Accuracy based Rewarding for Sensors in Noisy Collaborative Point Cloud Acquisition Environments

Sumiko Miyata
School of Engineering
Institute of Science Tokyo
Tokyo, Japan
sumiko@ict.eng.isct.ac.jp

Takamichi Miyata
Faculty of Advanced Engineering
Chiba Institute of Technology
Chiba, Japan
takamichi.miyata@it-chiba.ac.jp

***Abstract*—A common challenge in point-cloud acquisition environments, such as realizing traffic safety and autonomous driving, is to determine the placement of sensors and workers and the rewards to be paid to them. Game theory is available as an extremely powerful tool for the problem of determining appropriate rewards for deployed sensors and workers and several previous studies have proposed methods for determining rewards using game theory. However, these methods do not consider the affection of the recognition accuracy of downstream tasks by the AI to develop these appropriate rewards. In this paper, we propose a novel characteristic function of game theory by considering the recognition accuracy of AI. To define our function, we investigate how the number of viewpoints and the noise level of the point cloud affect the classification accuracy. In addition, we analyze which part of the point cloud the recognition model focuses on by using SHAP, a method to improve the explainability of machine learning based on the Shapley value.**

***Index Terms*—Point cloud, Characteristic function, Shapley value, SHAP, Zero-shot point cloud recognition model**

I. Introduction

Cooperative point-cloud acquisition environments over a wide area by multiple LiDAR and other sensors is an important component of LiDAR-based intersection monitoring [1], [2] , sensor fusion in autonomous driving [3], etc. A common challenge in point cloud capture environments is determining the placement of sensors and workers and the rewards to be paid to them [4], [5].

When solving the problem of sensor and worker placement, the accuracy of the point cloud data for downstream tasks, such as object recognition, detection, and segmentation, plays an important role [2]. Moreover, we need to pay attention that point cloud acquired by sensors incur unavoidable noise during the acquisition process. In the field of image recognition, an image classification network obtained by learning is strongly affected by noise present in the input image [6]. This research sheds light on the relationship between noise in point cloud capture and recognition accuracy, which has not been given much attention so far.

Game theory is available as an extremely powerful tool for the problem of determining appropriate rewards for deployed sensors and workers. While previous studies have presented utility functions (*characteristic functions*) for game theory by considering the number of point clouds [5] or network bandwidth [7], this paper proposes a characteristic function that is determined by the recognition accuracy of downstream tasks by the AI.

The major contributions of this paper are as follows

- We investigated how the number of viewpoints and the noise level of the point cloud affect class classification accuracy using a zero-shot point cloud recognition model that can be applied to the recognition of various classes of objects.
- We defined an appropriate characteristic function based on the effect of sensors coalition on accuracy, and quantified imputation to each sensor using Shapley value. This is useful for setting appropriate rewards based on contributions to sensor installers and cloud workers.
- To qualitatively investigate the effect of point cloud quality and number of viewpoints on recognition accuracy, we analyzed which part of the point cloud the recognition model is focusing on using SHAP, a method for improving the explainability of machine learning based on Shapley value.

This paper is organized as follows. Section 2 describes related work and preliminaries. Section 3 propose our model using coalitional game theory including of our design of characteristic function. Section 4 shows the the effect of point cloud quality and number of viewpoints on recognition accuracy. Section 5 concludes our paper.

II. Related Work And Preliminaries

Cooperative Game Theory. Let $N = \{1, 2, ..., n\}$ be the set of n *players* and $S \subseteq N$ be a *coalition*. The *characteristic function* $v : 2^N \to \mathbb{R}$ is defined to represent the value of each coalition S. For the empty set $\emptyset$, we define $v(\emptyset) = 0$. The characteristic function v is *superadditive* if,

$$S \cap T = \emptyset \Rightarrow v(S \cup T) \geq v(S) + v(T) \ \ \forall S, T \subseteq N. \tag{1}$$

The *payoff vector* $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is called an *imputation* if it satisfies the following two properties: *collective rationality* ($\sum_{i \in N} x_i = v(N)$) and *individual rationality* ($x_i \geq v(\{i\})$ for all $i \in N$).

