Qatar
[2]Higher School of Communications of Tunis, University of Carthage, Tunis, Tunisia
[3]Department of Informatics, University of Sidi Mohamed Ben Ab-dellah, Fez, Morocco
E-mail: omar.elharouss@qu.edu.qa

Received: 21 March 2023; **Revised:** 5 June 2023; **Accepted:** 5 June 2023

Abstract: Three-dimensional (3D) point cloud analysis has become one of the attractive subjects in realistic imaging and machine visions due to its simplicity, flexibility and powerful capacity of visualization. Actually, the representation of scenes and buildings using 3D shapes and formats leveraged many applications among which automatic driving, scenes and objects reconstruction, etc. Nevertheless, working with this emerging type of data has been a challenging task for objects representation, scenes recognition, segmentation, and reconstruction. In this regard, a significant effort has recently been devoted to developing novel strategies, using different techniques such as deep learning models. To that end, we present in this paper a comprehensive review of existing tasks on 3D point cloud: a well-defined taxonomy of existing techniques is performed based on the nature of the adopted algorithms, application scenarios, and main objectives. Various tasks performed on 3D point could data are investigated, including objects and scenes detection, recognition, segmentation, and reconstruction. In addition, we introduce a list of used datasets, discuss respective evaluation metrics, and compare the performance of existing solutions to better inform the state-of-the-art and identify their limitations and strengths. Lastly, we elaborate on current challenges facing the subject of technology and future trends attracting considerable interest, which could be a starting point for upcoming research studies.

***Keywords*:** point cloud, 3D point cloud, 3D object recognition, 3D object segmentation, 3D object and scene reconstruction

1. Introduction

The development of acquisition methods and sensing technologies contributed considerably to capturing objects and monitoring scenes using Three-Dimensional (3D) representation. As a result, these tasks have gotten much easier and more cost-effective: the use of 3D images provides efficient information about the captured objects and scenes, which supports many applications such as object reconstruction, automatic driving, scene scanning, etc.

The most basic shape of a 3D representation is a Point cloud: it is formed by a group of single points drawn in a 3D area [1] to represent objects. To each point are associated several measurements, including coordinates along the X, Y, and Z axes as well as more specific data such as the color value, recorded in Red Green Blue (RGB) format [2], and the

DOI: https://doi.org/10.37256/ccds.4220232722

lightness measure, which determines the point's brightness.

The point cloud is captured by laser scanners or LiDAR sensors. These latter use different technologies when it comes to generating the 3D shape of objects or scenes: LiDAR sensors use various beams to capture the density of points cloud, while laser scanners use light speed to generate the 3D points cluster. The efficiency of each sensor lead depends on the number of dots being generated. This number is also a determining factor when it comes to the effectiveness of the methods working on the 3D point cloud. Figure 1 represents some examples of objects and scenes depicted in 3D point clouds.

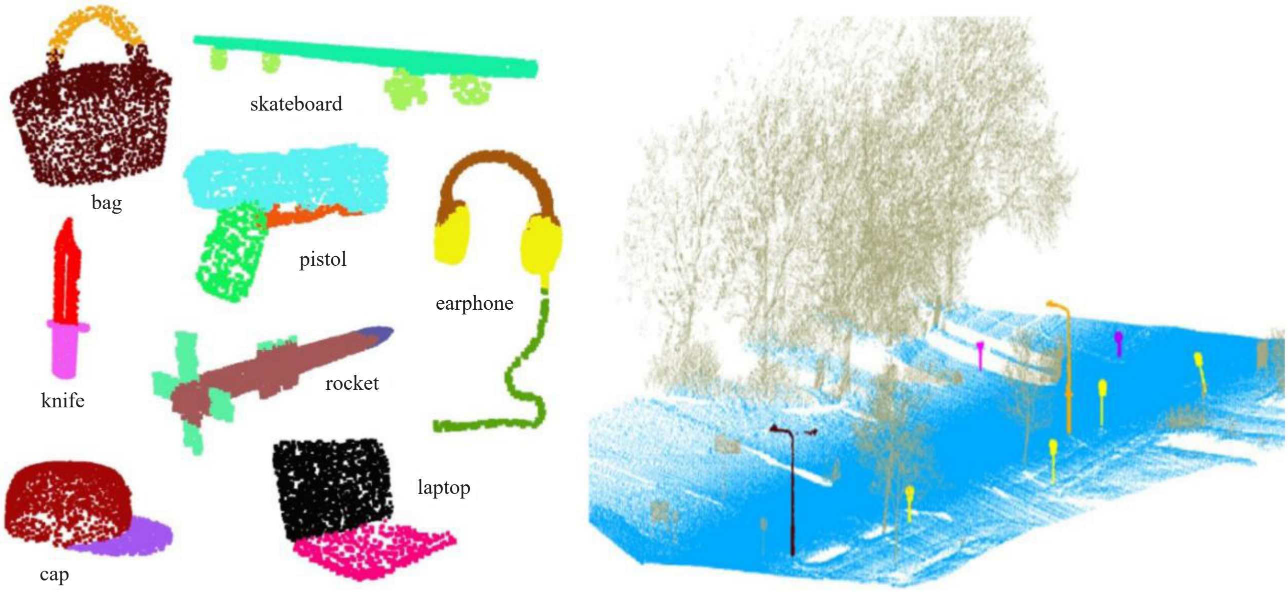


Figure 1. Object and scene with 3D point cloud representation

For the human eye, shapes and objects represented in 3D point clouds are easily detected and recognized even if the density of the gathered points is sparse. For example,

the swarm of dots that represents a tree can be recognized easily by the human eye, yet, a machine can consider this set of dots as noise which makes it hard to learn for the recognition algorithm. That being said, close attention was paid to finding effective methods to handle 3D point clouds growing up in a wide range of research fields [3-4]. Fortunately, with deep learning techniques and large-scale datasets, models became more capable of learning from various features and obtaining the desired outputs; whether it is about segmentation reconstruction or recognition of 3D shapes and scenes.

Unlike triangular meshes, the point cloud does not require storing or maintaining polygonal network connectivity [5] or topological consistency [6]. In addition, point cloud technology brought significant improvements in performance and reduced overhead costs. Therefore, this research area has drawn growing interest for a wide variety of applications [7]. Several sensing technologies have been used to create 3D models from 3D Point Clouds. This involves sensors like laser scanners, LiDAR, as well as Kinect, equipped with depth sensors to provide the distance and the depth of the object.

Even though the quick development of these sensing technologies makes the extraction of point clouds much easier [8], the analysis of point clouds is still one of the key aspirations [9]. In fact, point clouds usually suffer from noise pollution and the presence of outliers. This is usually due to limitations and noise inherent in sensing devices [10] and the nature of lightning, surface reflection, or artifacts in the respective areas [11]. As a way to redress this issue, filtering was implemented on key point clouds in order to achieve accurate results suitable for further processing. Yet, most of these approaches targeted meshes and only a few studies dealt directly with point clouds [12].

3D point cloud processing, on the other hand, attracted more attention and became a subject of interest and analysis

[13] worldwide. Different tasks were designed to process 3D point cloud data like, 3D object detection, recognition and segmentation as well as object and scene reconstruction [14].

In this regard, several studies have been performed, for instance, Dong et al. in [15] proposed a new 3D object recognition method based on sequential point cloud coding. This method first transforms a point cloud into an ordered multi-channel 2D array suitable for efficient 2D convolutional operation. While authors in [16] developed a localization method that integrates a supervised object recognition method to predict probabilistic distributions on classes of objects for individual sensor measurements, into the Bayesian network for localization.

At this stage, reviewing different tasks and techniques pertaining to the 3D point cloud should serve as a good starting point for researchers. Actually, it helps initiate basic concepts and understanding of leading ideas of the subject matter. Even though this paper is not an in-depth investigation of 3D point Clouds field of study, it does perform an overview of all existing tasks and approaches for 3D point cloud analysis. Within, we conduct an experiment to compare the performance of several methods on current datasets. Hence, the contributions of this work could be outlined as follows:

• A thorough taxonomy of the existing works is conducted with reference to various aspects, such as the methodology deployed to design the proposed models, types of image data that could be processed by the subject technology, and their application scenarios.

• Public datasets deployed to validate the proposed models on 3D point cloud data are discussed and compared with regard to different parameters.

• Several comparisons of the most significant works identified in the state-of-the-art have been conducted to demonstrate their performance across multiple datasets and metrics.

• Current challenges and issues that remained unresolved have been targeted. Insights about future directions and trends have been discussed along with their potential impact on research and development in the near and distant future.

The remainder of this article is organized as follows: first, we provide an overview of the filtering techniques applied to 3D point clouds. Then we describe the experiments performed on this kind of data before we detailed and discussed the respective results. Finally, we conclude the paper, and we identify potential challenges and future perspectives.

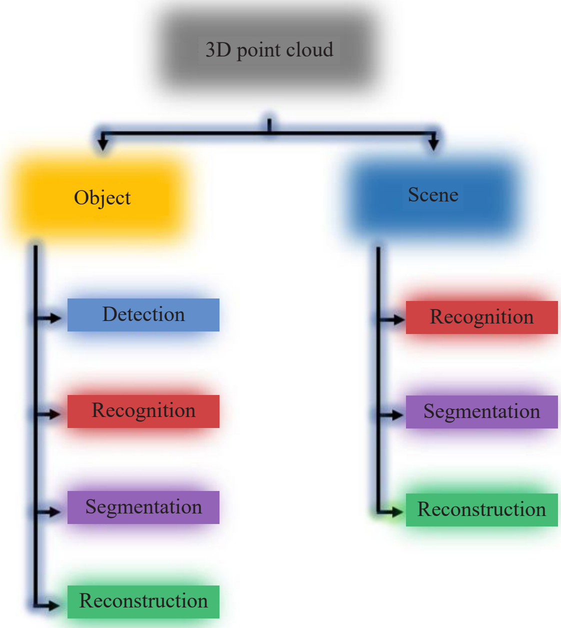


Figure 2. Computer vision tasks on 3D point cloud data

2. Related works

Table 1. Object detection recognition and segmentation methods and their characteristics

Task	Method	Technique	Dataset
	[11]	Statistical algorithm	Self-collected
	[19]	STEM, CaRFs, VGG	Washington RGB-D, Cornell grasping
	[20]	PPF, LRF	UWA, Queen's, bin picking
	[21]	Appearance-Based Object Recognition (ABOR)	KITTI
	[22]	Cross-modal, SHOT and ESF descriptors	self-collected
	[23]	Covariance descriptor	RGB-D
	[24]	LonchaNet, CNN, GoogleNet	ModelNet-10
	[25]	Global Orthographic Object Descriptor	Washington RGB-D
	[26]	iCLAP, tactile features	-
	[27]	3D FCN architecture	KITTI
	[28]	RNN,MLP	ModelNet40
	[29]	3D CNN	ModelNet40, MNIST
	[30]	3DSmoothNet, SDV, LFR	3DMatch, ETH
	[31]	voxels-based shape descriptor	self-collated MLS point cloud
Object detection and recognition	[32]	3D cluster descriptor	ImageNet
	[33]	PointNet, NetVLAD VGG/AlexNet	Oxford RobotCar
	[34]	ShapeContextNet (SCN)	ModelNet40, MNIST
	[35]	PVNet, Embedding Attention Fusion	ModelNet40
	[36]	VoxelNet, RPN	KITTI
	[37]	SO-Net	ModelNet10, ModelNet40, MNIST
	[38]	Saliency map, Point cloud conversion, Hough voting	RGB-D converted to 3d point cloud
	[39]	3DCNN, MLP	S3DIS
	[40]	Adopted 3D ZSL	ModelNet-10-40, Mcgill, SHREC2015
	[41]	TZSL, GZSL, 3D CNN, ResNet-101	ModelNet-10-40, McGill, SHREC2015
	[42]	3DCNN, ResNet	ModelNet10, ModelNet40
	[43]	GAN	KITTI
	[44]	RNN, K-D tree searching method	Bin-picking, UWA
	[45]	RS-CNN, MLP	ModelNet40
	[46]	MeteorNet	MSRAction3D

Table 1. (cont.)

