Bayesian Estimation for 3D-Point Object Identification Based on Probabilistic Field Representation

Hiroaki Sato, Shin'ichi Arakawa, Masayuki Murata
Graduate School of Information Science and Technology
Osaka University
Osaka, JAPAN
{hi-sato, arakawa, murata}@ist.osaka-u.ac.jp

Abstract—New network services using real spatial information are expected to emerge in remote areas. For the advancement of services, it is important to understand real space from real spatial information, and for this purpose, object identification using machine learning is widely employed. Instead of directly identifying real space information using machine learning techniques, in this study, we represented real space as a field of probabilistic superposition of objects, which incorporates the results of the object identification as well as information about neighboring objects. We obtained the neighbor information based on positional relationships of objects in real space from a dataset used by the machine learning algorithm, and built an empirical knowledge base of the positional relationships. Then, we developed a method to use the empirical knowledge to modify the object identification results. Our results show that the outcomes of object identification are changed by the empirical knowledge and the accuracy of the identification is improved when confidence in the machine learning algorithm is low.

Index Terms—point cloud; object detection; machine learning; semantic segmentation; Bayesian estimation

I. Introduction

New network services using real spatial information are expected to emerge in remote areas. Services such as remote shopping services that can provide a shopping center experience from home require understanding real space from real spatial information and information processing. Specifically, it is necessary to transfer real spatial information acquired remotely via a network and provide users with services obtained from real spatial information through applications. These network services need to understand real space from real spatial information. In fact, much research has been done on object identification in different situations, for example, for data on roads with a mixture of features such as vehicles, passersby, and buildings, and for indoor data with various types of furniture [1], [2]. These are among the tasks required in environments where information in real space must be recognized with high accuracy, such as automatic driving of automobiles and autonomous control of self-driving robots [3].

Currently, research is underway to acquire real-world information as 2D data such as video images and apply machine learning techniques such as deep learning to present real-world information to the user [4]. However, because real-

world information is reduced to 2D data, it cannot handle 3D information, such as overlapping objects. Therefore, in recent years, there has been much research on object identification using machine learning for 3D spatial information that can be acquired by Lidar sensors or depth cameras [5], [6]. Existing object identification methods are implemented by automatically learning features within the data from a large amount of training data. However, it has been pointed out that it is difficult for humans to understand the rules that guide the object identification results of existing machine learning methods [7]. Another issue is that object identification ignores real-space knowledge that we humans have, such as the fact that a kitchen and a sofa are unlikely to be adjacent to each other.

In this study, we developed a way to represent real space as a field of probabilistic superposition of objects, instead of directly identifying real space information using machine learning techniques and other methods. A field of probabilistic superposition of objects is a field that does not uniquely identify objects in real space but is represented by the probabilistic superposition of object categories. Specifically, we focused not on estimating objects for specific scenes, but rather on using probability to quantify the empirical knowledge that we potentially acquired during our lives. To represent the field of probabilistic superposition, we focused on the information of neighboring objects. By collecting object proximity information from a large amount of data, for example, a chair and a desk being in close proximity, we quantified latent and empirical knowledge in real space as statistical information. We obtained the neighbor information based on the positional relationships of objects in real space from a dataset used by machine learning, and the correlation between object adjacencies in real space was statistically determined. We used 3D point cloud data collected for the machine learning to obtain statistical information in real space, and point cloud data in real space was represented as a probabilistic field. In this study, we obtained probability fields from point cloud data reflecting the indoor environment of over thousands rooms, focusing on the number of objects and their adjacencies for 20 different objects, such as desks, chairs, beds, and

windows. Furthermore, to examine the application of the field of probabilistic superposition, we examined the modification of object identification results by a Bayesian estimation. This will aim to improve recognition of real-space information for point cloud data with the empirical knowledge.

II. PROBABILISTIC FIELD REPRESENTATION OF REAL SPACE INFORMATION AND ACQUISITION METHOD

A. Real Space Information

As real space information, we focused on information on the position and category of objects that exist in real space. For example, we handled information on the position and category of objects such as furniture for indoor data and objects such as cars and people for outdoor data. By handling such information, various applications are possible, for example, detection of obstacles necessary for automatic driving and autonomous control, and recognition of information on products desired by users in new network services, such as remote shopping services.