Shapley value is a solution in cooperative game (N, v) that assigns a value to each player in a game and it defined by the

average marginal contribution of a player across all possible coalitions [8]. The Shapley value of player i is defined as follows:

$$\phi(v)_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \big(v(S \cup \{i\}) - v(S)\big) \quad (2)$$

The Shapley value is imputation if the characteristic function v satisfies the individual rationality.

Characteristic Function of Point Cloud Data. In order to derive incentives for obtaining point clouds, we have used cooperative game theory in our analysis [4], [5]. In this approach, the characteristic function is expressed in terms of the number of points that could be acquired refereed to [7]. In particular, by setting the threshold at the number of point clouds that must be satisfied as a system, point cloud overlapping by multiple players is considered.

Hotoyama *et al.,* [2] solve the bandwidth allocation scheme for transmitting point cloud data at intersections using game theory. However, the gain function in non-cooperative game theory, which has similar meaning to the characteristic function, only takes into account the number of points and distortion in the frame.

The characteristic functions of these methods fail to take into account object recognition, which is necessary when considering services. In this paper, we propose a characteristic function derived from the accuracy of class classification, one of the high-level vision tasks.

High-level Vision Tasks for Point Clouds. Advances in deep learning have enabled the development of various high-level vision tasks for point clouds, including point cloud classification [9], [10], object detection [11], [12], and segmentation [13], [14].

In recent years, the zero-shot approach, which means that the model can perform some tasks without task-specific training, has been applied to point cloud classification and segmentation [15], [16]. The advantage of the zero-shot approach is that it can classify or segment objects with very diverse classes without retraining the model. In particular, PointCLIP-V2 [16] achieved state-of-the-art performance in zero-shot point cloud classification by projecting point clouds into images from multiple viewpoints and feeding these images into CLIP (Contrastive Language-Image Pre-training) [17], which is a widely and successfully used for various tasks such as object detection [18], image editing [19], and image quality assessment [20], [21].

SHAP. Lundberg *et al.,* proposed SHAP [22] (SHapley Additive exPlanations) as a method to explain the output of trained models. The definition of SHAP are based on Shapley values in cooperative game theory and calculated based on the average marginal contribution of each feature across all possible coalitions of features. The gist of SHAP is assuming feature additivity that the output of a model can be explained by the sum of the contributions of each feature. SHAP are widely used in various fields, including computer vision, natural language processing, and healthcare, to interpret the output of deep learning models.

In this paper, we use SHAP to interpret the classification process of PointCLIP-V2 and investigate which part of the point cloud (more specifically, which part of the projected image) the model focuses on when classifying objects.

III. Method

We assume that multiple sensors, such as LiDAR, are placed around the point cloud of an object. In this case, the point cloud data acquired by each sensor represents one side of the object. When using a rendering-based point cloud classification model such as PointCLIP-V2, it is best to consider that the information from a single sensor is a single image rendered from the viewpoint of the sensor. By feeding these images into the pre-trained CLIP, we can estimate which class the point cloud belongs to. The Fig. 1 shows an overview of the three experiments we performed in this paper. The details of each experiment are described below.

Effect of Number of Viewpoints and Noise on Classification Accuracy. To investigate the effect of multiple sensors placement on the classification accuracy when noise occurs in the point cloud acquisition process, we add i.i.d. Gaussian noise $\mathcal{N}(0, \sigma)$ to the coordinates of the all points in the original (clean) point cloud data. When k sensors observe one object's point cloud with noise of standard deviation σ, integraging the point cloud data from each sensor results in the noise standard deviation of $\sigma/\sqrt{k}$. On the other hand, when k sensors observe the object's point cloud from different viewpoints, the noise standard deviation is still σ, but obtaining the images from different viewpoints can improve the classification accuracy by providing more information about the target object. Thus, there is a trade-off between the number of viewpoints and the noise level in terms of classification accuracy.

Note that in this study, we assume that the correspondence between points in the point cloud data acquired by different sensors with the same viewpoint is known, and that the points obtained from different viewpoints do not overlap with each othre, for simplicity.

Design of Characteristic Function. Measuring the contribution of each sensor to the classification accuracy requires designing a characteristic function v. The most straightforward way to design the characteristic function is to define it equals to classification accuracy. Since this naive characteristic function is not always superadditive, there i (as shown in our experimental results in the following section)s no guarantee that it satisfies individual rationality. Which means that fair imputation does not always exist.