Task	Method	Technique	Dataset
Object detection and recognition	[47]	2D CNN, JointNet	HDRObject9, SMDObject6
	[48]	InterpCNNs, MLP	ModelNet40
	[49]	Attentional PointNet	KITTI
	[50]	SFCNN	ModelNet40
	[51]	CNN, YOLO	KITTI
	[52]	PointNet++, PGM, PointsPool	KITTI
	[53]	DGCNN	ModelNet40
	[54]	MVF, CNN	KITTI
	[55]	Point pair feature	UWA, Real scene
	[56]	Octree, DeepCNN, VoxRec	Shapenetcore Part
	[57]	two-stage Weakly Supervised	KITTI
	[58]	VoxNet, 3DCNN	RGB-D, LiDAR
	[59]	3D to 2D projection, Inception-v3	KITTI
	[60]	MVG, RPN	KITTI
	[61]	RPN, segmentation network, estimation network	KITTI
	[21]	CNN, Contextual Labeling Refinement (CLR)	KITTI
	[62]	Voxel-based 3D-CNN	LiDAR point cloud Ottawa
Object segmentation	[28]	3D CNN,MLP	ShapeNet part
	[29]	3D CNN	ModelNet40
	[32]	2D CNN labeling	ImageNet
	[34]	Attentional ShapeContextNet (A-SCN)	ShapeNet part
	[37]	SO-Net	ShapeNet part
	[45]	RS-CNN, MLP	ShapeNet part
	[46]	MeteorNet	KITTI, Synthia
	[48]	InterpCNNs, MLPs	ShapeNet Part
	[50]	SFCNN	ShapeNet Part
	[53]	DGCNN	ShapeNet Part
	[56]	Octree, DeepCNN, VoxRec	Shapenetcore Part

The impact of point clouds on many computer vision applications pushed researchers toward finding the best techniques to classify, recognize or segment the objects pictured using point clouds like presented in Figure 3. In this context, many approaches have been proposed, and from which we described and analyzed the ones we presumed more relevant. These approaches are split into categories in accordance with the tasks performed on point clouds: this could be object classification and recognition, object scene segmentation, and object scene reconstruction as presented in Figure 2. A summarization of the main characteristics of each method and for each task is detailed in Tables 1 and 2.

Table 2. Summarization of scenes recognition and objects reconstruction methods

Task	Method	Technique	Dataset
Scenes and places recognition	[69]	PointNetVLAD, LPD-Net, RN-VLAD, GCA	Oxford RobotCar
	[33]	PointNet, NetVLAD VGG/AlexNet	Oxford RobotCar
	[61]	RPN, segmentation network, estimation network	KITTI
Scenes and obejcts reconstruction	[78]	Building Information Modeling (BIM)	2D-3D-S, S3DIS
	[79]	unsupervised Building Information Modeling (BIM)	2D-3D-S, S3DIS
	[80]	Vectorized model, Multistep 2D optimization	MC and Office
	[81]	CNN, building element segmentation and reconstruction	S3DIS, 2D-3D-S
	[83]	Mesh deformation network and Graph Unpooling Network	ShapeNet
	[84]	point set generation network (PSGNet)	ShapeNet
	[85]	Point Cloud Reconstruction Network, Silhouette Completion Net-work	ShapNet
	[86]	CNN	ShapNet
	[87]	Encoder-decoder	ShapNetCore, Pix3D
	[88]	RfD-Net, 3D detector, spatial transformer, shape generation	ScanNet [89], Scan2CAD [77]

2.1 *Objects detection and recognition*

Recently, computer vision and pattern recognition communities have been focalizing object detection. It has always been considered an essential task in computer and robot vision systems. Despite the progress achieved in this area, a lot of improvements would have to be made to achieve human-level performance. In 3D representation of data, the detection and recognition of objects represent a different, more complicated task compared to the work performed on 2D data [17-18].

Recognizing 3D objects could be more problematic if the 3D point cloud data includes missing parts and or noisy data. To reach accurate recognition, researchers were compelled to find methods to recover or estimate these parts of obscured data [19-20]. However, the particularity of 3D point cloud representation, compared to RGB images and the usual types of data, makes it hard to operate without proper data preparation [21]. The 3D point cloud data should undergo a series of pre-processing steps such as data denoising and cleansing before initiating the recognition process. This does not change the fact that working with 3D point clouds, instead of conventional data, has driven computer vision applications towards new strategies of data analysis and optimized consumption of memory and storage [22]. For that, and using hierarchical cascading for low-quality, authors in [19] introduced an object recognition and detection solution from RGB-D images. This method consists of analyzing comprehension probabilities and object class probabilities by extracting features using VGG architecture. Also, to recognize objects using a unified representation based on point clouds, authors in [22] proposed a cross-modal Visio-Tactile object recognition method using tactile data retrieved from a tactile skin reader. SHOT and ESF descriptors were used for the recognition process.

Basically, capturing a 3D point cloud requires accurate sensors like laser, LiDAR, and range finder to generate precise measurements. Yet, in many cases, low-quality sensors are the ones being used. Here is where things get complicated and the need to achieve 3D object recognition despite low-quality representation arises. To handle the imperfect acquisition, authors in [23] suggested the use of descriptors (usually exploited in other computer vision domains) to process 3D point clouds. This method establishes an object recognition process based on the Covariance descriptor, which proved to be effective in many similar subject studies. In [24] authors designed a CNN-based architecture named LonchaNet for real-time recognition represented in 3D point cloud. The proposed model exploited the GoogleNet network for features extraction and was tested on the ModelNet-10 dataset.

In the same low-quality context, authors in [25] proposed a new object descriptor named Global Orthographic Object Descriptor (GOOD) where robots recognize objects on a real-time basis. This approach aimed at real-time recognition, easy descriptive use, and low computational cost. The evaluation was performed on the RGB-D Object dataset. On the other hand, authors in [26] built up an algorithm named iterative Closest Labeled Point (iCLAP) based on kinesthetic and tactile information. This solution combined two sensing modalities to create a composite perception of objects and yielded a significant improvement in recognition rate despite low-quality conditions.

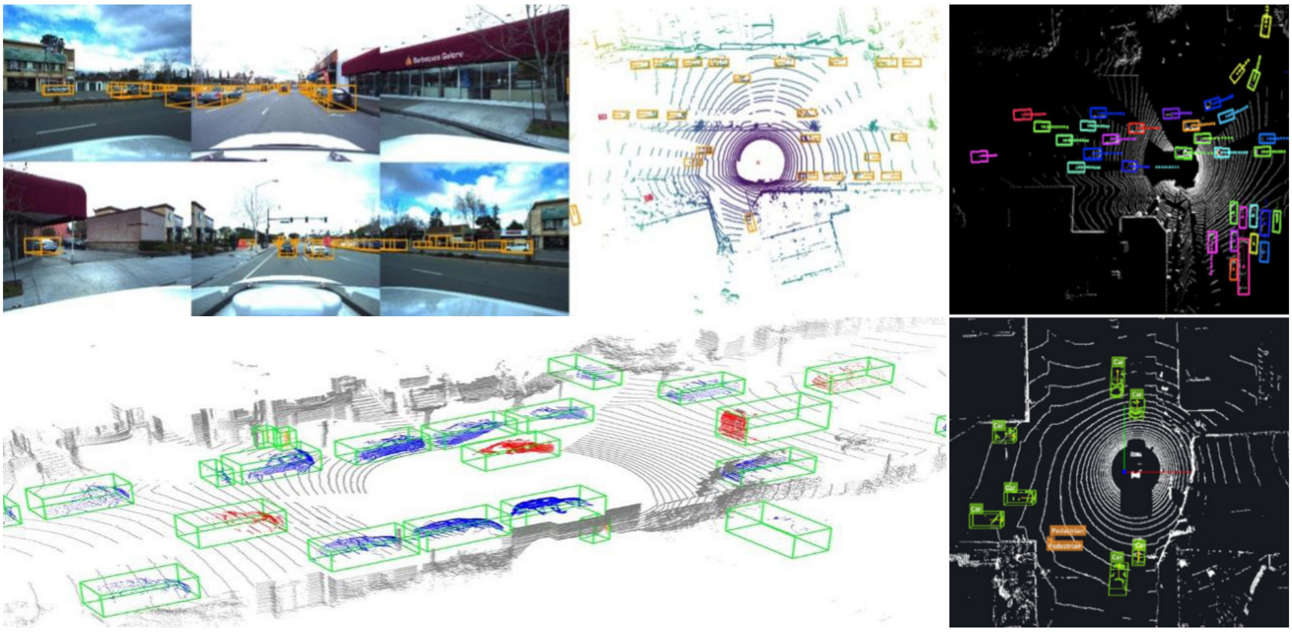


Figure 3. Car detection and recognition from 3d point cloud data

On KITTI dataset, authors in [27] adopted a 3D vehicle detection using 3D boundary boxes by extending 2D FCN architecture to 3D FCN. Another low-quality case is camera self-occlusions caused by hand obstruction.

For classification purposes, authors in [28] designed an RNN architecture named Point-Net using Multi-layer perception (MLP). Their purpose was to overcome small perturbations of input points as well as errors through point insertion or deletion. To improve the performance of PointNet architecture, authors in [29] used metric space distances to enable learning from local features as well as increasing contextual scales by learning from multiple scales. Also, the learning can be done from different modalities like LIDAR [30]. Authors in [31] used mobile laser scanning point cloud data to design 3D multi-scale shape descriptor for object recognition named SigVox. This method used a shape description exploiting voxels to identify different types of lamp poles and traffic signs in streets. In [32] authors' main objective was to create classification solutions for 3D object recognition from point cloud data. Also, these types of methods are implemented for place and scene recognition. They implemented 2D Convolutional Neural Network (CNN) for labeling, 3D clustering descriptor for object segmentation, and a deep learning-based model for recognition. Also, this is used for scene and place recognition like in [33]. Another 3D object recognition method was proposed in [34] which involves a deep learning-based technique named Attention ShapeContextNet (SCN). The architecture concatenates a list of ShapeContextNet blocks followed by Multilayer Perceptron (MLP) layers for the final classification. The proposed method has been evaluated on ModelNet40 classification dataset. With the same leading idea, authors in [35] used a deep-learning-based method for 3D shape recognition over point cloud data and multi-view 3D shape representation of the object. This combination was executed through a unified network while an Embedding Attention Fusion (EAF) block was used over the features on point cloud and multi-view 3D shape; features extraction was carried out using AlexNet backbone.

Detecting 3D objects from LiDAR point cloud data was the focus in [36]. In this work, authors introduced a generic 3D detection network named VoxelNet which consists of extracting features before predicting boundary boxes. The region proposal network (RPN) was adapted to interface with the 3D point cloud while generating boundary boxes and probability maps. Authors in [37] proposed a 3D point cloud classification method named SO-Net. SO-Net architecture is a multi-task network that extracts global features to be used in the classification network. ModelNet10, ModelNet40, and MNIST datasets were used for assessment purposes. In [30] proposed compact learned local feature descriptors for 3D point cloud matching. They introduced the smoothed density value (SDV) voxelization as a novel input data representation in order to reduce the sparsity of the input voxel grid, mitigate the boundary effects and smooth out miss-alignments due to local reference frame (LRF) estimation errors.

