2024 IEEE 27th International Conference on
Intelligent Transportation Systems (ITSC)
September 24- 27, 2024. Edmonton, Canada

Domain-Adaptive Point Cloud Semantic Segmentation from Urban to Off-road Scenes based on Knowledge-Augmented Deep Learning

Fengyu Xu[1], Yongxiong Xiao[1,*], Jilin Mei[2], Yu Hu[2,*] and Qiang Fu[1,*]

***Abstract*— Domain-adaptive point cloud semantic segmentation (PCSS) is crucial for high-level autonomous driving. However, supervised deep learning methods are often constrained by training data and suffer from poor generalization in unknown environments. To address these challenges, we propose a domain-adaptive PCSS approach leveraging knowledge-augmented deep learning (KADL). Specifically, we introduce three strategies: (1) point cloud data augmentation based on the compact bird's-eye view (CBEV) map, which is a novel point cloud organization method; (2) implicit knowledge augmentation based on knowledge distillation; (3) explicit knowledge augmentation based on attribution analysis and network modulation. For experimental validation, we utilize two distinct datasets, namely the urban dataset SemanticKITTI and the off-road dataset RELLIS-3D, which are involved in the training and testing phases, respectively. Additionally, we have added road labels to the RELLIS-3D dataset, which originally lacked a road category. To our knowledge, this work is the first to investigate domain-adaptive PCSS from urban to off-road scenes. The experimental results demonstrate that our method is effective and has promising performance. The code and data are available at https://github.com/xfy0032/kadlpcss**

***Index Terms*— autonomous driving, point cloud semantic segmentation, domain adaptation, knowledge-augmented deep learning**

1. INTRODUCTION

With the rapid development of deep learning, autonomous driving (AD) has made great progress, making assisted driving technology nearly standard in modern vehicles. Due to its precise perception capabilities for surrounding objects, 3D LiDAR has become one of the critical sensors for AD technology. Concurrently, 3D PCSS has emerged as a hot topic in intelligent vehicles, and PCSS methods based on deep learning have demonstrated remarkable performance across various evaluation datasets. However, most of these methods are trained and tested on the same dataset, leading to overfitting or poor generalization. In reality, traffic scenes are complex and diverse, making the excellent generalization of perception algorithms a fundamental prerequisite for vehicle safety.

To address these challenges, various domain-adaptive PCSS methods have been proposed. Domain adaptation aims to enhance the model's generalization, ensuring its high performance on the test dataset despite significant differences from the training datasets, such as datasets collected under different cities or weather conditions. Unlike existing works, we aim to train our model exclusively on urban scenes and apply it in off-road scenes. Therefore, we adopt the urban dataset SemanticKITTI [1] and the off-road dataset RELLIS-3D [2]. There is a pronounced distinction between these scenes, and correspondingly, the geometric characteristics of point clouds are also diverse as shown in Fig. 1. Nevertheless, humans can extract the common features of similar objects, such as obstacles or roads, from these diverse scenes and utilize them to understand unfamiliar scenes effectively. This motivates our research to explore the application of knowledge-augmented deep learning in domain-adaptive PCSS from urban to off-road scenes.

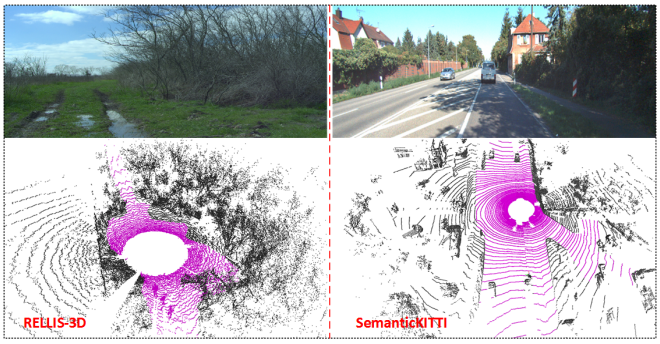

Fig. 1. Comparison of SemanticKITTI and RELLIS-3D. The top row shows the scene images, and the bottom row corresponds to the point clouds, where the colored points represent the drivable road areas.

Knowledge-augmented deep learning has been researched and applied in AD perception and decision-making to overcome the limitations of purely data-driven approaches [3]. From the perspective of knowledge embedding, KADL can be categorized into the following four levels: data level, architecture level, training level, and decision level [4]. In this work, we implement a domain-adaptive PCSS approach using knowledge augmentation. At the data level, we perform guided filtering on point cloud intensity data to make it consistent with the human common sense of continuity and smoothness; At the architecture level, we use a knowledge distillation framework to integrate some key features; At the decision level, we implement explicit knowledge embedding based on Shapley additive explanations (SHAP) [5] attribution analysis. Although it is difficult to mathematically encode human prior for neural networks in terms of the model parameters [6], this approach directly addresses the

This work was supported by the National Natural Science Foundation of China No. 52105128 and No. 62203424.