To address this issue, we design our characteristic function $v(S)$ for a coalition S as,

$$v(S) = a(S) - \frac{1}{|S|} \sum_{j \in S} a(j) \quad (3)$$

where, $a(S)$ is the classification accuracy obtained by integrating the information from all sensors in coalition S. This characteristic function was inspired by the characteristic function of the cab game [23]. Note that in our characteristic function, the utility of each sensor when participating in the game individually is 0 ($|S| = 1 \rightarrow v(S) = 0$), and as long as the payoff is non-negative, individual rationality is always

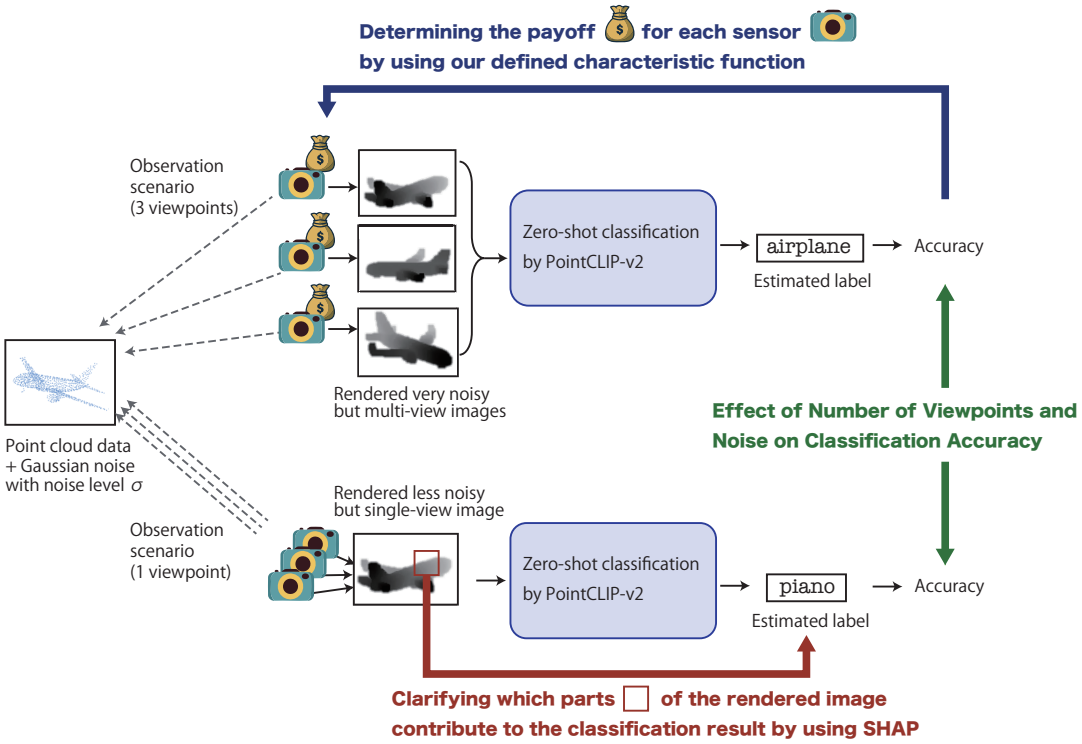


Fig. 1: Overview of the three experiments performed in this paper.

guaranteed. We use Shapley value to set the appropriate reward for each sensor based on this characteristic function. As we mentioned earlier, our characteristic function does not satisfy superadditivity, but the good properties of Shapley value are retained even in such cases [24].

Using SHAP for Interpretation of Classification Process. The model obtained by machine laerning is often a black box, making it difficult to explain its output. Our main interest in this paper is to investigate the effect of sensor placement on point cloud classification under noisy conditions and to set the appropriate reward for each sensor. However, it is difficult to understand how the placement of sensors affects the accuracy by discussing only the final clasiffication accuracy.

Therefore, we use SHAP, which is widely used to explain the behavior of trained models, to interpret the point cloud classification process. By focusing on characteristic cases of classification results, we quantitatively investigate which part of the image obtained by rendering the point cloud contributes to the classification result, and from this contribution, we expect to understand how the images seen from each sensor change under noisy conditions and how they affect the classification accuracy.