In order to recognize objects from RGB-D images, a segmentation process was per-formed on RGB and in-depth images [38]. This step was followed by a conversion and a generation to 3D point cloud representation. On resulting 3D point cloud data, features extraction followed by Hough voting operation are performed to recognize existing objects. Using 3D laser-scanned point clouds, authors in [39] established a 3DCNN-based deep learning method (MLP used in the last layer) to recognize building elements and perform object reconstruction. To evaluate the model, the authors used the S3DIS dataset. Despite the wide scope of solutions for object recognition, a concern about identifying unseen classes of 3D point cloud data leads to using some of the techniques initially applied to 2D data: one of which is Zero-Shot Learning (ZSL). For instance, authors in [40] adopted a 2D ZSL method to 3D point cloud data by changing the setting parameters of the original ZSL approach. In that same regard, authors in [41] combined Transductive Zero-Shot Learning (TZSL) and Generalized Zero-Shot Learning (GZSL) techniques for 3D point cloud classification. In [42] authors used the ResNet backbone for features extraction on multi-view of 3D presentation of an object, afterward, they implemented a deep-learning-based approach named View-based Weight Network (VWN) for 3D object recognition. Projections of different views of a given object were used for its identification. Exploiting Generative Adversarial Network (GAN) was also one of the ways to handle 3D point cloud object recognition: authors in [43] proposed a method called Point Cloud Up-sampling Adversarial Network, PU-GAN which consists in identifying an object from a whole scene captured using LiDAR sensors. Considering that the scene may comprise other objects like buildings, trees, etc. PU-GAN is designed to fill the detected object by up-sampling the point sets. In [44], instead of transforming point clouds to voxel grids, authors introduced a technique of extracting 3D key point features directly from point clouds. These features are used to recognize this object by exploiting and preserving its geometric information. In a different work, [45], a Relation-Shape Convolutional Neural Network (RS-CNN)-based solution was devised for object classification, segmentation, and reconstruction. Besides, to classify, segment and estimate the scene flow from 3D point clouds, authors in [46] introduced a method named MeteorNet. Unlike previous works which used normal 3D point cloud vectors [47], MeteorNet's purpose was to classify objects from 3D point cloud sequences using KITTI dataset for performance assessment. Authors in [48] proposed a CNN-based deep learning model named Interpolated Convolution Network (InterpConv). InterpConv was implemented for object recognition and segmentation purposes. While authors in [49] relied on PointNet architecture to detect and recognize 3D objects.

Even though most of the existing methods operated directly on 3D point cloud representation for object recognition, other approaches chose to convert, first, 3D data into 2D data. That is, to benefit from the existing object recognition methods with deep learning techniques. For instance, some authors developed a solution for object recognition using 2D representation while converting 3D shape of point clouds to 2D multi-view representation cloud data. Then, work on the new data with a 2D CNN-based method to recognize objects. A fusion module is implemented to collect different multi-view features for classification parts. The proposed method was tested using three datasets namely HDRObject9, SMDObject6, and SUObjects. Also, using an adaptive spheri cal projection of 3D point cloud with a fractal structure, authors in [50] implemented a convolutional neural network for 3D object recognition. Using 2D representation, the spherical projection enables the deep model to learn from various features.

From a sensing perspective, it is commonly known that methods based on LiDAR information are good to detect obstacles, yet limited for other types of objects. For better outcomes, combining the vision and point cloud (LiDAR) information is used in [51] by exploiting a fusion method for vehicle detection. In this solution, obstacles are detected using point cloud data, then mapped to the image. The YOLO technique is used for the detection and localization of vehicles. While the boundary boxes are mapped to the same image. In [52], authors defined a two-stage 3D object detection framework, named sparse-to-dense 3D Object Detector (STD). The proposed network goes through two stages: PointNet++, as backbone, detects the regions where the objects are, then PointsPool uses the first stage's features to detect objects with a compact representation. Also, to Work with convolutional neural networks on point cloud data, an adaptation is made by authors in [53] using a spatial transformation for point cloud segmentation and classification. Authors proposed a deep learning module named EdgeConv which was replicated between the model's layers. In another research work namely, [54], authors designed a multi-view Fusion (MVF) method by introducing a dynamic voxelization algorithm with the purpose of overcoming the problems affecting existing Voxel-based methods. The proposed voxelization method consists of two branches in the MVF network including the Bird-eye view and perspective view used for features extraction. From each branch, a list of convolution and pooling layers is fused to obtain the final classification. Using a new descriptor based on point pair features of 3D point cloud objects, authors in [55] estimated and recognized the poses of an object. Authors in [56] used many features such as surface normal and surface curvature and 3D point cloud for classification and segmentation processes. The authors proposed a multitask-deep-learning-based method by extracting the histogram and voxelized features of an object to be used as input for the proposed CNN network. Authors in [57] operated on LiDAR 3D point cloud data and implemented a two-stage weakly supervised system for 3D object detection. This model is trained on a small dataset with few annotated scenes (500 annotated scenes). The first stage operates on cylindrical object generation learning while the second stage implements cuboids confidence score selection. On the same dataset, authors in [58] developed a 3DCNN model referred to as VoxNet, for 3D object recognition. Also, by converting 3D point cloud data to a 2D representation, authors in [59] used the projection technique for mapping 3D data into 2D images. Then Inception-v3 was used for data classification. Object recognition is performed using the fusion of image-based detection and point cloud data. While Authors in [60] implemented Multiple Views Generator (MVG) methods that observe the scene from four views, and a region proposal network (RPN) for object detection.

The unstructured distribution of LiDAR point clouds has also been an issue when it comes to pattern learning, especially for recognition purposes [61-63]. Henceforth, authors in [64] proposed a variant of Broad Learning System (BLS) that implements a Unified Space Auto-Encoder (USAE). Referred to as USAE-BLS, the result came out as a lightweight model for 3D object recognition. In [65] authors sought object classification from 3d point cloud representation. While combining point clouds and voxelized data, they used the Euclidean distance between the point and the object's center to extract connections between points within each voxel. Next, a deep learning network performed the task of object classification. A different deep learning-based method named Drop Channel Graph Neural Network (DC-GNN) was designed in [66] to extract meaningful features and depict an effective representation of objects before classification. This solution relied on k-NN-based drop channel and Multi-Layer Perceptron (MLP) Networks. In [67], another MLP-based method implemented spatial-shift MLP as the backbone to determine the relationship between patches from different views. A different approach, known as view, set pooling transformer (OVPT), [68], was designed to build a comprehensive image of the same object from different perspectives using view information entropy computation. This process can be found also in other tasks like scene and place recognition [69]. In [70] authors proposed a 3D local feature descriptor named Histograms of Point Pair Features (HoPPF) for 3D object recognition. The HoPPF method implies robust representation and efficient computation.

In [71] authors proposed a dynamic ball query (DBQ) network for subset selection of input points and then assigned the feature transform to each selected point. Yet, in many cases, objects are partially covered and the information gathered is barely sufficient for accurate detection. Accordingly, authors in [72-73] presented a LiDAR-based Behind the Curtain Detector (BtcDet) model to complete the shape of objects before detection. By learning the shape priors, this solution estimates the complete shape of partially obscured (curtained) objects. Authors in [74] introduced a Boundary-Aware 3D Object Detection BADet network. In the form of a local neighborhood graph, this method efficiently displays local boundary correlations of an object. For more informative RoI-wise representations, a lightweight Region Feature Aggregation Module was devised to exploit voxel-wise, pixel-wise, and point-wise features with expanding receptive fields. In [75-76] authors described a solution for 3D object detection and recognition based on two different inputs of the same object: the RGB image and the 3D point cloud. On the other hand, authors in [75] implemented a network for 3D object detection using encoder-decoder architecture, a refinement module, and a Cascade Bi-directional Fusion.

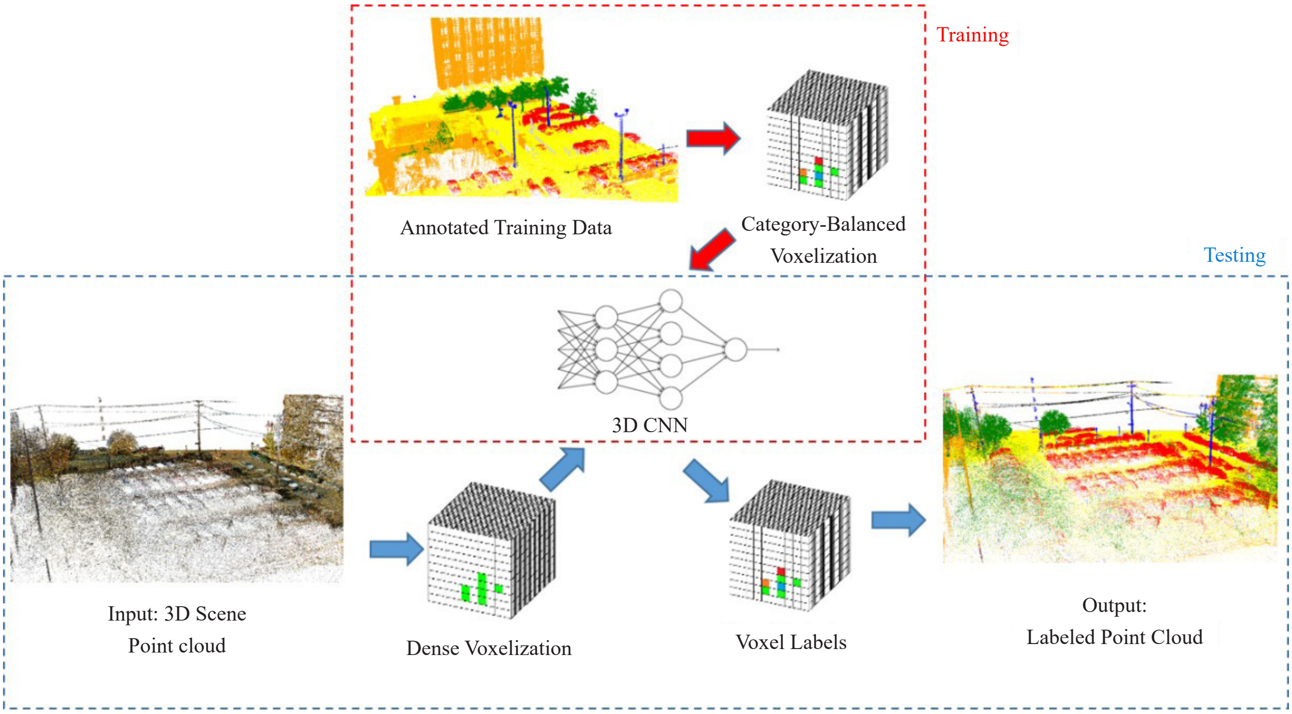


Figure 4. 3D point cloud segmentation proposed in [62]

2.2 *Scenes and place recognition*

Recognizing the observed scenes could be very helpful for navigation and localization purposes. In the case of conventional images and videos, the recognition of places and scenes could be easy but with 3D data presentation it becomes quite difficult: recognizing scenes and places from RBG images is easy using deep-learning-based techniques. But the representation of these scenes and places using 3D points makes it harder due to the potential presence of a swarm of dots as well as objects within the scene. To avoid possible confusion, many studies were conducted to differentiate places and scenes from other types of data.

Based on large-scale places datasets, authors in [33] introduced a 3D point cloud scene recognition approach named PointNetVLAD. The model encompasses three blocks, namely PointNet architecture, NetVLAD, and a fully connected network. NetVLAD architecture is used to aggregate image features extracted using VGG/AlexNet into a global descriptor vector. While in [61], authors depicted a 3-step Deep Neural Network (DNN) using RGB image and LiDAR data. RPN model is put in place to generate feature maps from RGB images. These features are lifted into 3D proposals to be used for segmenting object points. The estimation network is used to extract 3D objects from bounding boxes. In [69], authors propose the Point Cloud and Image Collaboration Network (PIC-Net), which combines the features of an image with those of a point cloud and mines the complementary information they provide. Additionally, the night image is converted into a daytime image in order to improve recognition performance at night.

2.3 *3D objects, scenes and buildings reconstruction*

The reconstruction of objects and scenes using different types of data including RGB-D images, Lidar, and Laser scanner data is a complex yet beneficial task for many applications such as medical imaging, virtual reality, computer imaging, etc. The reconstruction of objects, scenes, and buildings after being transferred makes the exchanges of this type of data very helpful compared to other types of data like images or videos.

Whether it is 3D laser scanning, RGB-D camera, stereo camera, or LiDAR, 3D point clouds usually represent 3D surfaces or objects in accurate manners. Using 3D point clouds, a building or a surface could be pictured and reconstructed [77]. Accordingly, multiple studies have been conducted on 3D Objects, scenes, and buildings reconstruction. while objects were reconstructed by filling the spaces between points. With the same scope, authors in [78] introduced a method for building Walls reconstruction using Building Information Modeling (BIM). Using the same method type, authors in [79] devised an unsupervised Building Information Modeling (BIM). This method also referred to as scan-to-BIM, consists of reconstructing the building including walls for scanned 3D point cloud data.