We used 3D point cloud data as real-space information. Point cloud data summarize information on points in space, and have components that represent six basic parameters for each point: spatial coordinate information (X, Y, Z) and color information (R, G, B). 3D point cloud data can capture complex shapes captured by an RGB-D camera capable of simultaneously acquiring color and distance images and a laser scanner that irradiates an object to obtain the object's coordinates from the time it takes for the laser to return and the angle of the irradiation.

B. Probabilistic Field Representation

A probabilistic field is not a unique identification of objects in real space, but a field represented by a probabilistic superposition of object categories. In this study, we obtained adjacency information for the positions of objects in real space from many data, statistically determined what kind of correlation exists between object adjacencies in real space, and used this as prior knowledge of real space. By statistically determining what kind of correlations exist among neighboring objects in real space, it is possible to quantify the correlations that humans have, or to incorporate them into other methods as prior knowledge. In this study, we obtained numerical information on the number of objects and adjacencies between object categories by determining the existence and adjacency probability distributions of the objects shown in Sections II-B1 and II-B2. In this process, it was necessary to cut out objects, acquire their positions, and acquire their adjacencies, and we explain the detection method in Section II-C.

1) Probability distribution of object existence: We obtained statistical information on object existence from real space information. Using the object information existing in real space, we obtained the existence status in the space by calculating the existence probability of each category based on the number of objects in each category and the total number of objects.

2) Probability distribution of adjacent objects: We obtained statistical information on object adjacency from real space information. Using the object information in real space, we searched for adjacent objects based on the distance between the center-of-gravity point of each object. The probability of which categories of objects are adjacent to the object of interest was calculated, and the objects determined to be adjacent were aggregated to obtain the state of adjacentness in the space. The adjacency probability was calculated by dividing the number of adjacencies between the object category of interest and the adjacent object categories by the number of detected adjacencies of the object category of interest.

C. Object Detection Method for Probabilistic Field Representation

In this section, we describe the method for obtaining information on the location and category of objects from real-space information. There are several methods for object detection, such as unsupervised learning (e.g., clustering) and supervised learning (e.g., deep learning), but in this study, we used density-based clustering. Density-based clustering is an algorithm that separates areas with a concentration of points from sparse areas and detects areas with a concentration of points as objects. By performing density-based clustering on a point cloud consisting of only points in each category, the object location in each category is detected.

1) Density-based clustering: We used DBSCAN clustering [8] from Open3D [9], an open-source library for 3D data. The algorithm performs a search for points within a radius r of a certain point and determines a point to be a cluster if the reachable points exceed a threshold n, while reachable points that do not exceed the threshold n are considered noise. In our study, the threshold n of the cluster is set to be 10, which means that the radius r is 5 [cm], and the distance between the center of gravity points of object adjacencies is 1 [m].

2) Nearest neighbor search: As in Section II-C1, we used the construction of a KDTree using the fast nearest neighbor search library FLANN [10] from Open3D [9], an open-source library for 3D data. KDTree is a spatially partitioned data structure that classifies points in a multidimensional space, and FLANN is a library that speeds up the search, which varies with the number of dimensions and amount of data.

III. EXAMPLE OF PROBABILISTIC FIELD REPRESENTATION USING 3D POINT CLOUD DATASET

A. Real Space Data

A public dataset of indoor space was used to obtain a real space stochastic field representation. Specifically, we used ScanNet [11], an indoor 3D point cloud dataset. This dataset includes coordinate and color information for 1613 room scenes surveyed by an RGB-D camera and labeled by scene type, such as office, apartment, or bathroom. Figure 1 shows an example of 3D point cloud data.

Of the total of 1613 room scenes, 1201 scenes were used for training data to obtain the probabilistic field representation (the remainder were split between 312 scenes for validation

Fig. 1: Indoor 3D point cloud data
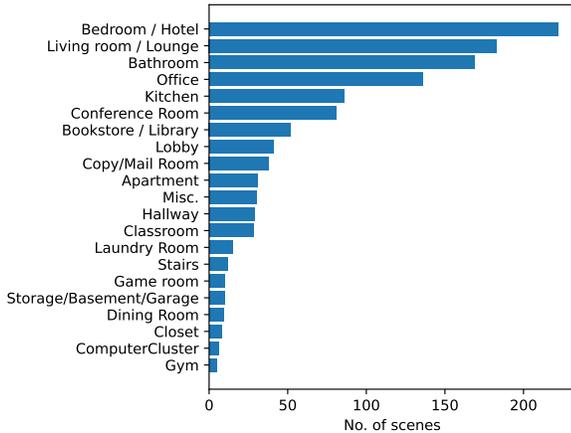
Fig. 2: Distribution of scene types in the ScanNet dataset

Fig. 3: Examples of object adjacencies

and 100 scenes for testing). These training scenes are annotated with surface reconstruction and semantic segmentation annotations in addition to the spatial coordinate and color information of the point cloud.