IV. Experimental Results

A. Common Experimental Settings

Observation scenarios In all experiments, we conider the three LiDAR sensors ($N = \{1, 2, 3\}$) observing point cloud of an object. Each sensor is placed by selecting one from multiple viewpoint candidates shown in Fig. 2. We consider two scenarios: **P3V3**: each sensor observes the object from a different viewpoint, and **P3V1**: all sensors observe the object from the same viewpoint (viewpoint 1 in Fig. 2).

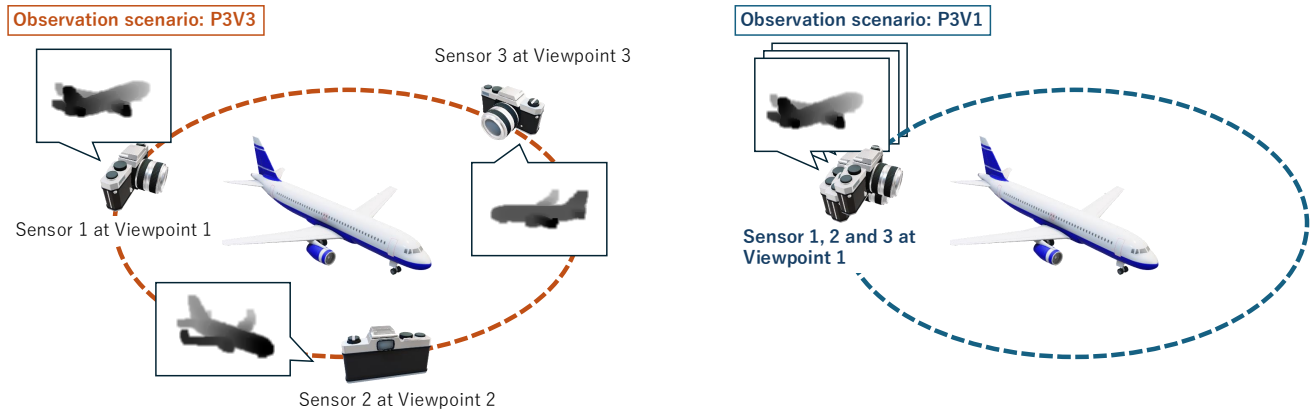
Evaluation We use the official implementation of PointCLIP-V2 to investigate the effect of choosing multiple sensor viewpoints on the accuracy of point cloud classification. In the original PointCLIP-V2 model, there are 10 viewpoints for each object. Thus, we choose 3 viewpoints (original viewpoint ID 0, 7 and 1) from them for our experiments as viewpoint 1, 2 and 3, respectively (Fig. 2).

For the point cloud dataset, we use the ModelNet40 test set, which contains 2,468 objects in 40 classes. We use the classification accuracy (%) as the evaluation metric. Since ModelNet40 includes 40 classes, the chance level accuracy is 2.5%.

B. Effect of Number of Viewpoints and Noise on Classification Accuracy

The classification accuracy for each observation scenario and noise level σ is shown in Fig. 3. The noise level σ is varied from 0 to 0.04 in steps of 0.01. The classification accuracy for each observation scenario and noise level σ is shown in Fig. 3.

The results show that 1) the accuracy decreases as the noise level increases, 2) P3V3 is more accurate than P3V1 when the noise level is low, and the opposite is true when the noise level is high. In this case, the accuracies of both scenarios are reversed at about $\sigma = 0.02$.



(a) Observation scenario P3V3

(b) Observation scenario P3V1

Fig. 2: Viewpoints and sensors' position in our experimental setting for each observation scenario.

TABLE I: Classification accuracy $a(S)$ (%) for each combination of coalition S, observation scenario, and noise levels $\sigma = 0.01$ and 0.03.