[1]Research Center for Novel Computing Sensing and Intelligent Processing, Zhejiang Lab, Hangzhou, China. xfy0032@foxmail.com, {xiaoyx, qfu}@zhejianglab.com

[2]The Research Center for Intelligent Computing Systems, Institute of Computing Technology, Beijing, China. {meijilin, huyu}@ict.ac.cn

*Corresponding authors

2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3315-0592-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ITSC58415.2024.10919619

black box problem of deep learning networks. Since the annotation categories of SemanticKITTI and RELLIS-3D are quite different, we add road labels to RELLIS-3D and set the categories of semantic segmentation into four categories, namely Road, Obstacle, Neutral area (its traversability lies between that of the roads and the obstacles.) and Others, among which Others fall into the ignorable category.

The main contributions of this work are listed as follows:

- We study the application of different levels of knowledge augmentation in PCSS networks, effectively and explicitly infusing domain knowledge into deep learning through SHAP attribution analysis.
- We have carried out a pioneering effort to research the domain adaptation PCSS task from urban to off-road scenes.
- We propose a novel point cloud organization method for data enhancement and provide road labels for the RELLIS-3D dataset to facilitate future research.

2. RELATED WORKS

2.1 Point Cloud Semantic Segmentation

3D PCSS is a fundamental task in the field of autonomous vehicles, aimed at assigning point-wise semantic labels to point clouds [8]. MV3D [9] takes the BEV map and range map as the model input and integrates image information to achieve the detection of specific objects. However, projection-based methods inevitably fail to fully utilize 3D information and currently meet the bottleneck of the segmentation accuracy [10]. Another mainstream PCSS methods are point-based methods, which employ multilayer perceptron (MLP) to directly process point clouds, such as PointNET [11] and RandLA-Net [12]. Obviously, point-based methods are generally time-consuming since the point clouds contain hundreds of thousands of measurement points. Researchers have also proposed voxel-based methods like SparseConv [13] to balance efficiency and accuracy. Recently, hybrid methods that combine the transformer with the above methods have also achieved promising results [14], [15]. However, these supervised algorithms generally have poor accuracy when applied to unfamiliar or unknown scenes, which reflects the limitations in terms of generalization.

2.2 Domain Adaptation

Domain adaptation has been widely used in vision tasks, while few studies apply them to point clouds [16]. 3DLP [17] is designed to tackle domain generalization challenges in PCSS, and it enhances generalization performance by leveraging the geometry and sequentiality of LiDAR data. PointDR [16] is proposed to achieve PCSS under different weather conditions through geometric style randomization and embedding aggregation. In addition, CosMix [18] is the first unsupervised domain adaptation method based on sample mixing, and it consists of a two-branch symmetric network that can process source data and target data concurrently. Since CoSMix employs unlabelled target data during the training phase, it is limited to completely unknown scenes. Moreover, the above methods are trained and tested on different urban datasets. As far as we know, no domain-adaptive PCSS method from urban to off-road datasets has been proposed in the literature.

2.3 Knowledge-Augmented Deep Learning

Although deep learning has achieved great success in the field of PCSS, data-driven supervised learning methods cannot handle long tail and hallucination problems well, and these methods also have challenges in interpretability [19]. Therefore, researchers have adopted different methods to extract and represent various human prior knowledge and integrate them into supervised learning models to achieve low data dependence, high generalization, and interpretable deep learning models. These approaches are called knowledge-augmented deep learning [4], which aim to realize deep learning driven by both data and knowledge. Similar concepts include knowledge-infused learning (KIL) [20], [21] and neural symbolic learning (NSL) [22]. Moreover, knowledge-augmented deep learning is increasingly applied to perception in the field of autonomous driving [3], [23]. Still, most existing methods belong to data-level knowledge augmentation, while architecture-level knowledge augmentation generally embeds domain knowledge into the deep learning model through explicit representation [24], [25]. To the best of our knowledge, applying architecture-level or decision-level knowledge augmentation to PCSS remains a relatively unexplored and scarce practice in the research community.

3. METHODOLOGY

3.1 Framework Overview

As shown in Fig. 2, our knowledge-augmented deep learning method mainly consists of the following three parts: (1) Point clouds data augmentation based on compact bird's-eye view; (2) Implicit knowledge augmentation based on knowledge distillation; (3) Explicit knowledge augmentation based on attribution analysis and network modulation.

For data augmentation, we adopt guided filtering based on CBEV to process the intensity data of point clouds instead of ignoring the intensity channel [17]. Implicit knowledge augmentation is designed to integrate artificial features into the network through knowledge distillation. This branch exclusively participates in the training phase and does not contribute to the inference phase. In contrast, the network modulation in our work belongs to an explicit knowledge augmentation that does not participate in the training of the neural network but intervenes in the inference phase. Different from modulating the weights of deep learning [26], the object modulated in our method is the intermediate feature vector. The following sections provide a detailed explanation of the above three components.

