2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Multi-Confidence Guided Source-Free Domain Adaption Method for Point Cloud Primitive Segmentation

Shaohu Wang, Yuchuang Tong, Xiuqin Shang, Zhengtao Zhang*, *Member, IEEE*

***Abstract*— Point cloud primitive segmentation aims to segment the surface point cloud into various geometric types of primitives, which plays a vital role in robot operation and industrial automation. However, differences in object structures and shapes across industrial datasets create domain shift issues, compounded by privacy concerns preventing dataset sharing. To address these challenges, we propose a novel source-free domain adaptation method for point cloud primitive segmentation, which follows the popular pseudo-label based self-training framework. Unlike previous works using single-model uncertainty to refine pseudo labels, our method leverages multi-confidence, including transformation consistency, task confidence, and geometric saliency to provide more informative guidance. Specifically, the transformation consistency is first utilized to vote pseudo-labels and task confidences. Furthermore, to filter out high-confident noises and obtain more reliable pseudo-labels, we investigate the geometric curvature properties of primitives and propose a geometric saliency guided dynamic prototype matching and label graph aggregation strategies for pseudo-label reassignment with different task confidence. For this novel task, we construct several datasets and verify the effectiveness of the proposed methods through a series of experiments.**

I. Introduction

3D primitive segmentation task refers to decomposing object's surface point cloud into a set of geometirc primitives, such as planes, spheres and complex surfaces, etc [1], which plays a crucial role in many downstream robotics tasks [2], like region-level measurements and surface quality inspections. Traditional RANSAC-based methods [3]–[6] are effective for handling simple and regular shapes, but is limited by time-consuming parameter tuning and cannot deal with complex shapes. Recently, the deep learning (DNN) based primitive segmentation approaches have developed rapidly and significantly surpassed traditional methods [7].

However, there are three main challenges limiting further development of primitive segmentation. Firstly, deep learning methods heavily rely on the amount of labeled data, which is difficult for point cloud segmentation [8], especially for fine primitive annotation. Secondly, as shown in Fig. 1, although the types of geometric primitives are finite, the structure and distribution of the point clouds in different tasks may differ greatly. In particular, the real collected point cloud may have much deformations and noises, which leads to the domain gap between source and target dataset. Furthermore, privacy issues are of great importance in industrial production, datasets from different sources are usually unable to share, thus it is difficult to avoid the domain-shift. Therefore, how to enhance the performance of primitive segmentation models on unlabeled target data where source data is inaccessible is the primary motivation of this paper.

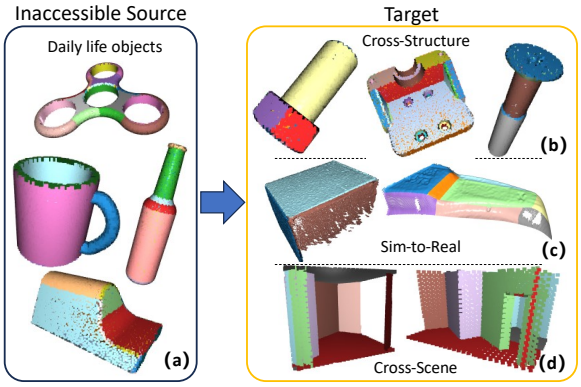

Fig. 1. Source-Free Domain Adaption for 3D Primitive Segmentation. Point cloud primitives are labeled by different colors. (a) Objects in Source dataset (Parts0), which are inaccessible during adaption. (b-d) Target objects with different structure (Parts1), large noise (ReaL) and different scene (Scene) from Source, which are unlabeled during adaption.

Source-Free Unsupervised Domain Adaptation (SF-UDA) methods [9] are introduced to address the aforementioned challenges. SF-UDA aims to perform unsupervised domain adaptation using only source-trained model without the access to source data. The pseudo-label based self-training method is most commonly used method for SF-UDA [10], which follows a teacher-student architecture, leveraging the ensemble predictions of the teacher model to generate reliable pseudo-labels for training the student model, and iteratively optimizing the weights of both models. However, relevant studies on point cloud primitive segmentation are still lacking. Moreover, these methods often yield unsatisfactory results when directly applied to primitive segmentation tasks, as they do not utilize the geometric properties of primitives.