However, generating data using image-based 3D reconstruction or using LiDAR is prone to errors, missing parts or noisy points. Using Multi-step 2D optimization and Levels of Detail (LoDs) which perform a progressive vectorized reconstruction of models, authors in [80] reconstructed surfaces from 3D point cloud data. To recognize the scene components and reconstruct them from 3D point cloud data, authors in [81] proposed a deep learning method to detect and recognize building elements from laser-scanned data. Then the detected elements are segmented before being reconstructed. Another paper [82], investigated an alternative approach, which is based on the alignment of convex hulls of segments detected in a depth image with convex hulls of target 3D object models. The input of the algorithm is a triangular mesh obtained from a 3D point cloud and the output is a set C of convex surfaces Ci. In the same context, yet to create 3D point objects from a single image, authors in [83] developed a CNN-based model named Pixel2Mesh with a Mesh deformation network and a Graph Unpooling Network. To generate 3D point cloud from RGB images, authors in [84] implemented a deep learning method named point set generation network (PSGNet). Also, authors in [85] proposed a method for object reconstruction using RGB images and a silhouette map. An encoder-decoder network was dedicated to the extraction of the complete object silhouette. The completed silhouette and RGB images are introduced as input of the ResNet-50 network. A reconstruction network was set up to build the point cloud representation of the object. The training and the validation were completed on the ShapeNet dataset using Chamfer Distance (CD) and Earth mover's distance (EMD) metrics. Another method to reconstruct a 3D object from a single image is proposed in [86]. The method named pixel2point exploited an encoder-decoder network to generate a point cloud. In [87] an encoder-decoder network was deployed to produce RGB information in latent space and then use it to reconstruct 3D objects.

The reconstruction of different components of a scene including objects, walls and buildings has become a subject of interest for many applications including interior design and robot navigation. For this reason, many researchers strived to improve reconstruction mechanisms over 3D point clouds. For example, using point cloud, authors in [88] presented a deep learning solution named RfD-Net for the detection and reconstruction of object surfaces in a scene on the ScanNet [89] dataset.

2.4 *Objects and scenes segmentation*

Segmenting objects in a scene is convoluted when the data is captured using LIDAR sensors. This data is presented with a swarm of dots and makes it hard to find a pattern to segment these objects. To segment the point cloud into regions of ground: low foreground (i.e. short objects), and high foreground (tall objects), authors in [21] built an object segmentation solution using CNN and Contextual Labeling Refinement (CLR) to complete the connected components of an object. As proposed for classification purposes, authors in [28] also presented a 3DCNN architecture for object part segmentation. The architecture is simple with 19 convolutional layers and assessed using the ShapeeNet part dataset.

A new upgraded version of PointNet architecture was described in [29] and labeled PointNet++. This new version consists of Multi-scale group (MSG) and Multi-resolution group (MRG) layers and provides a comprehensive solution for part segmentation. Also, for recognition purposes, the solution in [32] was designed to segment objects for easier recognition using 3D spatial information and a CNN-based model. Meanwhile, in [34], authors adopted ShapeContextNet (SCN) for point cloud recognition and attention ShapeContextNet (A-SCN) for segmentation. The evaluation of A-SCN was performed on the ShapeNet part segmentation dataset.

In [37], a 3D point cloud classification and segmentation system, SO-Net, was developed upon a multi-task network that extracts spatial distribution features before segmenting objects by a segmentation network. For evaluation purposes, authors used the ShapeNet part dataset as well. Another method was devised in [45] using a deep learning architecture named Relation-Shape Convolutional Neural Network (RS-CNN). This solution was tailored to segmentation as well as classification and object reconstruction. In [46] authors presented a deep learning method named MeteorNet. Here, semantic segmentation was performed using Synthia and KITTI datasets. In [48] another deep learning model named Interpolated Convolution operation (InterpConv) was deployed to tackle point cloud feature learning and understand issues of object segmentation.

As mentioned earlier, the conversion of 3D point clouds to a 2D representation can make the processing of a deep learning model much more effective. For this reason, an adaptive spherical projection using a 2D CNN model, also referred to as Spherical Fractal Convolutional Neural Networks (SFCNN), is used before segmenting objects [50]. Also, to Work with convolutional neural networks on point cloud data, an adaptation is made by authors in [53] using a spatial transformation for point cloud segmentation and classification. Authors proposed a deep learning module named EdgeConv which was replicated between the model's layers. Unlike standard methods implementing deep learning and using 3D point cloud for object recognition, authors in [56] utilized different features such as surface normal and surface curvature to classify and segment 3D objects. The authors proposed a deep-learning-based method using OCTREE encoder features for segmenting different parts of the object. With the aim of labeling a scene that involves objects pictured with 3D point clouds, authors in [62] proposed an object segmentation mechanism centered on a 3D convolutional neural network. The proposed method is presented in Figure 4. This approach minimizes prior knowledge of the labeling problem and does not require a segmentation step or artisanal features.

Segmenting objects in a scene some authors proposed a 3D object segmentation method using the Dual Attention Network (DANet) to overcome the irregularity and sparsity of point clouds for 3D object classification and segmentation [90-91]. DANet combines two modules: a Local Feature Extraction module (LFE) for local feature learning and a Global Feature Fusion module (GFF) for global information fusion. In [92] authors adopted a Graph Convolutional Neural Network (GCNN) to learn from contextual-point relationships between neighboring points for a structured 3D point cloud segmentation. Also using a GCNN-based network, authors in [93] attempted to locate the attentional regions using spatial information with different distances.

A Deep Feature Transformation Network (DFT-Net) was proposed in [94]. It was intended to capture local features and exploit the relationship between neighborhood points. In the same context, and to extract spatial and semantic consistencies between point cloud data, authors in [95] proposed a deep learning-based network that encompasses a data augmentation module and a segmentation network.

3. Datasets

The growth of ML and DL algorithms led to the creation of large-scale datasets [96]. These datasets have had a big impact on models outcomes and effectiveness. A dataset with thousands or billions of annotated data offers a huge amount of information model can learn from. For example, ImageNet [97] helps to evaluate visual recognition algorithms, while VGGFace [98] is designed to validate face recognition methods. For 3D point clouds, KITTI, ModelNet-10, and ModelNet-40 are the most famous data sets for object classification and recognition. For 3D object segmentation, a more recent research area, few datasets include annotation. Figure 5 portrays samples from the most famous datasets used for 3D object recognition, segmentation, and reconstruction. In the following paragraph, we will briefly discuss the individual properties of these datasets.

KITTI (http://www.cvlibs.net/datasets/kitti/)The images pertaining to this dataset are captured from various places in the metropolis of Karlsruhe (Germany), including highways and rural regions. It contains more than 12K images illustrating varied objects. In each image, we can find around 30 pedestrians and 15 vehicles.

MSRAction3D (https://www.uow.edu.au/ wanqing/#Datasets) is an action recognition dataset or 3D point representation that contains in-depth images grouped in 567 map sequences with a resolution of 640 × 240 pixels.

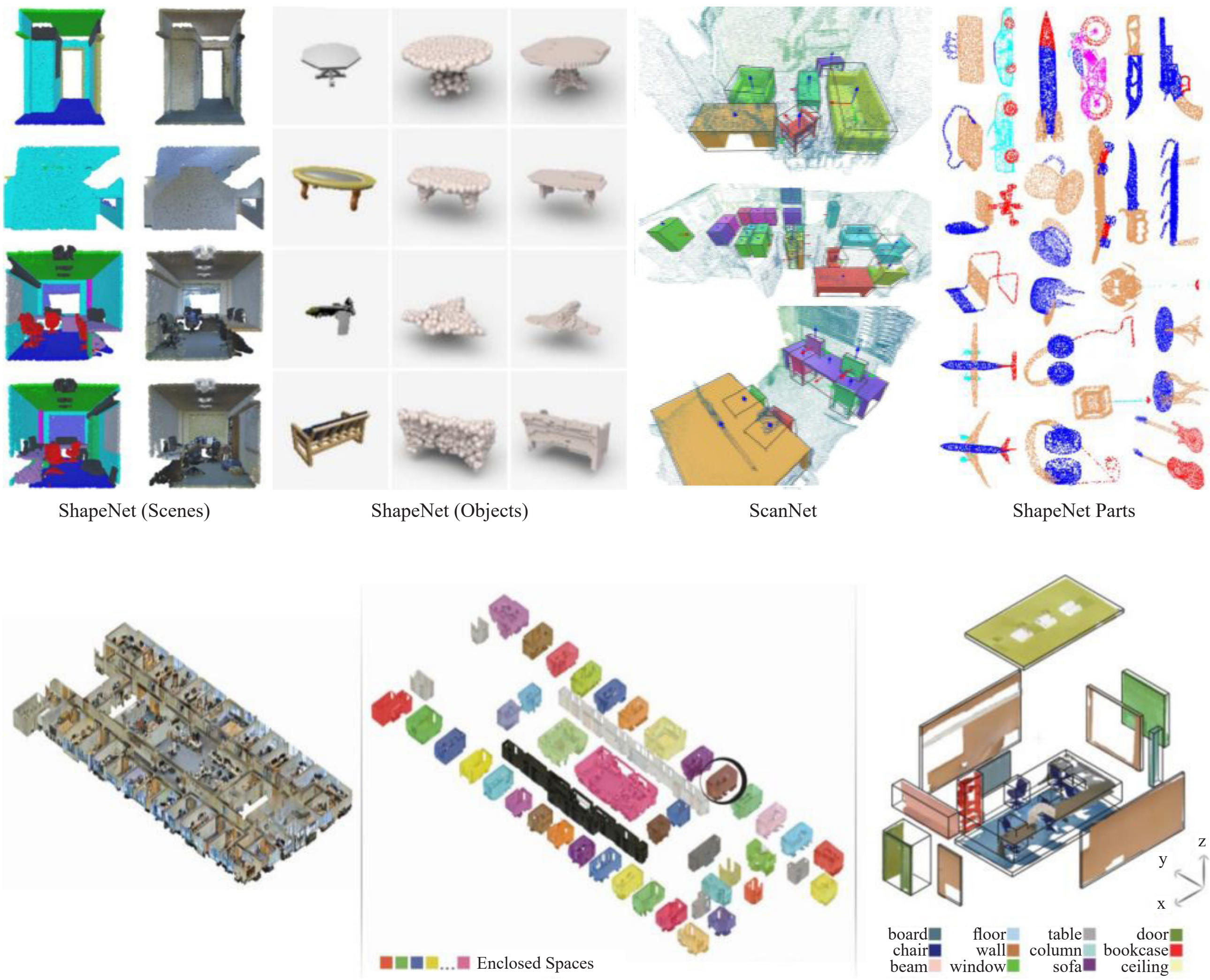
Images are captured using a Kinect sensor.

ModelNet (https://modelnet.cs.princeton.edu) ModelNet dataset derives a set of versions including ModelNet10 and ModelNet40. ModelNet10 is composed of 10 classes of different 3D point cloud objects. While ModelNet40 contains 40 classes. These two datasets are used by multiple deep-learning methods for 3D object classification and recognition.

3D ShapeNets (https://shapenet.org) or ShapeNet Parts is a 3D object segmentation dataset composed of more than 31K meshes of 16 classes of objects. Object shapes comprise 2 to 5 labeled parts.

ImageNet (http://image-net.org/) This data set is dedicated to Large Scale Visual Recognition Challenge (ILSVRC). ImageNet is also a large-scale dataset with thousands of images and subnets. Each subnet is represented by 1,000 images. The current version of the dataset contains more than 14,197,122 images wherein 1,034,908 images of the human body are annotated with bounding boxes.

2D-3D-S (https://github.com/alexsax/2D-3D-Semantics) is a dataset including 2D and 3D data with segmented geometric and semantic annotations. This dataset is collected from 6 large-scale areas of more than 6,000 m^2 of buildings with annotated 3D point cloud representations. It does, also, contains 70K RGB depth images of the same scenes.



2D-3D-S

Figure 5. Sample examples from 3D point cloud datasets

4. Experimental results and discussion

In this section, we present the experimental results of the state-of-the-art works eval uated on standard 3D point cloud benchmarks, including KITTI, ShapeNet, ShapeNet parts, ModelNet10, ModelNet40, and MNIST. The accuracy and efficiency are compared between methods performing the same task. Methods performance is often measured on the validation set and test set for each dataset: if so, this distinction is taken into consideration in our comparison.