Figure 2 shows the distribution of scene type occurrence for the 1201 scenes in the ScanNet dataset. The vertical axis indicates the scene type name, and the horizontal axis indicates the number of scenes of each type. The ScanNet dataset, which is constructed from 21 scene types, contains a variety of spaces ranging from small-scale types (e.g., bathroom, closet, and unit bath) to large-scale types (e.g., apartment, classroom, and library).

Each scene contains several point cloud objects corresponding to objects in the room. Hereafter, point cloud objects are referred to as objects. The objects are manually annotated with an object category identifier (henceforth, object category) that represents the object. Table I shows the object category classifications assigned to objects in the ScanNet dataset. The category "other furniture" contains furniture not listed in the table, such as piano and display case. Although the ScanNet dataset includes information on walls and floors, the center of gravity of walls and floors can be far from the center of gravity of objects, making it impossible to properly extract the

adjacency relationship, so this paper excludes walls and floors to obtain real-space information.

B. Example of Probabilistic Field Representation

An example of a probabilistic field obtained by the method described in Section II is shown below.

Figure 3 shows the object adjacency in one scene when we obtain the probabilistic field representation. The points represent object center-of-gravity points, the colors represent object categories, and the lines represent adjacencies. From this figure, it can be seen that an object that is identical in real space is divided into multiple objects (e.g., areas **a** to **c**) due to noise and clarity differences that occur when acquiring point cloud data for objects using the method shown in Section II-C1. In addition, we can confirm the adjacency situations that appear in real space, such as the adjacency between sink, counter, and cabinet (area **c**); desk and chair (area **d**); and cabinet, toilet, bathtub, and shower curtain in a unit bath (area **e**). This indicates that we can obtain knowledge information for our daily life.

Information on the object presence probability distribution and object adjacency probability distribution for all 1201 scenes in the training data is shown in Tables II and III.

Table II shows a high probability of the presence of common room furniture, such as table, chair, and door. Next in probability are other furniture, cabinet, and desk furniture. This is because "other furniture" includes a variety of furniture such as pianos and display furniture, while "cabinet" and "desk" are common furniture often included in the scene type shown in Figure 2.

TABLE II: Existence probability for all scenes (%)

Label	Category	Prob	Label	Category	Prob
0	wall	-	10	picture	2.35
1	floor	-	11	counter	2.11
2	cabinet	7.54	12	desk	6.26
3	bed	4.98	13	curtain	1.27
4	chair	19.61	14	refrigerator	1.28
5	sofa	3.44	15	shower curtain	0.41
6	table	22.52	16	toilet	0.81
7	door	8.56	17	sink	1.45
8	window	5.55	18	bathtub	0.57
9	bookshelf	3.55	19	other furniture	7.76

We describe the characteristics obtained from Table III. First, we see that the diagonal values of Table III are generally large. For example, bookshelf (Label 9) has a probability of 92%, reflecting how common it is for a bookshelf to be placed next to another bookshelf. Shower curtains, toilets, sinks, and bathtubs (Labels 15-18) have lower probabilities but are still considered to be important. Multiple objects of the same type appear, even though one is sufficient for their function. This is because when an object in real space is acquired by the method shown in Section II-C1, it is divided into multiple identical objects owing to noise and differences in sharpness that occur during the acquisition of the point cloud data. In addition, unit bath-related shower curtains, toilets, and bathtubs (Labels 15, 16, and 18, respectively) have a high probability of mutual adjacency. Moreover, chairs (Label 4) have an adjacency of 38% for tables (Label 6), excluding diagonals, followed by 8% for desk (Label 12). In addition, there is no refrigerator (Label 14) adjacent to the toilet (Label 16). This is thought to reflect the characteristics of the real space as we understand it, and the characteristics of the interior space can be expressed in a probabilistic manner. One case (0.005%) of toilet (Label 16) occurs next to bed (Label 3), but this reflects the adjacency of objects along the opposite wall in the scene, because the adjacency was extracted from the recognized center-of-gravity distance.