3.2 Data Augmentation

In vision tasks, data denoising has been established as a potent technique for enhancing model convergence stability and improving detection accuracy. Therefore, the point cloud

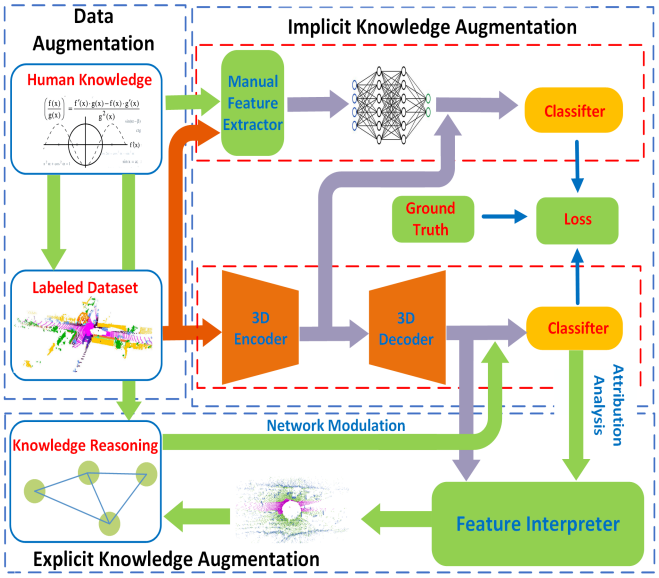

Fig. 2. Illustration of our knowledge-augmented deep learning architecture, which consists of the following three modules: (1) data augmentation; (2) implicit knowledge augmentation; and (3) explicit knowledge augmentation.

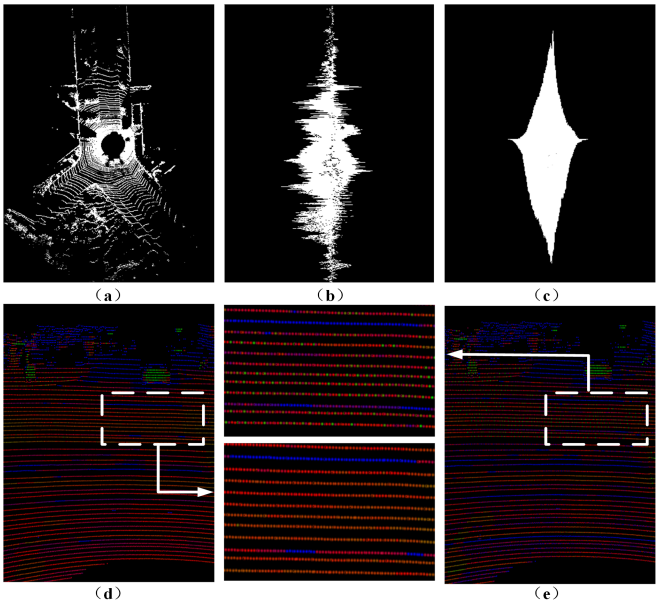

Fig. 3. Point clouds intensity filtering based on CBEV. (a) Original BEV map; (b) BEV map after lateral aggregation; (c) CBEV map; (d) filtered intensity data; (e) raw intensity data.

intensity is filtered to satisfy the smoothness constraint in this work. Currently, the methods for projecting point clouds into 2D images are BEV projection and range map [7] projection. The former projects point clouds to a 2D plane according to its real coordinates, and the latter projects point clouds to a cylinder plane. Although the BEV map is discrete, it reflects the true distribution state of the point cloud. Meanwhile, the range map is compact and has no overlap, but the neighborhood information on the range map suffers from longitudinal distortion. Given the above reasons, we propose the CBEV map, which can combine the advantages of both BEV maps and range maps at close distances. As shown in Fig. 3(b) and (c), the BEV map is compressed horizontally and then compressed longitudinally. Subsequently, guided filtering [27] is performed as illustrated in Fig. 3(d).

Another optimization is data normalization. For the filtered intensity,

$$i_{\text{new}} = \begin{cases} \dfrac{i}{I_{\text{th}}} & \text{if} \quad i < I_{\text{th}} \\ 0.999 & \text{else} \end{cases} \tag{1}$$

where I_{th} is a threshold that ensures the ratio of the number of points with intensity less than I_{th} to the total is p_I. Since the intensity of point clouds is significantly influenced by the material properties of targets, the normalization of intensity is to improve the adaptability to diverse scenes. Simultaneously, we also leverage normalization of the ground reference height to enhance generalization capabilities.

3.3 Implicit knowledge augmentation

In this work, the implicit knowledge augmentation is achieved through knowledge distillation. As illustrated in Fig. 4(a), the knowledge augmentation features F^{ka} used for knowledge distillation encompass manually designed features and the original attributes of the point cloud, namely:

$$F^{ka} = [x, y, z, i, h, d] \tag{2}$$

where (x, y, z) and i represent the original point cloud coordinates and normalized intensity; h and d are the height difference of the grid and the nearest obstacle distance, respectively, both of which are based on the BEV grid map.

The resolution of the grid map corresponds to a specific distance in the real environment, thereby determining the precision of the map. The grid map M established in our work has the size of 500 with the resolution r_{grid}, and M adopts the bottom left coordinate system, so the unmanned vehicle is located at $M_{(250,250)}$. After projecting the point cloud onto the grid map, $S_{(i,j)}$ represents the set of points falling into the grid $M_{(i,j)}$. The height difference h of a point is the difference between the maximum and minimum z values of the points in $S_{(i,j)}$. If the height difference exceeds the threshold H_{th}, the corresponding grid is marked as an obstacle grid. All obstacle grids constitute an obstacle map, and the nearest obstacle distance (NOD) d is obtained on the obstacle map using a breadth-first search algorithm, as illustrated in Fig. 4(c).