In this paper, we propose a novel source-free domain adaptation method for point cloud primtive segmentation, which is guided by multi-confidence consisting of transformation consistency, task confidence and geometric saliency. Com-

*This work is supported by the National Natural Science Foundation of China (62303457, U21A20482), the National Key Research and Development Program of China (2022YFB3303800), Project funded by China Postdoctoral Science Foundation (2023M733737). *(Corresponding author: Zhengtao Zhang.)*

Authors are with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, also with the School of Artificial Intelligence, University of Chinese Academy of Science, Beijing 100049, China, also with CAS Engineering Laboratory for Intelligent Industrial Vision, Beijing 100190, China. Zhengtao Zhang is also with Binzhou Institute of Technology, Binzhou City 256601, Shandong Province, China (e-mail: wangshaohu2020@ia.ac.cn; zhengtao.zhang@ia.ac.cn).

pared to using only model uncertainty [11], multi-confidence guidence can provide richer information for generating more robust pseudo-labels. In particular, the proposed geometric saliency reflects the curvature properties of different primitives, which can help learning more discriminative features.

In summary, our contributions are as follows:

1) We propose a multi-confidence guided SF-UDA method for point cloud primitive segmentation, which leverages transformation consistency, task confidence and geometric saliency to obtain more robust pseudo-labels. Besides, an instance confidence is proposed to assess the credibility of instance embeddings.
2) Based on the proposed geometric saliency, we design a geometric saliency-guided dynamic prototype matching and label graph aggregation method for reassigning pseudo-labels with different task confidences.
3) To the best of our knowledge, this is the first work to explore SF-UDA for point cloud primitive segmentation, we built several datasets for this task using both online public and real collected point clouds. Our method achieves SOTA performance on all datasets.

II. Related Work

A. Deep Learning-based Primitive Segmentation

With the rapid development of deep learning on point cloud [12]–[14], many DNN-based 3D primitive segmentation have emerged [15]–[21]. SPFN [18] first proposed an end-to-end supervised primitive segmentation method to predict the attributes of each point in a point cloud. However, only four basic primitive types can be handled. ParseNet [19] use B-spline surfaces to represent complex primitives other than the four basic primitives, which is used as the backbone of our study. [20]–[22] optimized the network structure based on [19] to further improve the primitive segmentation performance. These methods reflect the efficacy of deep learning in primitive segmentation but are constrained by the amount of annotated samples, while our method can effectively alleviate this issue.

B. Domain Adaption

Unsupervised Domain Adaptation: Unsupervised Domain Adaptation (UDA) aims to transfer models trained on source domain to unlabeled target domain, usually by aligning feature distributions between two domains or by leveraging generative models to mitigate domain shifts [23]. There were also some works addressing domain adaptation for 3D point clouds [24]–[27]. For point cloud segmentation, [28] proposed a dynamic feature graph matching method to mitigate domain shift, while PMAN [29] introduced a density-aware network to align the target and source domains. However, these methods still require access to the source domain, which is impractical for industrial scenarios with high privacy requirements.

Source-Free Domain Adaptation: Source-Free Domain Adaptation (SF-UDA) enables model adaptation using only source-trained model without accessing source domain data. Current SF-UDA paradigms mainly consists of generative modeling [30], auxiliary self-supervised training [31] and self-training [10], [11], [32]–[36], where pseudo-label based self-training method is the commonly used technique. SHOT [32] obtained more reliable pseudo-labels by information maximization loss and category prototyping, which is used as our baseline. SR [31] added an auxiliary branch for category ratio estimation, but ignored the difference between the category distributions of the source and target domains, need to additionally train new branching in the source domain. [10], [11], [33]–[36], on the other hand, denoised and optimized the pseudo-labels by quantifying the uncertainty of the model, but all of them stay at the feature level of the model, while our approach introduced a more intuitive geometric saliency, which further enhances the reliability of the pseudo-labels. In addition, researches on source-free domain adaptation methods for point cloud primitive segmentation or instance segmentation remains scarce.

III. Method

A. Preliminaries and Overview

Point cloud primitive segmentation network, denoted PSNet takes point coordinates $\boldsymbol{p} \in \mathbb{R}^{N\times 3}$ and normals $\mathbf{n} \in \mathbb{R}^{N\times 3}$ as input X and segments the points into M primitives with types, where N is the number of points. The outputs include the primitive instance $I \in \{0,\dots,M\}^{N\times 1}$ and geometric type $T \in \{0,\dots,5\}^{N\times 1}$ (corresponding to six types: plane, sphere, cylinder, cone, open and closed B-splines). The structure of PSNet is based on [19], as illustrated in Fig .2 (a). It consists of an encoder and a decoder, denoted as $\Phi = \mathcal{F}_e \circ \mathcal{F}_d$, where encoder contains three EdgeConv block from [14]. The input point cloud is first processed by the encoder to get the latent features, and then the features enter semantic segmentation and instance embedding learning branches to obtain the point-level categories T and instance embeddings $\mathbf{e}$, respectively. Finally, the embeddings are clustered to obtain primitive instances I.