Similar trends are seen for presence and adjacency probabilities by scene type. In addition, for example, in the bathroom scene, the number of shower curtains, toilets, sinks, and bathtubs was high, and the mutual adjacency between shower curtains, toilets, and bathtubs, which is seen in unit baths, as well as between counters and sinks around the water area, was also high. The scene also shows many mutual adjacencies between shower curtains and toilets and bathtubs, which are seen in unit bathrooms, and between counters and sinks, which are similarly seen around bathrooms.

IV. OBJECT ESTIMATION METHOD USING PRIOR KNOWLEDGE BASED ON PROBABILISTIC FIELD REPRESENTATION

This section presents examples of the use of the probability fields for all 1201 scenes in object estimation based on prior knowledge.

A. Existing Methods

We used the SparseConvNet [12] model as an existing method for object identification using 3D point cloud data. The SparseConvNet model identifies objects using deep learning, and it has one of the top accuracy rankings for semantic segmentation tasks on the ScanNet dataset [13]. Semantic segmentation for 3D point cloud data is the task of labeling each point with an object category. The probability of each object category is calculated, and in general, the object with the largest probability is labeled as the object identification result.

Table IV shows the identification accuracy of SparseConvNet using the ScanNet dataset. The identification accuracy of semantic segmentation uses a metric called Intersection over Union (IoU), which indicates the percentage of regions where the object category was correctly recognized. While the existing method can identify objects with a correct identification rate of over 80% for several object categories, some object categories have a predicted probability in the 20% range.

Object identification methods based on deep learning, such as SparseConvNet, are implemented by automatically learning features within data from a large amount of training data. However, object identification with these methods ignores real-space knowledge that we humans have, such as the fact that a kitchen and a sofa are unlikely to be adjacent to each other.

B. Segmentation Prediction Considering Prior Knowledge

We used the obtained probabilistic field representation to perform object estimation for semantic segmentation predictions. From the semantic segmentation prediction, we used Bayesian estimation [14] to calculate the posterior probabilities, including prior knowledge.

We focused on a certain object A in the point cloud data. Let $p_A(x)$ be the predicted probability of the object category $x \in X$ for object A by semantic segmentation prediction. Then, if $Z_{1:k} = Z_1, Z_2, \cdots, Z_k$ represents an object's neighboring object A, the posterior probability of object A under observation $Z_{1:k}$, $P(x|z_{1:k})$ can be estimated by Bayesian estimation:

$$
\begin{aligned}
P(x|z_{1:k}) &= \frac{G(z_k|x) \times P(x|z_{1:k-1})}{\sum_{x' \in X} G(z_k|x') \times P(x'|z_{1:k-1})}, \\
G(z_k|x) &= p_{Z_k}(z_k) \times g(z_k|x) \\
&\quad + [1 - p_{Z_k}(z_k)] \times [1 - g(z_k|x)], \\
P(x|z_{1:0}) &= p_A(x).
\end{aligned}
$$

As prior knowledge, let $g(b|a)$ be the probability that category a is adjacent to category b in one room scene type.

C. Evaluation

We calculated performance indicators before and after Bayesian estimation was applied.

Accuracy, precision, recall, and F1-score are shown for the performance measures of multiclass classification. Accuracy is the percentage of all data that were correct, precision is the

TABLE III: Adjacency probability for all scenes (%)