The design idea of our knowledge distillation module is inspired by the multi-scale fusion-to-single knowledge distillation (MSFSKD) [10], which is used to fuse image and point cloud and enhance the feature extraction capabilities of point clouds. As shown in Fig. 4(a), the knowledge augmentation features F^{ka} are dimensioned by a 3-layer encoder to generate intermediate features f^{ka}. The encoder is designed with a few layers to transfer as much human knowledge directly to subsequent modules as possible. Meanwhile, the point cloud's 3D features $f^{3d}_{scale_i}$ come from the 3D encoder

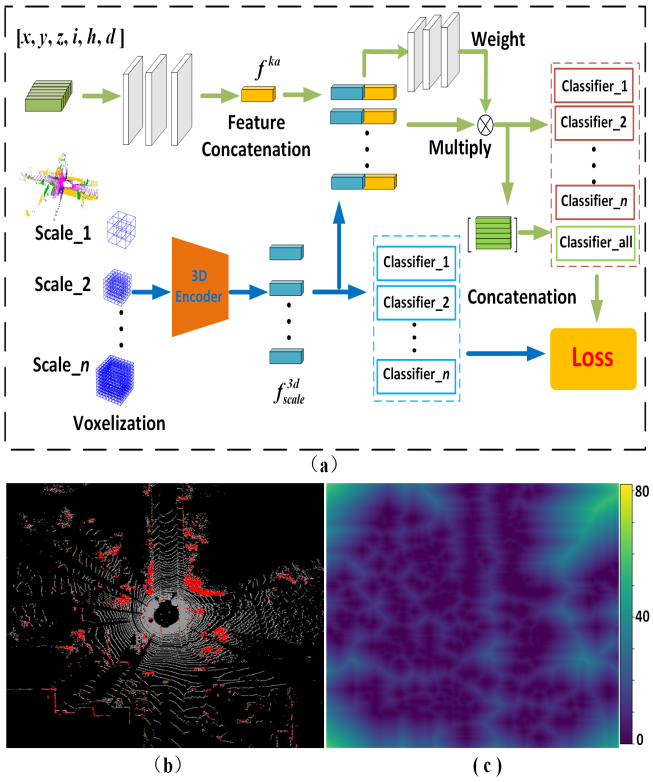

Fig. 4. Implicit knowledge augmentation: (a) the architecture of implicit knowledge augmentation; (b) Obstacle map; (c) NOD map.

module. After that, f^{ka} is connected with point cloud 3D features $f^{3d}_{scale_i}$ of different scales, and the mixture features $f^{3dka}_{scale_i}$ used for classification are calculated as follows:

$$f^{3dka}_{scale_i} = L_{\text{relu}}(W_{scale_i} * L_{\text{linear}}(L_{\text{cat}}(f^{3d}_{scale_i}, f^{ka}))) \quad (3)$$

where W_{scale_i} are weight vectors, which are also derived from mixture features:

$$W_{scale_i} = L_{\text{sigmoid}}(L_{\text{MLP}}(L_{\text{cat}}(f^{3d}_{scale_i}, f^{ka}))) \quad (4)$$

In addition, the module contains two independent sets of multi-layer perceptron classifiers, each containing classifiers for processing classification features obtained at different scales. In particular, the classifier C^{3dka}_{all} is used to process the aggregated feature of all scale features. The prediction results are as follows:

$$\begin{aligned} p^{3d}_{scale_i} &= C^{3d}_{scale_i}(f^{3d}_{scale_i}) \quad p^{3dka}_{scale_i} = C^{3dka}_{scale_i}(f^{3dka}_{scale_i}) \\ p^{3dka}_{all} &= C^{3dka}_{all}(L_{\text{cat}}(f^{3dka}_{scale_0}, f^{3dka}_{scale_1}, \ldots f^{3dka}_{scale_n})) \end{aligned} \quad (5)$$

We also adopt the combination strategy of the loss function in MSFSKD, which is composed of cross-entropy, Lovasz losses, and KL divergence, and the above loss function is

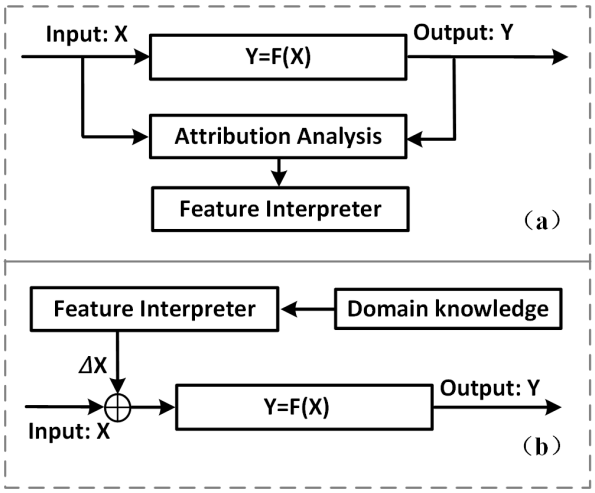

Fig. 5. Explicit knowledge augmentation: (a) Attribution analysis of the model; (b) Model modulation.