σ	scenario	classification accuracy a						
		$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{2,3\}$	$\{1,3\}$	$\{1,2,3\}$
0.01	P3V3	40.32	37.72	41.73	49.23	50.28	45.18	**52.47**
	P3V1		40.32			43.72		44.49
0.03	P3V3	17.30	22.08	21.35	24.35	25.73	19.65	25.49
	P3V1		17.30			25.04		**29.90**

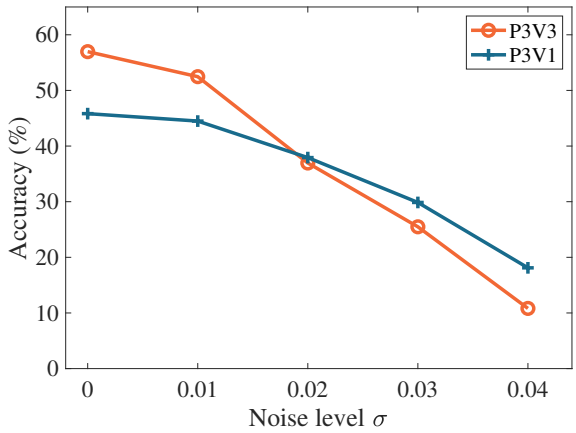


Fig. 3: Classification accuracy for each number of viewpoints and noise level σ.

Based on these results, we focused on two noise levels, $\sigma = 0.01$ and 0.03. TABLE I shows the class classification accuracy for each coalition in each scenario. The table shows that 1) classification accuracy generally improves with increasing coalition size, but there is exception (i.e., monotonicity is not satisfied), and 2) classification accuracy does not exhibit superior additivity in Eq. (1), while all cases have inferior additivity.

C. Imputation by Shapley Value

The Shapley values for each sensor, computed using our proposed characteristic function, are shown in Table II. From this table, we can see that:

TABLE II: Shapley value for each sensor with different noise levels ($\sigma = 0.01$ and 0.03) and number of viewpoints.

σ	scenario	$\phi(v)_1$	$\phi(v)_2$	$\phi(v)_3$	$v(\{1,2,3\})$
0.01	P3V3	**3.06**	**6.26**	**3.23**	**12.55**
	P3V1	1.39	1.39	1.39	4.17
0.03	P3V3	1.24	3.08	0.92	5.24
	P3V1	**4.20**	**4.20**	**4.20**	**12.60**

- Sensor 2 has the highest Shapley value of $\sigma = 0.01$ and P3V3. In fact, sensor 2 is the least accurate on its own, but when combined with other sensors, the accuracy improves significantly. Thus, we can understand that the rewards from the partnership are very large.
- For all noise levels, the reward received by each sensor tends to be greater when the scenario with the highest overall coalition utility is chosen. This is a good result in the sense that it allows for the natural selection of scenarios with overall benefits.

D. Interpretation of Classification Process of PointNet-v2 by using SHAP

By using SHAP, we investigated how sensor 2, observing from viewpoint2, contributes to the results in the P3V3 scenario with $\sigma = 0.01$. Based on a point cloud of guitars selected from ModelNet40, we investigated from what perspective each sensor 1-3. Fig. 4 shows the viewpoints from which sensors 1 to 3 viewed the point cloud, which class was estimated to have the highest probability, and where in the image the results were

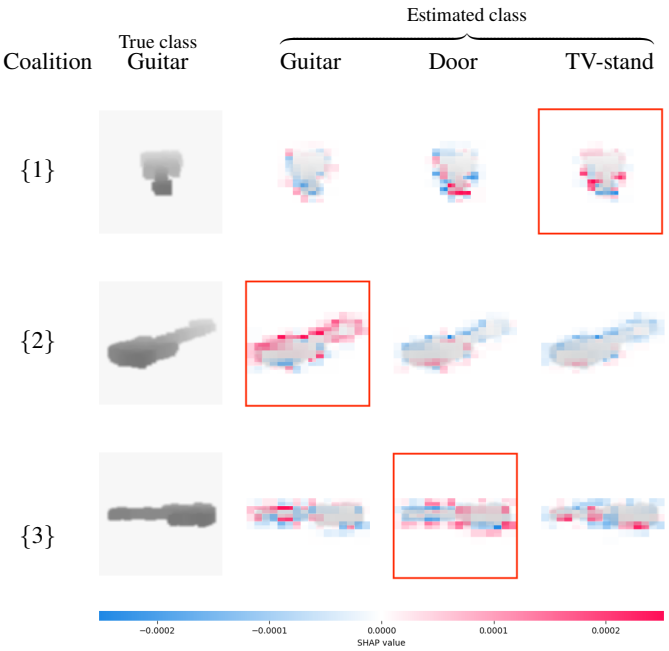


Fig. 4: SHAP values for the classification result for the guitar point cloud from three viewpoints ($\sigma = 0.01$ and P3V3 scenario).