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	71.9	0.1	3.4	0.3	1.2	1.4	0.8	0.4	0.5	7.7	3.2	0.3	1.6	0.0	0.8	3.9	0.1	2.7
3	0.1	95.3	1.0	0.0	0.2	0.2	0.4	0.2	0.3	0.0	1.3	0.4	0.0	0.0	0.0	0.0	0.0	0.7
4	1.1	0.3	48.1	0.2	38.9	0.5	0.9	0.2	0.1	0.0	8.0	0.3	0.1	0.0	0.0	0.0	0.0	1.5
5	0.7	0.1	1.3	80.5	11.4	0.2	0.6	0.0	0.4	1.4	0.5	0.3	0.0	0.0	0.0	0.0	0.0	2.5
6	0.2	0.0	17.7	0.7	79.8	0.2	0.3	0.0	0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.7
7	2.1	0.2	2.1	0.1	1.6	81.9	1.1	1.1	0.6	0.5	1.1	0.6	0.5	0.6	0.8	0.7	0.3	4.1
8	1.5	0.7	5.0	0.4	4.2	1.4	75.2	0.5	0.5	0.9	4.0	1.7	0.1	0.0	0.1	0.5	0.0	3.3
9	0.9	0.3	1.6	0.0	0.1	1.6	0.6	92.6	0.2	0.0	0.7	0.1	0.0	0.0	0.0	0.0	0.0	1.2
10	3.1	1.6	1.7	0.9	4.2	2.6	1.8	0.6	74.9	0.5	3.2	0.3	0.2	0.6	0.4	0.5	0.0	2.8
11	27.1	0.0	0.2	2.0	0.3	1.1	1.6	0.0	0.3	50.0	0.0	0.2	3.4	0.0	0.1	10.8	0.0	2.9
12	3.1	1.2	23.2	0.2	1.2	0.7	2.1	0.3	0.5	0.0	64.6	0.3	0.1	0.0	0.0	0.0	0.0	2.6
13	3.8	4.3	9.5	1.6	6.6	4.7	11.8	0.7	0.6	0.6	4.1	45.4	0.5	0.0	0.1	0.5	0.0	5.0
14	11.7	0.0	1.2	0.1	0.3	2.6	0.3	0.0	0.3	7.3	0.7	0.3	67.8	0.0	0.0	2.7	0.0	4.7
15	1.9	0.0	0.0	0.0	1.3	17.7	0.2	0.0	4.2	0.0	0.0	0.0	0.0	23.0	19.0	2.7	25.0	5.0
16	15.1	0.1	0.1	0.0	0.6	10.7	0.7	0.0	1.3	0.4	0.0	0.2	0.0	8.3	25.1	8.4	11.0	18.1
17	30.4	0.0	0.3	0.0	1.3	3.4	2.1	0.0	0.7	24.3	0.0	0.3	2.8	0.5	3.3	22.0	0.8	7.8
18	2.4	0.0	0.0	0.0	1.1	6.9	0.4	0.0	0.0	0.0	0.0	0.0	0.0	17.6	17.8	3.1	45.9	4.7
19	3.8	1.0	6.3	1.4	6.1	3.8	2.5	0.7	0.6	1.2	3.7	0.6	0.9	0.2	1.3	1.4	0.2	64.1

TABLE IV: Result of Semantic Segmentation on SparseConvNet

Label	Category	IoU	Label	Category	IoU
0	wall	0.7951	10	picture	0.2436
1	floor	0.9509	11	counter	0.5920
2	cabinet	0.6021	12	desk	0.5850
3	bed	0.7874	13	curtain	0.6768
4	chair	0.8907	14	refrigerator	0.4372
5	sofa	0.8249	15	shower curtain	0.6142
6	table	0.7151	16	toilet	0.9161
7	door	0.5354	17	sink	0.6636
8	window	0.5052	18	bathtub	0.8807
9	bookshelf	0.7651	19	other furniture	0.4921
				avg. IoU	0.6737

TABLE V: Evaluation Results

(a) Comparison of each indicator

	Accuracy	Precision	Recall	F1-score
predicted label	65.3%	60.1%	59.9%	60.0%
posterior label	67.5%	70.8%	62.4%	66.3%

(b) No. of changes for each label

	posterior label = annotated	posterior label ≠ annotated
predicted label = annotated	6678 (88.7%)	850 (11.3%)
predicted label ≠ annotated	1099 (27.5%)	2896 (72.5%)

percentage of those predicted as class L_i that were correct, recall is the percentage of class L_i data that were correctly predicted, and F1-score is the harmonic mean of precision and recall.

For all objects in the 312-scene validation data, we validated the correlation between predicted and posterior probabilities for semantic segmentation. We show the performance indices in Table V before and after applying Bayesian estimation. Here, "predicted label" represents the label before applying Bayesian estimation, i.e., the label with the highest predicted probability among $p_A(x)$ and "posterior label" represents the label after applying Bayesian estimation. Precision, recall, and F1-score are the average of the indicators for each class.

Table Va shows that precision was improved by about 10%, which is a concrete beneficial effect. Recall, however, only improved by about 3%. A high rate of improvement in the precision rate is thought to indicate that the number of correctly compensated objects tends to be large. Conversely, the low improvement rate of recall suggests that there a certain number of objects in a particular class L_a have been assigned to another class. There are many classes for which precision increased significantly but recall is still low. Table Vb shows the distribution of posterior labels depending on whether the posterior label is consistent with the predicted label or not and on whether they are true label or not. Table Vb shows that, by the Bayesian estimation, 27.5% of predicted labels, which differ from annotated labels, can be modified to the annotated labels. This is a factor for improving the indicators in Table Va. However, 11.3% of predicted labels was mistakenly modified by the Bayesian estimation, which is why the improvement in Table Va was marginal. Although this suggests that Bayesian estimation can be applied to correct certain object predictions, some of object predictions cannot be corrected by Bayesian estimation.