calculated as follows:

$$\begin{aligned} Loss_{\text{ce}} &= \sum_{i=1}^{n}(L_{\text{ce}}(p^{3d}_{scale_i}, L_{\text{truth}}) + \lambda_1 L_{\text{ce}}(p^{3dka}_{scale_i}, L_{\text{truth}})) \\ &+ L_{\text{ce}}(p^{3dka}_{all}, L_{\text{truth}}) \\ Loss_{\text{ll}} &= \sum_{i=1}^{n}(L_{\text{ll}}(p^{3d}_{scale_i}, L_{\text{truth}}) + \lambda_1 L_{\text{ll}}(p^{3dka}_{scale_i}, L_{\text{truth}})) \\ &+ L_{\text{ll}}(p^{3dka}_{all}, L_{\text{truth}}) \\ Loss_{\text{kl}} &= \lambda_2 \sum_{i=1}^{n} D_{\text{KL}}(p^{3d}_{scale_i}, p^{ka}_{scale_i}) \end{aligned} \quad (6)$$

where L_{truth} represents ground truth; n means the number of scales; λ_1 and λ_2 are the weights of the loss function. Therefore, the final loss function of the knowledge distillation module is as follows:

$$Loss = Loss_{\text{ce}} + Loss_{\text{ll}} + Loss_{\text{kl}} \quad (7)$$

3.4 Explicit knowledge augmentation

In contrast to implicit knowledge augmentation, which is involved solely in the training phase, explicit knowledge augmentation is applied exclusively during the inference phase. The key to network modulation lies in calculating the correction ΔX of the input feature X. We propose two guiding principles for determining ΔX:

- After applying the correction ΔX, the new output should exceed the original output:

$$|F(X + \Delta X) - L_{\text{truth}}| < |F(X) - L_{\text{truth}}| \quad (8)$$

- The correction ΔX should be as sparse as possible. The correction ΔX_1 is considered superior to ΔX_2, if

$$\begin{cases} |F(X + \Delta X_1) - L_{\text{truth}}| = |F(X + \Delta X_2) - L_{\text{truth}}| \\ f_{\text{sparse}}(\Delta X_1) < f_{\text{sparse}}(\Delta X_2) \end{cases} \quad (9)$$

where f_{sparse} is a sparsity measure. A sparser correction ΔX represents a more efficient network modulation.

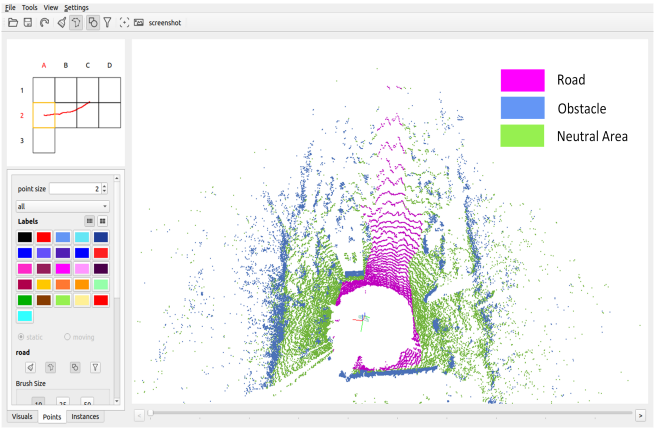

Fig. 6. Point clouds annotation tool. Road labels are new annotations in our work.

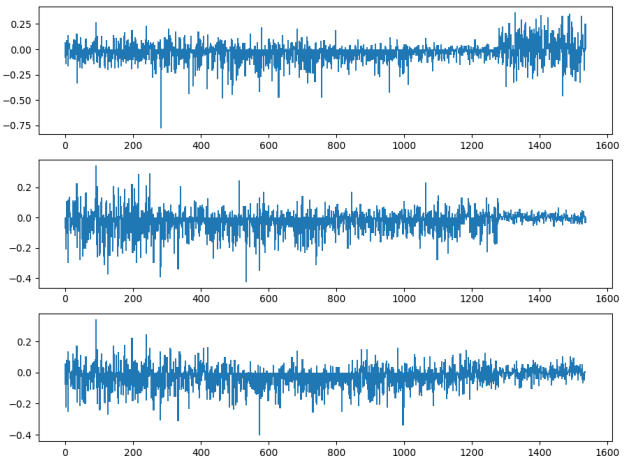

Fig. 7. Feature values of different classes. From top to bottom, there are road, obstacle, and neutral area classes.