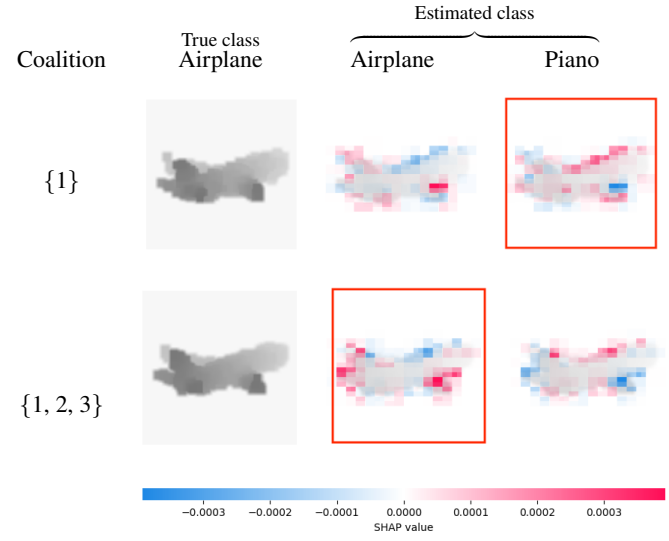


Fig. 5: SHAP values for the classification result for the airplane point cloud from one viewpoint and different noise levels, high (top:$\sigma = 0.03$ when S=$\{1\}$) and low (bottom:$\sigma = \frac{0.03}{\sqrt{3}}$ when S=$\{1, 2, 3\}$).

selected. The leftmost column shows the images visible from each sensor, and the three columns to its right show the SHAP values corresponding to the three classes. Areas colored red in these images represent positive contributions to being in that class. (Blue means the opposite). The red boxes in the images represent the most probable classes.

The results show the following:

- Sensors 1 and 3 misclassify the image to a class other than guitar. The angle of sensor 1 makes it clear from the image that it looks like a TV-stand. Especially, we can see more clearly from SHAP that such a judgment is made by looking at the nape of the center portion of the image.
- Sensor 2 correctly identifies the object as a guitar. From this angle, the characteristic shape and neck of the guitar are visible, and the SHAP value is positive and large at the boundary between the object and the background.

Then, an observation of the contribution of the resulting point cloud to the noise reduction by increasing the number of sensors in the P3V3 scenario with $\sigma = 0.03$ is shown in Fig. 5. True class object is Airplane. The obtained characteristics are as follows:

- The upper part of the image is the result when there is only one sensor such as $S = \{1\}$. This airplane model is unnaturally raised due to the large noise level, and the characteristic shapes such as wings and tail are also somewhat ambiguous. PointCLIP is unable to correctly classify the class from these factors and PointCLIP failed to correctly classify the class and misclassified “ Piano ” as the most probable.
- On the other hand, the image in the lower row is the result when overall coalition, and since the noise level is kept low ($\sigma = \frac{0.03}{\sqrt{3}}$), the characteristic shape of the aircraft is more clear, and the SHAP value also indicates that the correct decision is made on this basis.

Thus, when the noise level is high, the results show that more sensors looking at the same viewpoint would be more effective.

V. Conclusion

We experimentally demonstrated that the sensor placement that maximizes the classification accuracy by AI in a collaborative point cloud acquisition environment varies depending on the noise level of the acquisition. By considering this problem as a cooperative game, we also proposed a new characteristic function that focuses on the point cloud classification accuracy. Calculating the Shapley value based on the proposed characteristic function allows us to determine the appropriate reward for sensors and workers. In the experimental results in this paper, it was shown that selecting the coalition that maximizes the reward for each sensor from multiple observation scenarios is equivalent to selecting the scenario with the highest accuracy at the same noise level. Moreover, by using SHAP, we visualize where the point cloud classification method is based on in various combinations of noise levels and viewpoints in the classification process. This result can be used as a guide when considering more complex observation scenarios.

Our future work includes exploring more appropriate characteristic functions and applying our method to other high-level vision tasks such as object detection and segmentation.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers, 22K12015 and 23K03871.