Let $\mathcal{D}_s$ be the dataset in source domain $\mathcal{S}$, comprising input point clouds $\mathcal{X}_s$ and corresponding labels $\mathcal{Y}_s$, and $\mathcal{D}_t$ denotes the dataset in target domain $\mathcal{T}$, which only contains point clouds $\mathcal{X}_t$. In the source-free domain adaption setting, $\mathcal{D}_s$ is inaccessible during adaptation. The goal is to transfer the model Φ_s trained on Ds to $\mathcal{T}$ to obtain Φ_t while minimizing the performance degradation due to domain-shift.

The main framework of our adaption method is shown in Fig. 2 (b), which follows a pseudo-label based self-training framework, where teacher and student models are both initialized by Φ_s. Our method has three main steps: pseudo-label voting, pseudo-label reassignment and self-training, which is guided by multi-confidence of transformation consistency, task confidence and geometric saliency.

B. Consistency-based Pseudo-Label Voting

Inspired by [37], we adopt random transformations on X to vote pseudo-labels. Specifically, the input point cloud is first transformed by random rotations around z-axis to obtain a set of augmented samples, denoted as

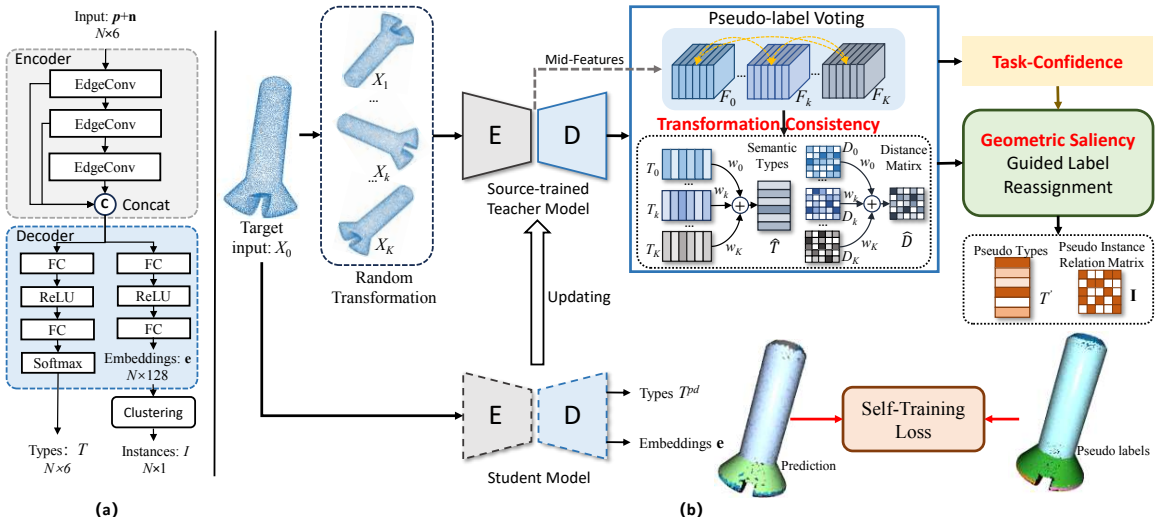

Fig. 2. (a) Structeure of point cloud primitive segmentation network. (b) Multi-confidence guided SF-UDA method for point cloud primitive segmentation.

$\{X_0, X_1, \cdots, X_K\}$, where K is the number of transformations. Then these samples are fed to teacher model and outputs $\{(\sigma_0(t), \mathbf{e}_0), \cdots, (\sigma_K(t), \mathbf{e}_K)\}$, where $\sigma_k^i(t)$ is the softmax type probability vector of kth transfromed point i.