In our work, the backbone of our PCSS network is from 2DPSS [10], which contains the 3D encoder module, 3D decoder module, and classifier module. Attribution analysis is performed on the trained classifier module to obtain a feature interpreter by evaluating the influence of different components of the input feature X on the output Y, as shown in Fig. 5(a). The attribution analysis method employed in explicit knowledge augmentation is SHAP, a pioneering approach in the field of machine learning interpretability [5]. Originating from cooperative game theory, SHAP offers a rigorous framework for explaining the output of any machine learning model by assigning each feature's contribution to the final prediction. The feature interpreter can map the human domain knowledge to the feature space to realize feature augmentation, as shown in Fig. 5(b). Concretely, for a classification problem, standard features for each category can be obtained using statistical methods in advance, and the classifier's outputs for these standard features are unambiguous. Generally, the mean values of the corresponding features for each category can serve as the standard features. Subsequently, domain knowledge can be transformed into prior judgments about the class of some points. For these points, the feature components with larger SHAP values are replaced with the corresponding components from the standard features by the correction ΔX.

4. EXPERIMENTS

4.1 Data labeling

This study aims to enhance the generalization of the PCSS network trained in urban scenes and applied to off-road scenes. Currently, there are many open datasets for urban scenes, and considering the dataset utility and the perfection of related auxiliary tools, we selected the SemanticKITTI dataset for training. However, there are relatively limited point cloud semantic segmentation datasets for off-road scenes [28]. Ultimately, we chose the RELLIS-3D dataset due to its compatibility with the SemanticKITTI data format. Nevertheless, the RELLIS-3D dataset lacks road labels, prompting us to utilize the point cloud semantic labeling tool[1] to annotate road areas, as illustrated in Fig. 6. To standardize the target categories in the scenes, we remapped the classes in the aforementioned two datasets into Road, Obstacle, Neutral area, and Others. Further details can be found on our website.

4.2 Experiment implementation

Our method was implemented in Python, with details provided below. For data augmentation, the code for the guided filter is derived from an open-source program[2]. We set key parameters window radius $r = 4$, and the guided image is the same as the input image. Additionally, the threshold control parameter p_I is set to 0.95; For implicit knowledge augmentation, the grid map resolution r_{grid} and obstacle height threshold H_{th} are set to 0.2m and 0.15m, respectively. For the weights in Eq.(7), we set $\lambda_1 = 0.17$ and $\lambda_2 = 0.008$, and the num of scale in the knowledge distillation module is 6. For explicit knowledge augmentation, we use the PCSS network proposed in [10] as the backbone, adopting the original training details with SGD optimizer and learning rate $lr = 0.001$. In addition, the code for SHAP is also sourced from an open-source program[3]. The mean SHAP value obtained after attribution analysis of input features is a 1,536-dimensional vector for different classes. The standard features of different classes are shown in Fig. 7, from which it is obvious that the features corresponding to different classes have distinct differences. In the explicit knowledge augmentation phase, our modulation strategy focuses solely on road class, with the road points' distribution obtained through domain knowledge and used to modulate the input features of the classifier module. Specifically, Tab. I presents two optimization strategies involved in this work, along with their corresponding domain knowledge and implementation methods.

[1] https://github.com/jbehley/point_labeler
[2] https://github.com/swehrwein/python-guided-filter
[3] https://github.com/shap/shap

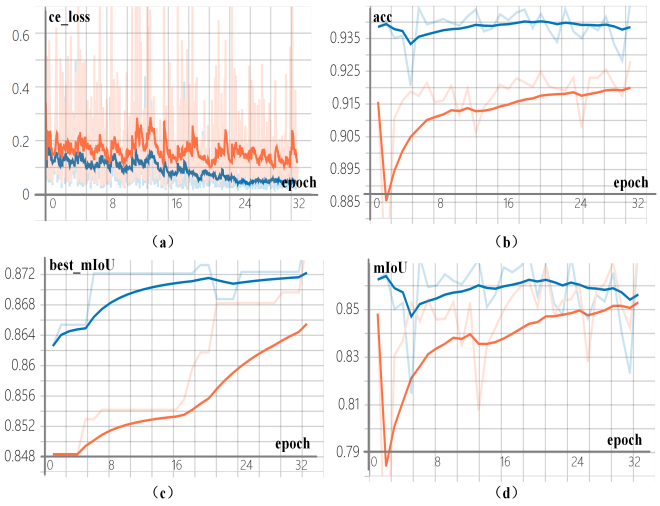

Fig. 8. Training process comparison. (a) cross-entropy loss; (b) accuracy; (c) best mIoU, and (d) mIoU. Blue is our method, and orange is 2DPASS. Moreover, the color of raw data is light, while smoothed data is dark.

TABLE I

DOMAIN KNOWLEDGE AND IMPLEMENTATION

Strategy	Domain Knowledge	Implementation
S1	Road areas are far from obstacles.	Road points correspond to locations on the NOD map where values exceed 4.
S2	Road areas are continuous and of low height.	Filter out road points.

To demonstrate the effectiveness of our method, we selected five previous PCSS methods as the comparison group. Since SemanticKITTI is used as the training dataset in this work, we adopted the top two methods in the leaderboard of the PCSS task on SemanticKITTI, namely 2DPASS [10] and PVKD [8]. Meanwhile, we also selected three recently published PCSS methods with domain adaptation ability, namely 3DLP [17], PointDR [16], and CoSMix [18]. Note that CoSMix needs a target sequence in RELLIS-3D in the adaptation stage, so this sequence (Seq. 35) is excluded from the test set. All experiments in this work were conducted on an NVIDIA 3090 GPU without pre-trained weight. Due to the hardware limitations, the batch sizes during training were set to the maximum value that satisfies the hardware constraints, and the maximum epoch was set to 32.