4.1 *Object detection and recognition evaluation on KITTI dataset*

Table 3. Evaluation results of car detection on KITTI dataset

Dataset	Test/Val	Method	3D Car			BEV Car		
			Easy	Moderate	Hard	Easy	Moderate	Hard
KITTI	val	Yang et al. [60]	86.31	77.32	73.21	89.36	81.52	79.62
		Li et al. [27]	84.20	75.3	68.0	-	-	-
		Meng et al. [57]	84.04	75.10	73.29	90.11	84.02	76.97
		A-PointNet [49]	58.62	52.28	47.23	-	-	-
		VoxelNet [36]	81.97	65.46	62.85	-	-	-
		MVF [54]	90.23	79.12	76.43	-	-	-
		DBQ-SSD [71]	89.60	79.56	78.51	-	-	-
		BtcDet [72]	93.15	86.28	83.86	-	-	-
	test	VoxelNet [36]	77.47	65.11	57.73	-	-	-
		Wang et al. [51]	70.58	62.71	55.17	-	-	-
		Zhang et al. [61]	80.25	69.21	60.15	-	-	-
		Meng et al. [57]	80.15	69.61	63.71	90.11	84.02	76.9
		Yang et al. [52]	86.61	77.63	76.06	89.66	87.76	86.89
		BADet [74]	89.28	81.61	76.58	95.23	91.32	86.48
		Li et al. [76]	78.68	61.46	60.09	88.07	86.26	69.79
		SIENet [73]	88.22	81.71	77.22	92.38	88.65	86.03
		EPNet++ [75]	91.37	81.96	76.71	-	-	-
		PointRCNN+SAT-GCN [99]	86.55	78.12	73.72	92.83	88.06	83.51

The 3D object detection and recognition have been evaluated on the KITTI dataset. Some researchers chose to evaluate their method on validation sets only while others opted for validation and test sets like VoxelNet in [36].

Results obtained for each method are assessed using Average Precision metric (AP) on two modalities including 3D car and BEV Car. Table 3 represents the respective results for validation and test sets.

On the KITTI validation set, we note that the approaches under investigation: [36], [54], and [71], succeeded to detect 3D objects with close results for easy, moderate, and hard metrics in 3D car modality. However, results of [49] are 32% below the utmost one, for the first metric (Easy). For the BEV Car, only two methods investigated this modality, and the outcome showed improvements as per the methods evaluated on 3D car modality. For example, in [57] the accuracy for 3D car(easy) is 84.04% whilst it hits 90.11% for BEV Car, with a difference of 6%.

On the KITTI test set, we notice that more methods performed the evaluation for BEV Car as opposed to the validation set. For 3D car modality, EPNet++ [75] reached the highest performance for easy and moderate modalities, while it came in second place for hard modality with a precision of 76.71. For BEV car modality, BADet [74] achieved the highest precision results for easy modality with 95.23 outperforming PointRCNN+SAT-GCN [99] by 2%. The same observation applies to moderate categories. Meanwhile, results recorded by [52] are the best for hard metrics.

The Average Precision metric (AP) is also considered for evaluation on Pedestrian and Cyclist modalities according to the same difficulty flags: easy, moderate, and hard. Table 4 details the results of different methods. Comparing the evaluation results on 3D Car and BEV Car, we note that the AP rates on Pedestrian and Cyclist fall behind those registered for 3D Car and BEV Car modalities. This difference demonstrates that Pedestrians and Cyclists are more challenging to process due to intrinsic characteristics and the complexity of these modalities compared to the Car modality. This makes real-time usage hard to be implemented at present.

With reference to the same table (Table 4), we observe that the solution detailed in [51] yields better results in terms of AP values on the test set, and outperforms EPNet++ [75] by a difference under 1% for Pedestrian and 7-13% for Cyclist modality. On the Val set, the method in [57] hits the best AP values for Pedestrian modality yet has not been performed on Cyclist. On the other hand, the method in [52] reached the second-best outcome for Pedestrians and SIENet [73] reach the highest one for cyclists.

Table 4. Evaluation results for Pedestrian and Cyclist on KITTI dataset

Dataset	Test/Val	Method	Pedestrian			Cyclist		
			Easy	Moderate	Hard	Easy	Moderate	Hard
KITTI	test	Zhang et al. [61]	51.04	43.23	40.77	67.26	55.34	45.26
		VoxelNet [36]	39.48	33.69	31.51	61.22	48.36	44.37
		Yang et al. [52]	53.08	44.24	41.97	78.89	62.53	55.77
		Li et al. [76]	39.11	34.42	27.13	62.09	43.65	36.80
		SIENet [73]	-	-	-	83.00	67.61	60.09
		EPNet++ [75]	52.79	44.38	41.29	76.15	59.71	53.67
	val	VoxelNet [36]	57.86	53.42	48.87	67.17	47.65	45.11
		Meng et al. [57]	74.65	69.96	66.49	-	-	-
		BtcDet [72]	69.39	61.19	55.86	91.45	74.70	70.08

4.2 *Object shape classification on ModelNet10, odelNet40, and MNIST datasets*

To perform 3D object classification from 3D point cloud data, the referred methods involved three public datasets namely ModelNet10, ModelNet50, and 3D MNIST. Relevant results are shown in Table 5, for ModelNet10 and ModelNet40 datasets, and in Table 6 for the MNIST dataset.

The results on the ModelNet10 (displayed in Table 5) demonstrate that most methods succeeded to classify 3D objects with approximate accuracy values except for TZSL [41] Which only reached 46.9%. The best results were recorded by SO-Net [37] and VWN [42], and they outperformed TZSL by a large margin of 49%.

Table 5. Object shape classification results of methods On ModNet10 and ModNet40

Type	Method	ModeNet10	ModNet40
Point	TZSL [41]	46.9	-
	Lonchanet [24]	93.3	-
	VWN [42]	95.1	93.8
	SO-Net [37]	95.7	93.4
	VoxNet [58]	93.9	89.7
	RS-CNN [45]	-	93.6
	PointNet++ [29]	-	91.9
	Mao et al. [48]	-	93.0
	SFCNN [50]	91.4	92.3
	Xie et al. [34]	-	90.0
	PVNet [35] AlexNet	-	93.2
	DGCNN [53]	-	93.5
	SCN [34]	-	90.0
	USAE-BLS [64]	92.6	-
	3DHMNN [63]	90.9	83.8
	Gezawa et al. [65]	93.4	88.2
	DC-GNN [66]	-	93.6
View	R2-MLP (20 views) [67]	97.7	99.6
	OVPT (9 views) [100]	97.2	99.3
	ReINView (12 views) [68]	96.1	97.8

Results achieved by Lonchanet [24], VoxNet [58], and SFCNN [50] are mostly close: the difference between remained below 2%. This proves that the aforementioned approaches were fairly efficient and accurate in terms of 3D object recognition, yet there is always scope for improvement.

Other methods chose ModelNet40 dataset for their evaluation as shown in Table 5. Respective results indicate that VWN reached the highest accuracy value of 93.8% while RS-CNN and DC-GNN achieved 93.6% with a difference of 0.2%. The same difference was recorded between RS-CNN and SO-Net [37], PVNet and SO-Net. Also, we notice that the difference between the highest and the lowest results stayed under 4% which is between VWN, VoxNet, and DC-GNN. This demonstrates the effectiveness of these solutions as well as the complexity of ModelNet40 compared to ModelNet10.

For the Multi-view-based approaches that were evaluated on ModelNet10 and Model-Net40 datasets are presented in Table 5. Within, we presented obtained accuracies while taking into account the number of views in the training process. We observe that R2-MLP [67] achieved the highest performance accuracies for both datasets using 20 views. While OVPT [100] method reached the second place with an accuracy of 79.2% on Mod-eNet10 and 99.3% on ModNet40 using 9 views. Using 12 views ReINView [68] reached high-performance accuracy for both datasets. In terms of the number of views, we can de-duce that OVPT [100] reached satisfactory results with less number of views compared to R2-MLP [67] and ReINView [68].

Table 6. MNIST digit classification

Method	Input size	Error rate
PointNet [28]	256*2	0.78
PointNet++ [29]	512*2	0.51
SO-Net [37]	512*2	0.44
SCN [34]	256*2	0.60

Table 6 presents the object classification of 3D MINST dataset. The table shows that SO-Net [37] achieved better performance in terms of error rate, followed by PointNet++ [29] with 0.51 as the error rate and a difference of 0.1%. These results were produced using an input size of 512. Methods with higher error rates used an input data size of 256. This demonstrates that the size of input data can affect system effectiveness on 3D MNIST dataset.

4.3 *Object segmentation evaluation On ShapeNet Parts*

Objects represented with 3D point clouds have made the segmentation more complicated because of confusing point clouds to noisy points. Therefore, the point cloud pre-processing step should precede the segmentation process. The proposed object segmentation methods on 3D point clouds were trained and evaluated on ShapeNet parts. Table 7 compares all these methods with 16 object categories and a presentation of the best results for each category.

In Table 7, we find that some methods reached a good performance for the 16 categories like RS-CNN, LSDTM-GCNN, VoxRec, and DC-GNN, which hit the best results. While other methods only achieved high accuracies for selected categories. In terms of Mean values, we find that SFCNN [50] came in second place by a value of 85.4% with a difference of 0.8% as opposed to RS-CNN [45]. From the means values we notice that all results are close and the difference between the highest and lowest values is only 5.3%. Also for some categories like Motor, Rocket, and Skate, results are lower than some other categories like Car, Gui, Mug, and Knife. Figure 6 illustrates some examples of the segmented part of objects from the ShapNet Part dataset obtained using the VoxRec method.

Table 7. Object part segmentation results on ShapeNetPart dataset.

Method	Mean	Aero	Bag	Cap	Chair	Car	Ear-ph	Gui	Knife	Lamp	Laptop	Motor	Mug	Pistol	Rocket	Skate	Table
PointNet [28]	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
VoxRec [56]	81.5	75.6	83.0	83.4	83.9	73.9	75.3	88.6	88.4	73.3	94.5	72.6	96.5	80.3	69.1	83.8	-
SO-Net [37]	84.6	81.9	83.5	84.8	78.1	90.8	72.2	90.1	83.6	82.3	95.2	69.3	94.2	80.0	51.6	72.1	82.6
RS-CNN [45]	86.2	83.5	84.8	88.8	79.6	91.2	81.1	91.6	88.4	86.0	96.0	73.7	94.1	83.4	60.5	77.7	83.6
SFCNN [50]	85.4	83.0	83.4	87.0	80.2	90.1	75.9	91.1	86.2	84.2	96.7	69.5	94.8	82.5	59.9	75.1	82.9
A-SCN [34]	84.6	83.8	80.8	83.5	79.3	90.5	69.8	91.7	86.5	82.9	96.0	69.2	93.8	82.5	62.9	74.4	80.8
DGCNN [53]	85.2	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
LSDTM-GCNN [92]	-	84.4	86.6	90.2	79.9	91.9	80.6	95.3	90.0	83.2	96.1	71.7	95.2	80.3	68.1	79.3	97.2
DC-GNN [66]	-	85.3	82.2	86.0	81.0	91.3	78.9	92.2	88.0	85.0	95.8	72.8	95.2	84.5	54.2	76.9	82.9
AGNet [93]	-	84.1	83.2	86.0	78.8	90.6	76.9	91.9	88.4	82.3	96.0	65.5	93.7	84.2	64.2	76.8	80.6
DANet [91]	-	83.9	83.2	85.0	79.7	91.1	77.3	91.9	88.4	84.8	95.7	71.9	94.7	83.2	58.0	75.1	82.8
3D-SelfCutMix [95]	-	82.4	72.0	79.1	76.9	90.6	75.2	91.6	88.6	85.3	96.1	67.8	92.0	79.2	57.8	72.9	83.5

4.4 *3D object reconstruction evaluation*

To evaluate 3D object reconstruction methods, authors used the ShapeNet dataset which contains segmented features. The metrics used for this evaluation for 3D point cloud reconstruction are Chamfer distance (CD), and Earth mover's distance (EMD). These metrics are defined in [72] as follows:

$$CD\left(P,\hat{P}\right)=\frac{1}{|P|}\sum_{p\in P}\min_{\hat{p}\in\hat{P}}\|p-\hat{p}\|_2^2+\frac{1}{|\hat{P}|}\sum_{\hat{p}\in\hat{P}}\min_{p\in P}\|p-\hat{p}\|_2^2$$

While given predicted point clouds $\hat{p}\in\hat{P}$ and the ground truth $p\in P$.

$$EMD\left(P,\hat{P}\right)=\min_{\phi:P\to\hat{P}}\frac{1}{|P|}\sum_{p\in P}\|p-\phi p\|$$

Where $\phi:P\to\hat{P}$ denotes a bijection that minimizes the average distance between corresponding points.