To see when and how such the mistake happens, in Figure 4, we show histograms of the cases where the predicted labels changed dependent on the highest predicted probability. In the figure, the histogram with magenta shows the case when the wrong label has been properly corrected, and the histogram with green shows the case when the correct label has been mistakenly modified to a wrong label. Figure 4 shows that the patterns that led to correct labels are concentrated around 0.5 to 0.6 of the prediction probability. The patterns that led to wrong labels are concentrated in 0.9 or more higher prediction probabilities before Bayesian estimation. These results show that our Bayesian estimation should not be applied to objects taking a high prediction probability by the machine learning techniques.

V. CONCLUDING REMARKS

In this study, we worked on the representation of real-space information as a probabilistic field. We obtained adjacency
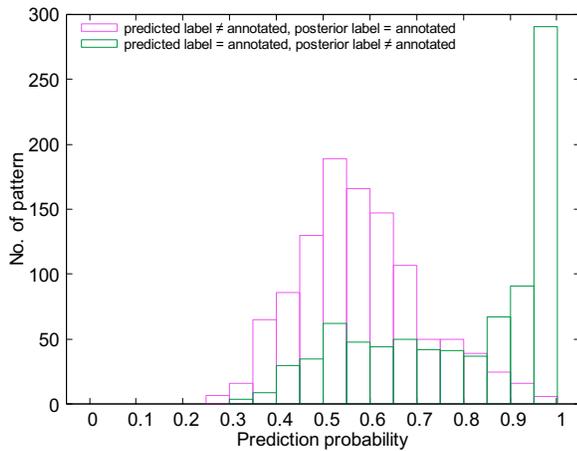
Fig. 4: Prediction probability distribution for patterns with changing prediction labels

information for the positions of objects in real space from an indoor 3D point cloud dataset, and obtained a real-space probabilistic field by statistically determining what kind of correlation exists between object adjacencies in real space. Next, we presented an object estimation method based on prior knowledge as an example of using the obtained probabilistic field for The evaluation results confirmed that the object estimation method based on prior knowledge improved the estimation accuracy to a certain degree in the accuracy, precision, recall, and F1-score metrics. It was also confirmed that object estimation based on prior knowledge is effective when segmentation prediction by existing methods is not sufficiently accurate.

Our future work is to improve our object estimation method by acquiring probabilistic fields of real-space information from multiple datasets such as the dataset of outdoor fields.

ACKNOWLEDGMENT

REFERENCES

[1] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[2] T. de Vries, I. Misra, C. Wang, and L. van der Maaten, "Does object recognition work for everyone?" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2019, pp. 52–59.

[3] L. Wang, X. Fan, J. Chen, J. Cheng, J. Tan, and X. Ma, "3D object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities," *Sustainable Cities and Society*, vol. 54, p. 102002, Mar. 2020.

[4] S. Takagi, J. Kaneda, S. Arakawa, and M. Murata, "An improvement of service qualities by edge computing in network-oriented mixed reality application," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, Apr. 2019, pp. 773–778.

[5] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud," *IEEE Access*, vol. 7, pp. 179 118–179 133, Dec. 2019.

[6] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International Journal of Computer Vision*, vol. 128, pp. 261–318, Oct. 2019.

[7] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, vol. 70, pp. 245–317, 2021.

[8] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Aug. 1996, pp. 226–231.

[9] "Open3D - A Modern Library for 3D Data Processing," http://www. open3d.org/, (Accessed: 2020-11-18).

[10] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, Feb. 2009, pp. 331–340.

[11] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2432–2443.

[12] B. Graham, M. Engelcke, and L. van der Maaten, "3D semantic segmentation with submanifold sparse convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018, pp. 9224–9232.

[13] "ScanNet Benchmark," http://kaldir.vc.in.tum.de/scannet_benchmark/, (Accessed: 2020-10-05).

[14] B. Ristic, C. Gilliam, M. Byrne, and A. Benavoli, "A tutorial on uncertainty modeling for machine reasoning," *Information Fusion*, vol. 55, pp. 30–44, Mar. 2020.