Furthermore, the latent features $\{F_0, \ldots, F_k\}$ of the model are utilized for measuring the transformation consistency. Specifically, we calculate the KL divergence KL_k of each feature F_k with the others, and the consistency weights w are then computed using the Softmax function:

$$w_k = \frac{e^{-KL_k}}{\sum_{j=0}^{K} e^{-KL_j}}, KL_k = \sum_{i \neq k} \mathrm{KL}(F_i \| F_k) \tag{1}$$

For semantic labels, we perform weighted average of $\sigma_k(t)$ to get the pseudo probability vector,i.e.,

$$\hat{\sigma}(t) = \sum_{k=0}^{K} w_k \cdot \sigma_k(t) \tag{2}$$

where the index and value of maximum probability refer to the semantic type pseudo-label $\hat{T}_i$ and confidence ρ_i, the weighted feature $\hat{F}$ can be obtained by the same way.

For instance labels, we compute the point-to-point distance matrix D_k for each instance embedding $\mathbf{e}_k$, where $D_k(i,j) = \|\mathbf{e}_k(i) - \mathbf{e}_k(j)\|_2$. Afterwards, the pseudo-instance distance matrix $\hat{D}$ is obtained by weighted average voting over all D:

$$\hat{D}(i,j) = \sum_{k=0}^{K} w_k \cdot D_k(i,j) \tag{3}$$

Besides, we propose a point cloud instance confidence representation based on the idea of metric learning, i.e., the distance between same classes is close while the distance between different classes is far. In scenarios where true labels are unknown, we expect the embedding distance to be either very large or small. Hence, the instance confidence is formulated as:

$$\rho_i^I = \frac{1}{N-1} \sum_{j \neq i} 1 - 2 \frac{\min(\hat{D}_{ij}^2, \max(0, \Theta - \hat{D}_{ij}^2))}{\Theta} \tag{4}$$

where the upper distance limit Θ is set to 1. It is evident that higher values will be obtained when D_{ij} is either large or small. Furthermore, this confidence also reflects the degree of points to the instance embedding center, which can be proved through Cauchy-Schwarz inequality.

C. Geometric Saliency Guided Label Reassignment

In this section, we propose label reassignment methods to denoise pseudo-labels with different task confidence by utilizing the geometric saliency, which is defined as the distribution characteristics of geometric curvature on different types of primitives, e.g., points on the plane usually have zero curvatures, whereas points on a sphere have non-zero and constant curvatures. We calculate the point-wise curvature c and curvature variance v within its local neighborhood as two geometric Saliency features. The type-wise geometric saliency $g(c_i, v_i | T_i)$ defined as:

$$g(c_i, v_i | T_i = t) = \begin{cases} \dfrac{1}{1 + e^{c_i \cdot v_i}}, & t = 0 \\ \dfrac{1}{1 + e^{v_i}}, & t \in \{1, 2\} \\ 1, & others \end{cases} \tag{5}$$

We use task confidence to divide points into high-confident P^H, mid-confident P^M, and low-confident points P^L by two thresholds η_h and η_l. Then the following methods are proposed for denosing pseudo-labels with different confidences.

1) **Geometric Saliency-Guided Prototype Matching:** To denoise wrong high-confident points and avoid low-confident oscillating points, we propose a geometric saliency-guided dynamic prototype matching method inspired by [32], as illustrated in Fig. 3. Firstly, the geometric saliency is leveraged for computing weighted average of high-confidence

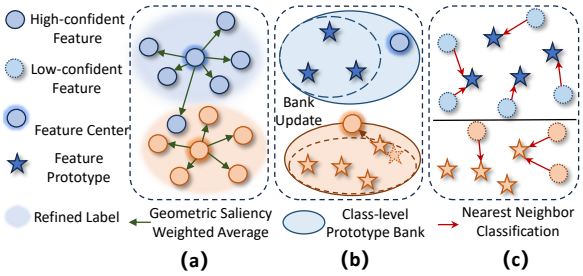

Fig. 3. Geometric Saliency-Guided Prototype Matching. Different colors represent differnt types. (a) Sample-level feature center computation and high-confident pseudo label denoising. (b) Dynamic feature prototype bank updating. (c) Nearest Neighbor algorithm is employed to determine types of low-confident points.

features to obtain class-level centroids $\mathbf{z}$, as formed by,

$$\mathbf{z}_t = \frac{1}{\sum_{T_i=t} g(c_i, v_i|T_i=t)} \sum_{T_i=t} g(c_i, v_i|T_i=t) \cdot \hat{f}_i \quad (6)$$

Subsequently, all high-confidence points are assigned new labels based on nearest neighbor computation.