4.3 Analysis of results

During the training phase, the cross-entropy loss of the proposed method exhibits a more rapid descent compared to the original backbone method. Simultaneously, our method demonstrates advantages in accuracy and mIoU (mean Intersection over Union) on the validation sequence of the SemanticKITTI dataset (Seq.08), as illustrated in Fig. 8. This improvement is attributed to the normalization and filtering of the training data, which prevents large oscillations and accelerates training convergence.

Table. II presents the statistical results of semantic segmentation with different methods on the RELLIS-3D dataset. Our method achieves the most effective obstacle detection when only leveraging data augmentation and implicit knowledge augmentation, but its performance in classifying road and neutral areas is inferior. When data augmentation (ground height normalization) is omitted, the semantic segmentation performance of implicit knowledge augmentation deteriorates, which again proves the foundation and necessity of data augmentation. Furthermore, among all comparative methods, only PVKD attains an IoU_{road} exceeding 0.5, while other comparison methods demonstrate weaker performance. An obvious conclusion is that all methods perform better at recognizing obstacles than at distinguishing road and neutral areas. The reasons for this phenomenon are as follows: (1) The training dataset lacks sufficient samples to clearly differentiate between road and neutral areas; (2) In off-road scenes, the boundary between road and neutral areas is often indistinct; (3) The features of obstacles are more consistent across different environments.

TABLE II

STATISTICAL RESULTS OF PCSS WITH DIFFERENT METHODS

Method	IoU_{obs}	IoU_{road}	IoU_{neu}	mIoU	Acc
PVKD [8]	0.759	0.584	0.376	0.573	0.756
2DPASS [10]	0.752	0.178	0.384	0.438	0.693
PointDR [16]	0.724	0.114	0.353	0.397	0.671
3DLP [17]	0.796	0.068	0.440	0.435	0.713
CoSMix [18]	0.757	0.101	0.331	0.397	0.679
Ours(I)	0.740	0.052	0.299	0.364	0.657
Ours(D+I)	**0.815**	0.170	0.442	0.476	0.735
Ours(D+I+E1)	0.805	0.629	0.493	0.642	0.808
Ours(D+I+E1+E2)*	0.812	**0.652**	**0.493**	**0.652**	**0.814**

* D and I denote data augmentation and implicit knowledge augmentation, respectively; E1 means explicit knowledge augmentation with S1, and so is E2.

TABLE III

STATISTICAL RESULTS OF GROUND SEGMENTATION

Method	IoU_{obs}	$\text{IoU}^{*}_{\text{ground}}$	mIoU	Acc
PVKD [8]	0.759	0.715	0.737	0.850
2DPASS [10]	0.752	0.708	0.730	0.845
PointDR [16]	0.724	0.671	0.698	0.828
3DLP [17]	0.796	0.768	0.782	0.878
CoSMix [18]	0.757	0.675	0.716	0.838
Ours(I)	0.740	0.645	0.692	0.823
Ours(D+I)	**0.815**	**0.779**	**0.797**	**0.888**
Ours(D+I+E1)	0.805	0.772	0.789	0.883
Ours(D+I+E1+E2)	0.812	0.778	0.795	0.887

* Both road and neutral areas are considered as ground.

Given these reasons, we apply explicit knowledge augmentation to identify road areas and test the two strategies proposed in Tab. I. In Strategy 1, we deploy the NOD map introduced in Section 3.3 to select road points by setting the distance threshold to 4, and the feature vectors of road points are partially modulated as standard features of the road class. Specifically, feature components with SHAP values in the top 36% are modulated. Experimental results demonstrate that this method can significantly improve the IoU_{road}; however, it concurrently reduces the precision of obstacle detection. To mitigate this issue, Strategy 2 is incorporated to correct

the bias caused by Strategy 1 and minimize misclassifying obstacle points as roads. By integrating these two strategies through explicit knowledge augmentation, our method achieves optimal performance in both mIoU and overall accuracy, as shown in Tab. II.

For a more comprehensive evaluation, we present the statistical results of ground segmentation in Tab. III, where both road and neutral areas are classified as ground. Because the boundary between the road and neutral areas is blurred in off-road scenes, the road belongs to the optimal drivable area subjectively determined by humans. The results in Tab. III indicate that the implicit knowledge augmentation and data augmentation proposed in this work achieve optimal obstacle detection, which is crucial for the safe passage of unmanned vehicles. Although obstacle detection performance IoU_{obs} slightly decreases after integrating explicit knowledge enhancement (from 0.815 to 0.812), its positive impact on improving IoU_{road} (from 0.170 to 0.652) and overall accuracy (from 0.476 to 0.652) is notably significant. We believe that this method will have greater potential after fusing more complex human knowledge.