Table 8. Quantitative comparison of different object reconstruction proposed models on ShapeNet

Category	CD				EMD			
	[84]	[83]	[85]	[86]	[84]	[83]	[85]	[86]
plane	0.430	0.477	0.386	0.329	0.396	0.579	0.527	0.382
bench	0.629	0.624	0.436	0.459	1.113	0.965	0.815	0.431
cabinet	0.439	0.381	0.373	0.607	2.986	2.563	2.147	0.494
car	0.333	0.268	0.308	0.439	1.747	1.297	1.306	0.361
chair	0.645	0.610	0.606	0.48	1.946	1.399	1.257	0.645
monitor	0.722	0.755	0.501	0.58	1.891	1.536	1.314	0.845
lamp	1.193	1.295	0.969	0.639	1.222	1.314	1.007	0.594
speaker	0.756	0.739	0.632	2.89	3.490	2.951	2.441	0.425
firearm	0.423	0.453	0.463	0.585	0.397	0.667	0.572	0.503
couch	0.549	0.490	0.439	0.390	2.207	1.642	1.536	0.737
table	0.517	0.498	0.589	0.626	2.121	1.480	1.340	0.605
cellphone	0.438	0.421	0.332	0.427	1.019	0.724	0.674	0.377
watercraft	0.633	0.670	0.478	0.555	0.945	0.814	0.730	0.489
mean	0.593	0.591	0.501	0.538	1.653	1.380	1.205	0.530

3D object reconstruction methods were trained and tested on the ShapeNet dataset. This dataset includes synthetic images from different viewpoints, wherein 13 categories are considered during the evaluation process. Individual results are presented in Table 8 for each category. From the Table, we find that some methods generated point clouds for 3D objects with considerable accuracy using the two metrics CD and EMD. This translates into generating completed shapes and surfaces despite the presence of varied objects in the dataset. The comparison of relevant results reported in Table 8, shows that for the CD metric the [84] method reached the highest values for all categories.

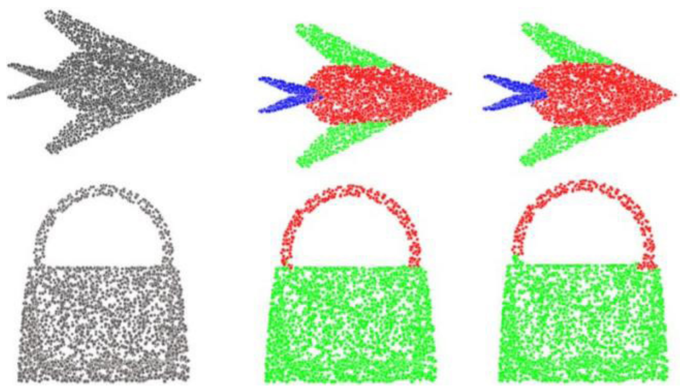


Figure 6. Segmented results of some example from Shapenet Parts dataset using the method in [56] (VoxRec). Left: 3D point cloud. Middle: Ground-truth. Right: the segmented results

While [83] method reached the second-best results for some categories like Plan, Monitor, Lamp, and Watercraft. Also, [86] yielded the best results for many categories like Cabinet, Car, Speaker, Firearm and Table (see Figure 7). When it comes to EMD metric, [84] reached the highest results for all categories except for Lamp, Speaker, and Firearm. While [84] method came in the second place for most categories. [86] results using EMD are the lowest compared to previous methods with a difference of 0.01 to 2.5 points. To summarize, it is clear that the experimental results are convincing for certain categories yet need more work for others.

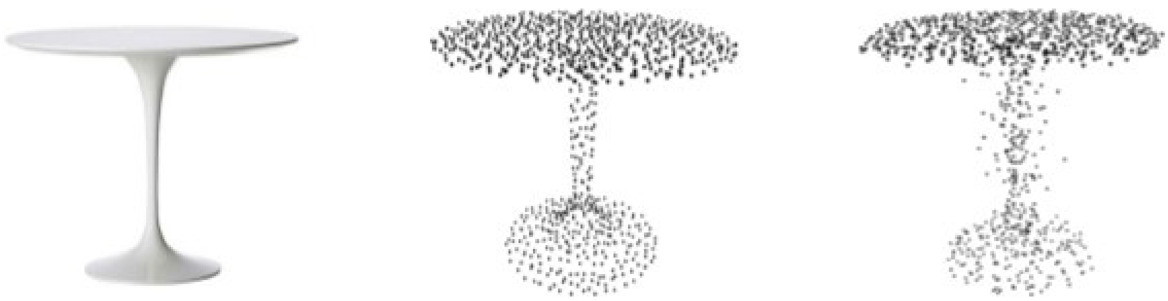


Figure 7. Generated 3D point cloud table using the method in [86]. Left: original image. Middle: Ground-truth. Right: the generated table

As an illustration of 3D object reconstruction, Figure 8 represents the output of [101] approach applied to 2D-3D-Semantics Dataset. The evaluation was carried out on a set composed of a tile size of 1 million 3D points runnable on a normal machine CPU (the tile size can be bigger if working with a high machine performance). The figure reveals that some parts are not well segmented due to the noise and occlusions in the dataset. Besides, due to feature variance, some parts are over-segmented.

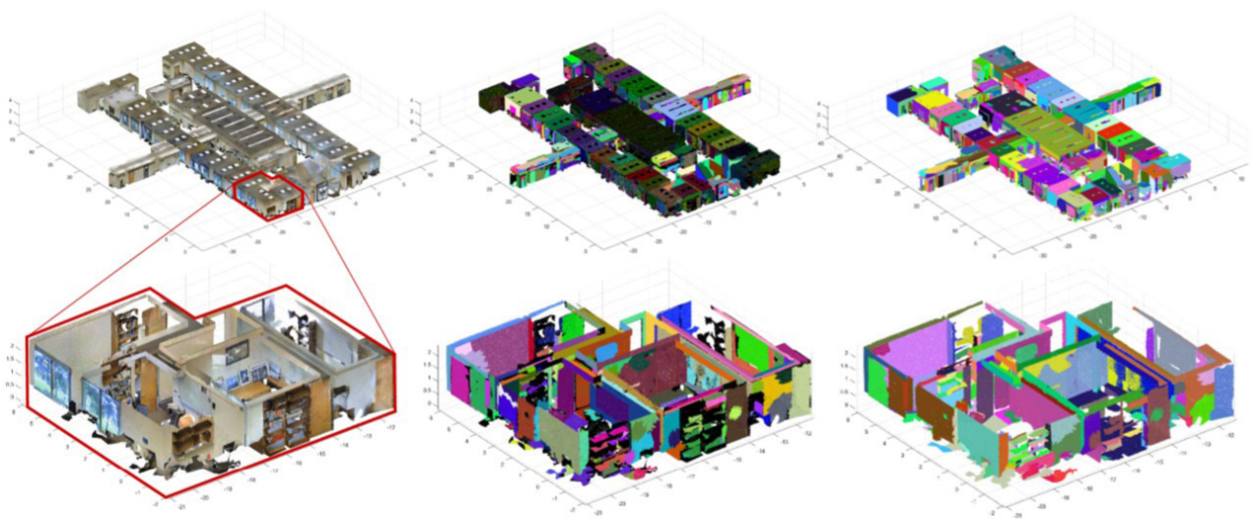


Figure 8. The obtained results of [101] on 2D-3D-S dataset

5. Challenges and future directions

Point cloud representation is a new type of data that has been used to support and enhance many computer-aided applications. It comes with various benefits in terms of data transfer cost, efficient shredding data including multi-view data representation of objects, and the ability to reconstruct and generate real objects from point cloud data. However, this technique often faces challenges pertaining essentially to data analysis due to the presence of noisy points, blind and cluttered scenes, uncompleted parts of the objects, etc. In the following, we will present current issues of the 3D point cloud as well as some of the solutions and future trends intended to solve each one of them.

5.1 *Incomplete, noisy, and unstructured data*

Using dots of XYZ coordinates can represent a 3D synthetic object, but using a sensor like LiDAR, these coordinates cannot be perfectly represented due to the number of points (dots) captured as well as the distribution of these points. The number of points generated can be considered as a cloud of noise while there are many objects in a complex scene. For the same reason, the gathered data can contain some uncompleted parts of the objects. These parts can be important for recognizing the object in order to segment it or reconstruct it. Also using LiDAR sensor, the generated point cloud represents just one side of the object. Also, we can find that some regions have a sparse swarm of dots while others have dense dots which makes the data irregular because the distance between the points is varying due to the fact that each point is scanned independently.

In order to overcome this problem, researchers attempted to use deep learning techniques to complete the missing parts of 3D point cloud objects with reconstruction methods [102]. Also, the proposed methods tried to recognize the objects even if represented with a minimum swarm of dots.

Another issue facing deep learning methods (specifically the ones including convolutional neural networks (CNN)) is that they are usually trained on structured, ordered and regular data. So, to face 3D point cloud data discrepancy, researchers decided to convert the point cloud into a structured grid presentation to serve as input for their models.

5.2 *Hard to store and process*

LiDAR data that contains information on the 3D objects (represented with points in 3D space), can result in massive files to be stored and transferred [103]. This tremendous amount of information makes the data hard to process and handle. When it comes to training a model with a large-scale dataset, the process is costly in terms of memory and demands high-performance machines. Nevertheless, the use of multi-dimensional vectors or tensors of the point cloud reduces the size of each file and enables the machine to process them more easily [104]. Also, the use of compact representations by encoding 3D shapes can make the classification and recognition tasks more doable [105].

6. Conclusion

3D point cloud is a new representation of data besides the existing ones such as image, video, and time-series data. This presentation can give much information using swarm dots about the object or the scenes represented in 3D shapes. Working on this data for recognizing the object segmentation andthe part of these objects or also reconstructing objects and scenes is a new breakthrough in computer vision. It has opened multiple opportuni ties in different research and development fields, wherein complete information about the objects is required, such as autonomous driving, remote sensing image mapping, object and scene reconstruction, etc. To inform the state-of-the-art, we proposed an extensive critical review of the tasks performed on 3D point cloud, which has been designed following a well-defined methodology. Accordingly, different tasks were first presented. Next, we listed the datasets used for performance evaluation. Also, we drew a comparison of the obtained results for each task using different evaluation metrics. A list of challenges of the methods working on 3D point cloud was provided with potential solutions and future directions.

Acknowledgments

This research work was made possible by research grant support (QUEX-CENG-SCDL-19/20-1) from Supreme Committee for Delivery and Legacy (SC) in Qatar.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Rusu R, Cousins S. 3d is here: Point cloud library (pcl). *2011 IEEE international conference on robotics and automation*. Shanghai, China: IEEE; 2011. p.1-4. Available from: https://doi.org/10.1109/ICRA.2011.5980567.

[2] Gao D, Liu J, Wu R, Cheng D, Fan X, Tang X. Utilizing relevant RGB-D data to help recognize RGB images in the target domain. *International Journal of Applied Mathematics and Computer Science*. 2019; 29(3): 611-621.

[3] Zheng Z, Yu T, Wei Y, Dai Q, Liu Y. Deephuman: 3d human reconstruction from a single image. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE; 2019. p.7739-7749. Available from: https://doi.org/10.1109/ICCV.2019.00783.

[4] Narayanan V, Likhachev M. Perch: Perception via search for multi-object recognition and localization. *In 2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE; 2016. p.5052-5059. Available from: https://doi.org/10.1109/ICRA.2016.7487711.

[5] Heisterkamp D, Bhatacharya P. Matching of 3D polygonal arcs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997; 19(1): 68-73.

[6] Li S. On topological consistency and realization. *Constrained*. 2006; 11: 31-51.

[7] Daftry S, Hoppe C, Bischof H. Building with drones: Accurate 3d facade re-construction using MAVs. *In 2015 IEEE International Conference on Robotics and Automation (ICRA)*. Seattle, WA, USA: IEEE; 2015. p.3487-3494. Available from: https://doi.org/10.1109/ICRA.2015.7139681.