However, due to structure disparities, sample-level points' centers cannot be able to cover all categories, which is disadvantageous for low-confident oscillating points. We design a class-level dynamic prototype matching method for robust label updating of low-confident points. For each class t, the prototype bank Z_t starts as an empty set. In r iteration, if Z_t is still empty, $\mathbf{z}_t^r$ will be directly added to it. Otherwise, the nearest prototype $\widehat{\mathbf{z}}_t$ in Z_t with cosine distance d_z are computed and the self-update strategy will be adopted:

$$\begin{cases} Z_t^r \leftarrow Z_t^{r-1} \cup \{\mathbf{z}_t^r\}, & d_z \geq \xi \\ \widehat{\mathbf{z}}_t^r \leftarrow \widehat{\mathbf{z}}_t^r + \frac{1}{n_{z_t}} \cdot (\mathbf{z}_t^r - \widehat{\mathbf{z}}_t^{r-1}), & d_z < \xi \end{cases} \quad (7)$$

where n_{z_t} denotes the times $\widehat{\mathbf{z}}_t$ has been updated, and feature distance threshold ξ is set to 0.2. For low-confident points, we compute nearest point in the feature bank Z to determine their corresponding labels.

2) **Geometric Saliency-Guided Label Graph Aggregation:** For mid-confident points, whose features may be ambiguous, we propose a geometric saliency guided label graph aggregation method by utilizing local consistency for label denoising. For each point $p_i \in P^M$, we first utilize KNN algorithm to obtain its K neighboring points in P: $\Omega_i = (f_{i0}, \cdots, f_{iK})$ and construct an adjacency graph $\mathcal{G}_i$ centered around p_i. To prevent the influence between different prmitives for boundary points, we use instance distances to adjust distance of KNN, the adjusted $d_{i,j}$ is given by,

$$d_{i,j} = \|p_i - p_j\|_2 \cdot (2 - \frac{\mathbf{n}_i \cdot \mathbf{n}_j}{\|\mathbf{n}_i\| \cdot \|\mathbf{n}_j\|}) \cdot \hat{D}_{i,j} \quad (8)$$

where $\mathbf{n}$ is the normal and $\hat{D}_{i,j}$ refers to instance distance.

The value of node ik in $\mathcal{G}_i$ corresponds to the pseudo-class probability vector $\sigma(p_{ik})$. The edge weight l of each node to the centroid is guided by the geometric saliency, namely,

$$l_{ik} = e^{-d_{i,ik}} \cdot \rho_{ik} \cdot g(c_{ik}, v_{ik}|T_{ik}) \quad (9)$$

Then all node values in $\mathcal{G}_i$ are aggregated to obtain the adjusted class probability vector $\sigma^{'}$,

$$\sigma^{'}(p_i) = \frac{1}{\sum_{k=0}^{K} l_{ik}} \sum_{k=0}^{K} l_{ik} \cdot \sigma(p_{ik}) \quad (10)$$

The refined pseudo type label is defined as $T_i^{'}$.

3) **Semantic-Aware Instance Relation Updating:** For instance segmentation, we design a semantic-aware instance relation updating strategy based on the coupling relationship between instance and semantics, i.e., points with different types are not likely belong to same instance. The refined instance distance is given by,

$$D_{i,j}^{'} = \hat{D}_{i,j} \cdot (1 + \max(0, \epsilon - \frac{\sigma^{'}(p_i) \cdot \sigma^{'}(p_j)}{2\|\sigma^{'}(p_i)\| \cdot \|\sigma^{'}(p_j)\|}) \quad (11)$$

where ϵ is set to 0.5. Finally, $D^{'}$ is binarized into an instance relation matrix $\mathbf{I}$, which indicates whether two points belong to the same instance.

D. Self-Training Strategy

The total training loss contains semantic and instance part, each part consists of a supervised loss and an information maximization loss. The cross-entropy loss from [21] and information maximization loss from [32] are applied for semantic segmentation, denoted as $\mathcal{L}_{sem}^1$ and $\mathcal{L}_{sem}^2$.