References

[1] J. R. Palit and O. A. Osman, "Application of lidar data for deep learning based near crash prediction at signalized intersection," *Journal of Transportation Technologies*, vol. 13, no. 2, 2023.

[2] H. Hotoyama, S. Miyata, and R. Shinkuma, "Point cloud data reduction method considering bandwidth limitation at an intersection," in *Proc. of KJCCS2024*, 2024.

[3] M. Nawaz, J. K.-T. Tang, K. Bibi, S. Xiao, H.-P. Ho, and W. Yuan, "Robust Cognitive Capability in Autonomous Driving Using Sensor Fusion Techniques: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 3228–3243, 2024.

[4] J. Watanabe, S. Miyata, K. Kanai, N. Kamiyama, T. Yamazaki, and E. Kamioka, "Reward Distribution Using Coalitional Game Considering Overlapping of Point Cloud Data," in *Proc. of INCoS2024*, 2024.

[5] J. Watanabe, S. Miyata, and K. Kanai, "Reward Distribution Using an Anti-Duality Game in a Point Cloud Data Trading Model Based on Collected Area," in *Proc. of International Conference on Artificial Intelligence in Information and Communication*, 2024, pp. 552–557.

[6] K. Mamiya and T. Miyata, "Few-class learning for image-classification-aware denoising," in *Proc. of IEEE International Conference on Image Processing*, 2020, pp. 948–952.

[7] S. Miyata, "Cost Sharing Method for a Mobile Tethering with a Coalitional Game Theory," in *Proc. of International Conference on Artificial Intelligence in Information and Communication*, 2024, pp. 291–296.

[8] L. S. Shapley, "A value for n-person games," pp. 307–317, 1953.

[9] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 77–85.

[10] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9621–9630.

[11] C. Xu, B. Wu, Z. Wang, W. Zhan, P. Vajda, K. Keutzer, and M. Tomizuka, "SqueezeSegV3: Spatially-adaptive convolution for efficient point-cloud segmentation," in *Proc. of European Conference on Computer Vision*, 2020, pp. 1–19.

[12] Q. Meng, W. Wang, T. Zhou, J. Shen, Y. Jia, and L. Van Gool, "Towards a weakly supervised framework for 3D point cloud object detection and annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[13] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[14] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions On Graphics*, vol. 38, no. 5, pp. 1–12, 2019.

[15] R. Zhang, Z. Guo, W. Zhang, K. Li, X. Miao, B. Cui, Y. Qiao, P. Gao, and H. Li, "PointCLIP: Point cloud understanding by CLIP," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8542–8552.

[16] X. Zhu, R. Zhang, B. He, Z. Guo, Z. Zeng, Z. Qin, S. Zhang, and P. Gao, "PointCLIP V2: Prompting CLIP and GPT for powerful 3D open-world learning," in *Proc. of IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2639–2650.

[17] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the International Conference on Machine Learning*, vol. 139, 2021, pp. 8748–8763.

[18] X. Gu, T.-Y. Lin, W. Kuo, and Y. Cui, "Open-vocabulary object detection via vision and language knowledge distillation," in *Proc. of the International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=lL3lnMbR4WU

[19] O. Patashnik, Z. Wu, E. Shechtman, D. Cohen-Or, and D. Lischinski, "StyleCLIP: Text-driven manipulation of StyleGAN imagery," in *Proc. of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2065–2074.

[20] J. Wang, K. C. Chan, and C. C. Loy, "Exploring CLIP for assessing the look and feel of images," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[21] T. Miyata, "ZEN-IQA: Zero-shot explainable and no-reference image quality assessment with vision language model," *IEEE Access*, vol. 12, pp. 70 973–70 983, 2024.

[22] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. of International Conference on Neural Information Processing Systems*, 2017, p. 4768–4777.

[23] K. Shin, "Introduction to cooperative game theory (in Japanese)," *Operations Research*, vol. 6, no. 37, pp. 343–350, 2015.

[24] A. Honda, E. Takahagi, Y. Narukawa, and K. Fujimoto, "Non additive set functions and their applications—iii — measurement and control for the degree of freedom in a fuzzy measure," *Systems, Control and Information*, vol. 65, no. 6, pp. 232–238, 06 2021. [Online]. Available: https://cir.nii.ac.jp/crid/1390290393639906176