[8] Han J, Shao L, Xu D, Shotton J. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics*. 2013; 43(5): 1318-1334.

[9] Xu W, Lee IS, Lee SK, Lu B, Lee EJ. MultiView-based hand posture recognition method based on point cloud. *KSII Transactions on Internet and Information Systems (TIIS)*. 2015; 9(7): 2585-2598.

[10] Park J, Ki H, Tai YW, Brown MS, Kweon I. High quality depth map upsampling for 3D-TOF cameras. *In 2011*

International Conference on Computer Vision. Barcelona, Spain: IEEE; 2011. p.1623-1630. Available from: https://doi.org/10.1109/ICCV.2011.6126423.
[11] Zaman F, Wong YP, Ng BY. Density-based denoising of point cloud. In: Ibrahim H, Iqbal S, Teoh S, Mustaffa M. (eds.) *9th International Conference on Robotic, Vision, Signal Processing and Power Applications. Lecture Notes in Electrical Engineering, vol 398*. Singapore: Springer; 2017. p.287-295. Available from: https://doi.org/10.1007/978-981-10-1721-6_31.
[12] Tian Y, Song W, Sun S, Fong S, Zou S. 3D object recognition method with multiple feature extraction from LiDAR point clouds. *The Journal of Supercomputing*. 2019; 75: 4430-4442. Available from: https://doi.org/10.1007/s11227-019-02830-9.
[13] Wang W, Zhou S, Li J, Li X, Yuan J, Jin Z. *Temporal pulses driven spiking neural network for fast object recognition in autonomous driving*. arXiv[Preprint]2001. Available from: https://doi.org/10.48550/arXiv.2001.09220.
[14] Kasaei H, Ghorbani M, Schilperoort J, Rest W. Investigating the importance of shape features, color constancy, color spaces, similarity measures in open-ended 3D object recognition. *Intelligent Service Robotics*. 2021; 14: 329-344.
[15] Dong S, Ren L, Zou K, Li W. 3D Object recognition method based on point cloud sequential coding. *In Proceedings of the 2020 12th International Conference on Machine Learning and Computing (ICMLC 2020)*. New York, NY, USA: Association for Computing Machinery; 2020. p.297-300. Available from: https://doi.org/10.1145/3383972.3383984.
[16] Akai N, Hirayama T, Murase H. Semantic localization considering uncertainty of object recognition. *IEEE Robotics and Automation Letters*. 2020; 5(3): 4384-4391.
[17] Elharrouss O, Almaadeed N, Al-Maadeed S. A review of video surveillance systems. *Journal of Visual Communication and Image Representation*. 2021; 77: 103116. Available from: https://doi.org/10.1016/j.jvcir.2021.103116.
[18] Akbari Y, Almaadeed N, Al-maadeed S, Elharrouss O. Applications, databases and open computer vision research from drone videos and images: A survey. *Artificial Intelligence Review*. 2021; 54(5): 3887-3938.
[19] Asif U, Bennamoun M, Sohel FA. Rgb-d object recognition and grasp detection using hierarchical cascaded forests. *IEEE Transactions on Robotics*. 2017; 33(3): 547-564.
[20] Buch AG, Kiforenko L, Kraft D. Rotational subgroup voting and pose clustering for robust 3d object recognition. *In 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE; 2017. p.4137-4145. Available from: https://doi.org/10.1109/iccv.2017.443.
[21] Börcs A, Nagy B, Benedek C. Instant object detection in lidar point clouds. *IEEE Geoscience and Remote Sensing Letters*. 2017; 14(7): 992-996. Available from: https://doi.org/10.1109/LGRS.2017.2674799.
[22] Falco P, Lu S, Cirillo A, Natale C, Pirozzi S, Lee D. Cross-modal visuo-tactile object recognition using robotic active exploration. *In 2017 IEEE International Conference on Robotics and Automation (ICRA)*. Singapore: IEEE; 2017. p.5273-5280. Available from: https://doi.org/10.1109/ICRA.2017.7989619.
[23] Fehr D, Beksi W, Zermas D. Papanikolopoulos N. Covariance based point cloud descriptors for object detection and recognition. *Computer Vision and Image Understanding*. 2016; 142: 80-93. Available from: https://doi.org/10.1016/j.cviu.2015.06.008.
[24] Gomez-Donoso F, Garcia-Garcia A, Garcia-Rodriguez J, Orts-Escolano S, Cazorla M. Lonchanet: A sliced-based cnn architecture for real-time 3d object recognition. *In 2017 International Joint Conference on Neural Networks (IJCNN)*. Anchorage, AK, USA: IEEE; 2017. p.412-418. Available from: https://doi.org/10.1109/IJCNN.2017.7965883.
[25] Kasaei SH, Tomé AM, Lopes LS, Oliveira M. GOOD: A global orthographic object descriptor for 3D object recognition and manipulation. *Pattern Recognition Letters*. 2016; 83: 312-320.
[26] Luo S, Mou W, Althoefer K, Liu H. Iterative closest labeled point for tactile ob-ject shape recognition. *In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, Korea (South): IEEE; 2016. p.3137-3142. Available from: https://doi.org/10.1109/IROS.2016.7759485.
[27] Li B. 3d fully convolutional network for vehicle detection in point cloud. *In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vancouver, BC, Canada: IEEE; 2017. p.1513-1518. Available from: https://doi.org/10.1109/IROS.2017.8205955.
[28] Qi CR, Su H, Mo K, Guibas LJ. Pointnet: Deep learning on point sets for 3d classification and segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE; 2017. p.652-660. Available from: https://doi.org/10.1109/CVPR.2017.16.
[29] Qi CR, Yi L, Su H, Guibas LJ. Pointnet++: Deep hierarchical feature learning on point sets in a metric space.

In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Red Hook, NY, USA: Curran Associates Inc.; 2017. p.5105-5114. Available from: https://dl.acm.org/doi/10.5555/3295222.3295263.
[30] Gojcic Z, Zhou C, Wegner JD, Wieser A. The perfect match: 3d point cloud matching with smoothed densities. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Beach, CA, USA: IEEE; 2019. p.5545-5554. Available from: https://doi.org/10.1109/CVPR.2019.00569.
[31] Wang J, Lindenbergh R, Menenti M. Sigvox-a 3d feature matching algorithm for automatic street object recognition in mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2017; 128: 111-129.
[32] Rangel JC, Martínez-Gómez J, Romero-González C, García-Varea I, Cazorla M. Semi-supervised 3d object recognition through CNN labeling. *Applied Soft Computing*. 2018; 65: 603-613. Available from: https://doi.org/10.1016/j.asoc.2018.02.005.
[33] Uy MA, Lee GH. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Salt Lake City, UT, USA: IEEE; 2018. p.4470-4479. Available from: https://doi.org/10.1109/CVPR.2018.00470.
[34] Xie S, Liu S, Chen Z, Tu Z. Attentional shapecontextnet for point cloud recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE; 2018. p.4606-4615. Available from: https://doi.org/10.1109/CVPR.2018.00484.
[35] You H, Feng Y, Ji R, Gao Y. Pvnet: A joint convolutional network of point cloud and multi-view for 3d shape recognition. *MM '18: Proceedings of the 26th ACM international conference on Multimedia*. New York, NY, United States: IEEE; 2018. p.1310-1318. Available from: https://doi.org/10.1145/3240508.3240702.
[36] Zhou Y, Tuzel O. Voxelnet: End-to-end learning for point cloud based 3d object detection. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA: IEEE; 2018. p.4490-4499. Available from: https://doi.org/10.1109/CVPR.2018.00472.
[37] Li J, Chen B, Lee G. So-net: Self-organizing network for point cloud analysis. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE. 2018. p.9397-9406. Available from: https://doi.org/10.1109/CVPR.2018.00979.
[38] Ahmed A, Jalal A, Kim K. RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and Hough voting. *In 2020 17th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. Islamabad, Pakistan: IEEE; 2020. p.290-295. Available from: https://doi.org/10.1109/IBCAST47879.2020.9044545.
[39] Chen J, Kira Z, Cho Y. Deep learning approach to point cloud scene under-standing for automated scan to 3d reconstruction. *Journal of Computing in Civil Engineering*. 2019; 33: 04019027. Available from: https://doi.org/10.1061/(asce)cp.1943-5487.0000842.
[40] Cheraghian A, Rahman S, Petersson L. Zero-shot learning of 3d point cloud objects. *In 2019 16th International Conference on Machine Vision Applications (MVA)*. Tokyo, Japan: IEEE; 2019. p.1-6. Available from: https://doi.org/10.23919/MVA.2019.8758063.
[41] Cheraghian A, Rahman S, Campbell D, Petersson L. Transductive zero-shot learning for 3d point cloud classification. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Snowmass, CO, USA: IEEE. 2020. Available from: https://doi.org/10.1109/WACV45572.2020.9093545.
[42] Huang Q, Wang Y, Yin Z. View-based weight network for 3d object recognition. *Image and Vision Computing*. 2020; 93: 103828.
[43] Li R, Li X, Fu CW, Cohen-Or D, Heng PA. Pu-gan: A point cloud upsampling adversarial network. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE; 2019. p.7203-7212. Available from: https://doi.org/10.1109/ICCV.2019.00730.
[44] Liu H, Cong Y, Yang C, Tang Y. Efficient 3d object recognition via geometric information preservation. *Pattern Recognition*. 2019; 92: 135-145. Available from: https://doi.org/10.1016/j.patcog.2019.03.025.
[45] Liu Y, Fan B, Xiang S, Pan C. Relation-shape convolutional neural network for point cloud analysis. *In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE; 2019. p.8887-8896. Available from: https://doi.org/10.1109/CVPR.2019.00910.
[46] Liu X, Yan M, Bohg J. Meteornet: Deep learning on dynamic 3d point cloud se quences. *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE; 2019. Available from: https://doi.org/10.1109/ICCV.2019.00934.
[47] Luo Z, Li J, Xiao Z, Mou ZG, Cai X, Wang C. Learning high-level features by fusing multi-view representation of mls point clouds for 3d object recognition in road environments. *ISPRS Journal of Photogrammetry and Remote*