5. CONCLUSIONS

This work proposes a domain-adaptive PCSS method based on knowledge augmentation, systematically achieving knowledge augmentation at the data, architecture, and decision levels and enhancing the interpretability of deep learning. Additionally, we added new annotations to the off-road dataset RELLIS-3D, which can be used to test the generalization of the PCSS method from urban to off-road scenes in the future. The proposed method shows competitive advantages on the test dataset, and we believe that it can be applied to more diverse scenes. Furthermore, due to the hardware limitations, subsequent testing on devices with better hardware can more comprehensively evaluate the method proposed in this work.

REFERENCES

[1] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, Semantickitti: A dataset for semantic scene understanding of lidar sequences, in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 9297-9307.

[2] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, Rellis-3D dataset: Data, benchmarks and analysis, in 2021 IEEE international conference on robotics and automation, 2021, pp. 1110-1116.

[3] J. Wörmann, D. Bogdoll, E. Bührle, H. Chen, E. F. Chuo, et at., Knowledge augmented machine learning with applications in autonomous driving: A survey, 2022, arXiv preprint arXiv:2205.04712.

[4] Z. Cui, T. Gao, K. Talamadupula, and Q. Ji, Knowledge-augmented deep learning and its applications: A survey, IEEE Transactions on Neural Networks and Learning Systems, 2023.

[5] S. M. Lundberg, and S. I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems, 2017.

[6] G. Erion, J. D. Janizek, P. Sturmfels, S. M. Lundberg and S. I. Lee, Improving performance of deep learning models with axiomatic attribution priors and expected gradients, Nature machine intelligence, 2021, pp. 620-631.

[7] F. Xu, H. Liang, Z. Wang and L. Lin, A Framework for Drivable Area Detection Via Point Cloud Double Projection on Rough Roads, Journal of Intelligent Robotic Systems, 2021, pp. 102(2): 45.

[8] Y. Hou, X. Zhu, Y. Ma, C. C. Loy, and Y. Li, Point-to-voxel knowledge distillation for lidar semantic segmentation, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 8479-8488.

[9] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, Multi-view 3d object detection network for autonomous driving, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1907-1915.

[10] X. Yan, J. Gao, C. Zheng, C. Zheng, R. Zhang, S Cui and Z. Li, 2DPASS: 2d priors assisted semantic segmentation on lidar point clouds, European conference on computer vision, 2022, pp. 677-695.

[11] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652-660.

[12] Q. Hu, B. Yang, et at., Randla-net: Efficient semantic segmentation of large-scale point clouds, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 11108-11117.

[13] 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, 2018, pp. 9224-9232.

[14] A. Ando, S. Gidaris, A. Bursuc, G. Puy, A. Boulch and R. Marlet, Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 5240-5250.

[15] X. Lai, Y. Chen, F. Lu, J. Liu and J. Jia, Spherical transformer for lidar-based 3d recognition, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, 17545-17555.

[16] A. Xiao, J. Huang, et al., 3D semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 9382-9392.

[17] J. Sanchez, J. E. Deschaud and F. Goulette, Domain generalization of 3D semantic segmentation in autonomous driving, in Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 18077-18087.

[18] C. Saltori, F. Galasso, G. Fiameni, N. Sebe, E. Ricci and F. Poiesi, CoSmix: Compositional semantic mix for domain adaptation in 3d lidar segmentation, European conference on computer vision, 2022, pp. 586-602.

[19] Y. Zhou, Techniques for Interpretability and Transparency of Black-Box Models, Massachusetts Institute of Technology, 2023.

[20] M. Gaur, U. Kursuncu, A. Sheth, R. Wickramarachchi and S. Yadav, Knowledge-infused deep learning, in Proceedings of the 31st ACM Conference on hypertext and social media, 2020, pp. 309-310.

[21] L. Von Rueden, S. Mayer, et al., Informed Machine Learning? A taxonomy and survey of integrating prior knowledge into learning systems, IEEE Transactions on Knowledge and Data Engineering, 2021, pp. 614-633.

[22] A. D. A. Garcez, S. Bader, et al., Neural-symbolic learning and reasoning: a survey and interpretation, Neuro-Symbolic Artificial Intelligence: The State of the Art, 2022, pp. 342(1).

[23] J. Mei, H. Zhao, Incorporating human domain knowledge in 3D LiDAR-based semantic segmentation, IEEE Transactions on Intelligent Vehicles, 2019, pp. 178-187.

[24] X. Liang, Z. Hu, H. Zhang, L. Lin and E. P. Xing, Symbolic graph reasoning meets convolutions, Advances in neural information processing systems, 2018.

[25] Z. Xu, W. Zhang, P. Ye, H. Chen, and H. Chen, Neural-symbolic entangled framework for complex query answering, Advances in Neural Information Processing Systems, 2022, pp. 1806-1819.

[26] S. Luan, C. Chen, et al., Gabor convolutional networks, IEEE Transactions on Image Processing, 2018, pp. 4357-4366.

[27] K. He, J. Sun and X. Tang, Guided image filtering, IEEE transactions on pattern analysis and machine intelligence, 2012, pp. 1397-1409.

[28] M. Liu, E. Yurtsever, X. Zhou, J. Fossaert, Y. Cui, B. L. Zagar and A. C. Knoll, A survey on autonomous driving datasets: Data statistic, annotation, and outlook, in IEEE Transactions on Intelligent Vehicles, 2024.