Sensing. 2019; 150: 44-58.
[48] Mao J, Wang X, Li H. Interpolated convolutional networks for 3d point cloud understanding. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE; 2019. p.1578-1587. Available from: https://doi.org/10.1109/ICCV.2019.00166.
[49] Paigwar A, Erkent O, Wolf C, Laugier C. Attentional pointnet for 3d-object detection in point clouds. *In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Long Beach, CA, USA: IEEE; 2019. p.1297-1306. Available from: https://doi.org/10.1109/CVPRW.2019.00169.
[50] Rao Y, Lu J, Zhou J. Spherical fractal convolutional neural networks for point cloud recognition. *In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE; 2019. p.452-460. Available from: https://doi.org/10.1109/CVPR.2019.00054.
[51] Wang H, Lou X, Cai Y, Li Y. Chen L. Real-time vehicle detection algorithm based on vision and lidar point cloud fusion. *Journal of Sensors*. 2019; 2019: 1-9. Available from: https://doi.org/10.1155/2019/8473980.
[52] Yang Z, Sun Y, Liu S, Shen X, Jia J. Std: Sparse-to-dense 3d object detector for point cloud. *In 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE; 2019. p.1951-1960. Available from: https://doi.org/10.1109/ICCV.2019.00204.
[53] Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*. 2019; 38(5): 1-12.
[54] Zhou Y, Sun P, Zhang Y, Anguelov D, Gao J, Ouyang T, et al. End-to-end multi-view fusion for 3d object detection in lidar point clouds. *In Conference on Robot Learning (PMLR)*. Osaka, Japan; 2020. p.923-932. Available from: http://proceedings.mlr.press/v100/zhou20a/zhou20a.pdf.
[55] Li D, Wang H, Liu N, Wang X, Xu J. 3d object recognition and pose estimation from point cloud using stably observed point pair feature. *IEEE Access*. 2020; 8: 44335-44345.
[56] Karambakhsh A, Sheng B, Li P, Yang P, Jung Y, Feng DD. Voxrec: Hybrid convolutional neural network for active 3d object recognition. *IEEE Access*. 2020; 8: 70969-70980.
[57] Meng Q, Wang W, Zhou T, Shen J, Dai D. Weakly supervised 3d object detection from lidar point cloud. *In Computer Vision-ECCV 2020: 16th European Conference*. Glasgow, UK: IEEE; 2020. p.515-531. Available from: https://doi.org/10.1007/978-3-030-58601-0_31.
[58] Sedaghat N, Zolfaghari M, Brox T. *Orientation-boosted voxel nets for 3D object recognition*. In: Kim TK, Zafeiriou S, Brostow G, Mikolajczyk KP. (eds.) *Proceedings of the British Machine Vision Conference (BMVC)*. BMVA Press; 2017. Available from: https://doi.org/10.5244/C.31.97.
[59] Melotti G, Premebida C, Gonçalves N. Multimodal deep-learning for object recog-nition combining camera and lidar data. *In 2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. Ponta Delgada, Portugal: IEEE; 2020. p.177-182. Available from: https://doi.org/10.1109/ICARSC49921.2020.9096138.
[60] Yang Y, Chen F, Wu F, Zeng D, Ji Y, Jing X. Multi-view semantic learning network for point cloud based 3d object detection. *Neurocomputing*. 2020; 397: 477-485. Available from: https://doi.org/10.1016/j.neucom.2019.10.116.
[61] Zhang K, Wang S, Ji L, Wang C. DNN based camera and lidar fusion framework for 3d object recognition. *Journal of Physics: Conference Series*. 2020; 1518: 012044.
[62] Huang J, You S. Point cloud labeling using 3d convolutional neural network. *In 2016 23rd International Conference on Pattern Recognition (ICPR)*. Cancun: IEEE; 2016. p.2670-2675. Available from: https://doi.org/10.1109/ICPR.2016.7900038.
[63] Lakhili Z, El Alami A, Mesbah A, Berrahou A, Qjidaa H. Rigid and non-rigid 3d shape classification based on 3d hahn moments neural networks model. *Multimedia Tools and Applications*. 2022; 81: 38067-38090. Available from: https://doi.org/10.1007/s11042-022-12125-6.
[64] Tian Y, Song W, Chen L, Fong S, Sung Y, Kwak J. A 3d object recognition method from lidar point cloud based on usae-bls. *IEEE Transactions on Intelligent Transportation Systems*. 2022; 23(9): 15267-15277. Available from: https://doi.org/10.1109/TITS.2021.3140112.
[65] Gezawa A, Bello Z, Wang Q, Yunqi L. A voxelized point clouds repre-sentation for object classification and segmentation on 3d data. *The Journal of Supercomputing*. 2022; 78(1): 1479-1500.
[66] Meraz M, Ansari M, Javed M, Chakraborty P. Dc-gnn: drop channel graph neural network for object classification and part segmentation in the point cloud. *International Journal of Multimedia Information Retrieval*. 2022; 11(2): 123-133.
[67] Chen S, Yu T, Li P. *R2-mlp: Round-roll mlp for multi-view 3d object recognition*. arXiv[Preprint]2022. Available from: https://doi.org/10.48550/arXiv.2211.11085.
[68] Xu R, Ma W, Mil Q, Zha H. Reinview: Re-interpreting views for multi-view 3d object recognition. *In 2022 IEEE/*

RSJ International Conference on Intelligent Robots and Systems (IROS). Kyoto, Japan: IEEE; 2022. p.6630-6636. Available from: https://doi.org/10.1109/IROS47612.2022.9981777.

[69] Lu Y, Yang F, Chen F, Xie D. *Pic-net: Point cloud and image collaboration network for large-scale place recognition*. arXiv[Preprint]2008. Available from: https://doi.org/10.48550/arXiv.2008.00658.

[70] Zhao H, Tang M, Ding H. Hoppf: A novel local surface descriptor for 3d object recognition. *Pattern Recognition*. 2020; 103: 107272.

[71] Yang J, Song L, Liu S, Li Z, Li X, Sun H, et al. Dbq-ssd: Dynamic ball query for efficient 3d object detection. *Computer Vision and Pattern Recognition*. arXiv[Preprint]2022. Available from: https://doi.org/arxiv-2207.10909.

[72] Xu Q, Zhong Y, Neumann U. Behind the curtain: Learning occluded shapes for 3d object detection. *In Proceedings of the AAAI Conference on Artificial Intelligence*. 2022; 36(3): 2893-2901.

[73] Li Z, Yao Y, Quan Z, Xie J, Yang W. Spatial information enhancement network for 3d object detection from point cloud. *Pattern Recognition*. 2022; 128: 108684.

[74] Qian R, Lai X, Li X. Badet: Boundary-aware 3d object detection from point clouds. *Pattern Recognition*. 2022; 125: 108524.

[75] Liu Z, Huang T, Li B, Chen X, Wang X, Bai X. Epnet++: Cascade bidirectional fusion for multi-modal 3d object detection. *IEEE Transactions on Pat-tern Analysis and Machine Intelligence.* 2022; 45(7): 8324-8341. Available from: https://doi.org/10.1109/TPAMI.2022.3228806

[76] Li J, Li R, Li J, Wang J, Wu Q, Liu X. Dual-view 3d object recognition and detection via lidar point cloud and camera image. *Robotics and Autonomous Systems*. 2022; 150: 103999.

[77] Avetisyan A, Dahnert M, Dai A, Savva M, Chang A, Nießner M. Scan2cad: Learning cad model alignment in RGB-D scans. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Beach, CA, USA: IEEE; 2019. p.2614-2623. Available from: https://doi.org/10.1109/CVPR.2019.00272.

[78] Bassier M, Vergauwen M. Topology reconstruction of bim wall objects from point cloud data. *Remote Sensing*. 2020; 12(11): 1800.

[79] Bassier M, Vergauwen M. Unsupervised reconstruction of building information modeling wall objects from point cloud data. *Automation in Construction*. 2020; 120: 103338.

[80] Han J, Rong M, Jiang H, Liu H, Shen S. Vectorized indoor surface reconstruction from 3d point cloud with multistep 2d optimization. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021; 177: 57-74.

[81] Chen J, Kira Z, Cho Y. Deep learning approach to point cloud scene understanding for automated scan to 3d reconstruction. *Journal of Computing in Civil Engineering*. 2019; 33(4): 04019027.

[82] Cupec R, Vidović I, Filko D, urović P. Object recognition based on convex hull alignment. *Pattern Recognition*. 2020; 102: 107199.

[83] Wang N, Zhang Y, Li Z, Fu Y, Liu W, Jiang Y. Pixel2mesh: Generating 3d mesh models from single rgb images. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y. (eds.) *Computer Vision-ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11215*. Springer, Cham; 2018. p.52-67. Available from: https://doi.org/10.1007/978-3-030-01252-6_4.

[84] Fan H, Su H, Guibas L. A point set generation network for 3d object recon-struction from a single image. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE; 2017. p.605-613. Available from: https://doi.org/10.1109/CVPR.2017.264.

[85] Zou C, Hoiem D. Silhouette guided point cloud reconstruction beyond occlusion. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Snowmass, CO, USA: IEEE; 2020. p.41-50. Available from: https://doi.org/10.1109/WACV45572.2020.9093611.

[86] Afifi A, Magnusson J, Soomro T. Hellwich O. Pixel2point: 3d object reconstruction from a single image using cnn and initial sphere. *IEEE Access*. 2020; 9: 110-121.

[87] Jin P, Liu S, Liu J, Huang H, Yang L, Weinmann M, et al. Weakly-supervised single-view dense 3d point cloud reconstruction via differentiable renderer. *Chinese Journal of Mechanical Engineering*. 2021; 34(1): 1-11.

[88] Nie Y, Hou J, Han X, Nießner M. Rfd-net: Point scene understanding by se-mantic instance reconstruction. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, TN, USA: IEEE; 2021. p.4608-4618. Available from: https://doi.org/10.1109/CVPR46437.2021.00458.

[89] Dai A, Chang AX, Savva M, Halber M, Funkhouser T, Nießner M. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. Honolulu, HI, USA: IEEE; 2017. p.5828-5839. Available from: https://doi.org/10.1109/CVPR.2017.261.

[90] Elharrouss O, Al-Maadeed S, Subramanian N, Ottakath N, Almaadeed N, Himeur Y. *Panoptic segmentation: A review*. arXiv[Preprint]2021. Available from: https://doi.org/10.48550/arXiv.2111.10250.

[91] Zhou C, Xie Y, He X, Yuan T, Ling Q. Dual attention network for point cloud classification and segmentation. *In*

2022 41st Chinese Control Conference (CCC). Hefei, China: IEEE; 2022. p.6482-6486. Available from: https://doi.org/10.23919/CCC55666.2022.9902385.

[92] Zhang WJ, Su SZ, Hong QQ, Wang BZ, Sun L. Long short-distance topology modelling of 3d point cloud segmentation with a graph convolution neural network. *IET Computer Vision*. 2023; 17(3): 251-264. Available from: https://doi.org/10.1049/cvi2.12160.

[93] Jing W, Zhang W, Li L, Di D, Chen G, Wang J. Agnet: An attention-based graph network for point cloud classification and segmentation. *Remote Sensing*. 2022; 14(4): 1036.

[94] Sheikh M, Asghar M, Bibi R, Malik M, Shorfuzzaman M, Mehmood R, et al. Dft-net: Deep feature transformation based network for object categorization and part segmentation in 3-dimensional point clouds. *Sensors*. 2022; 22(7): 2512.

[95] Xu Y, Ji Y, Huang S, Lin Z, Wang Y. 3d-selfcutmix: Self-supervised learning for 3d point cloud analysis. *In 2022 IEEE International Con-ference on Image Processing (ICIP)*. Bordeaux, France: IEEE; 2022. p.676-680. Available from: https://doi.org/10.1109/ICIP46576.2022.9897785.

[96] Wang Q, Tan Y, Mei Z. Computational methods of acquisition and processing of 3d point cloud data for construction applications. *Archives of Computational Methods in Engineering*. 2020; 27(2): 479-499.

[97] Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition*. Miami, FL, USA: IEEE; 2009. p.248-255. Available from: https://doi.org/10.1109/CVPR.2009.5206848.

[98] Cao Q, Shen L, Xie W, Parkhi O, Zisserman A. Vggface2: A dataset for recognising faces across pose and age. *In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. Xi'an, China: IEEE; 2018. p.67-74. Available from: https://doi.org/10.1109/FG.2018.00020.

[99] Wang L, Song Z, Zhang X, Wang C, Zhang G, Zhu L, et al. Sat-gcn: Self-attention graph convolutional network-based 3d object detection for autonomous driving. *Knowledge-Based Systems*. 2023; 259: 110080.

[100]Wang W, Chen G, Zhou H, Wang X. OVPT: Optimal Viewset Pooling Transformer for 3D Object Recognition. In: Wang L, Gall J, Chin TJ, Sato I, Chellappa R. (eds.) *Computer Vision-ACCV 2022. ACCV 2022. Lecture Notes in Computer Science, vol 13841*. Springer, Cham: IEEE; 2022. p.4444-4461. Available from: https://doi.org/10.1007/978-3-031-26319-4_29.

[101]Bassier M, Bonduel M, Van Genechten B, Vergauwen M. Segmentation of large unstructured point clouds using octree-based region growing and conditional random fields. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2017; 42: 25-30.

[102]Spurek P, Kasymov A, Mazur M, Janik D, Tadeja S, Struski L, et al. Hyperpocket: Generative point cloud completion. *Computer Vision and Pattern Recognition*. arXiv[Preprint]2021. Available from: https://doi.org/10.48550/arXiv.2102.05973.

[103]Spurek P, Zieba M, Tabor J, Trzcinski T. *Hyperflow: Representing 3d objects as surfaces*. Computer Vision and Pattern Recognition. arXiv[Preprint]2020. Available from: https://doi.org/10.48550/arXiv.2006.08710.

[104]Zamorski M, Zieba M, Klukowski P, Nowak R, Kurach K, Stokowiec W, et al. Adversarial autoencoders for compact representations of 3d point clouds. *Computer Vision and Image Understanding*. 2020; 193: 102921.

[105]Kania K, Zieba M, Kajdanowicz T. *Ucsg-net-unsupervised discovering of con-structive solid geometry tree. 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada: IEEE; 2020. p.8776-8786.