For instance training, we propose a confidence-weighted triple loss as the supervised training loss. Several embedding triplet groups set $\mathcal{I}$ are selected from the point cloud according to the instance relation matrix $\mathbf{I}$, given the triplet of embeddings (e^o, e^p, e^n), the triple loss [19] is defined as $L_{tri}(e^o, e^p, e^n)$, the instance supervised loss is given by:

$$\mathcal{L}_{ins}^1 = \frac{1}{\sum_{e^o} \rho^I(e^o)} \sum_{(e^o, e^p, e^n) \in \mathcal{I}} \rho^I(e^o) L_{tri}(e^o, e^p, e^n) \quad (12)$$

Information maximization loss of instance is formed by,

$$\mathcal{L}_{ins}^2 = -\frac{1}{N} \sum_{i=1}^{N} \rho_i^I \quad (13)$$

Thus, the total loss is the sum of above items,

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{sem}^1 + \lambda_2 \mathcal{L}_{sem}^2 + \lambda_3 \mathcal{L}_{ins}^1 + \lambda_4 \mathcal{L}_{ins}^2 \quad (14)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are the balancing weight.

When training on the target dataset, the network's decoder $\mathcal{F}_d$ is fixed, only the encoder $\mathcal{F}_e$ is updated, and the teacher model is updated once per epoch.

IV. Experiments

A. Experimental Setting

1) Dataset and Evaluation Metrics: We construct several datasets from both online public and real-collected point clouds for validation. ABCParts [38] contains numerous 3D CAD models with diverse shapes, which is the most comprehensive dataset for primitive segmentation.While ANSI [18] mainly contains point clouds of industrial parts, where

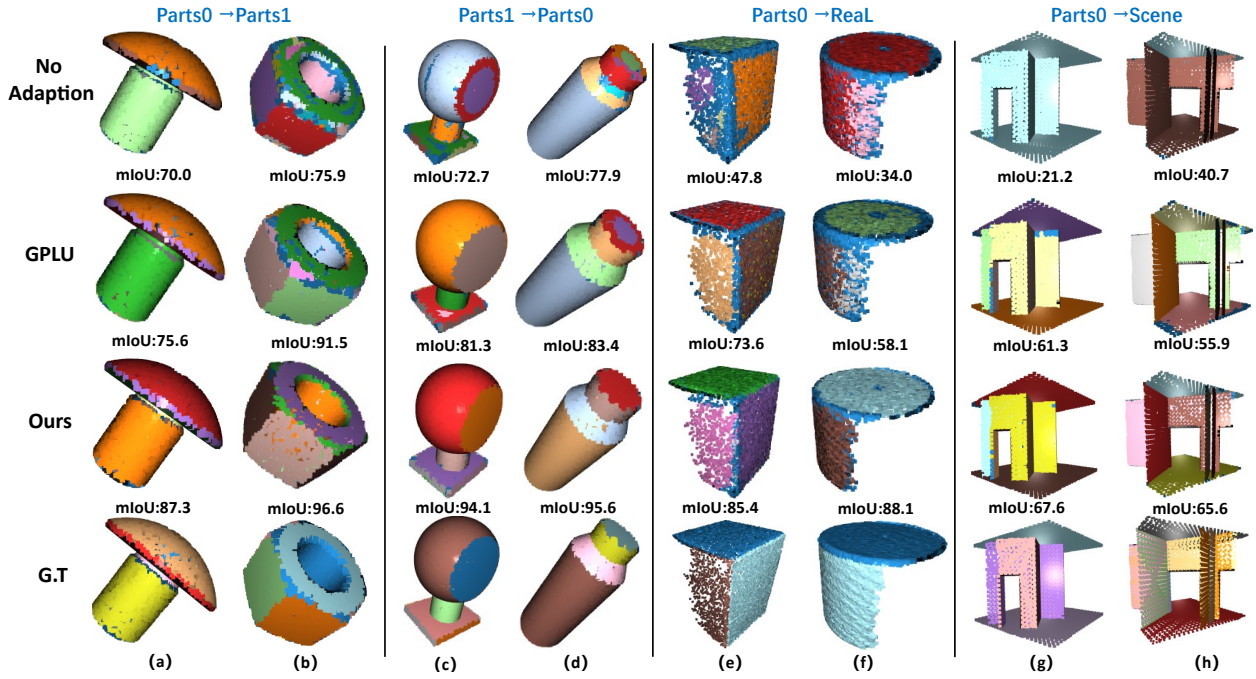

Fig. 4. Visualized primitive segmentation results of domain adaption. The primitive instances are shown in different colors. (a-d) Cross-structure setting. (e-f) Sim-to-real setting. (g-h) Cross-scene setting.

most models are composed by regular primitives. We select 20K models from ABCParts as the source dataset, denoted as Parts0 (Fig .4(c-d)) and choose remaining objects different from Parts0, as well as models from the ANSI to form Parts1 (Fig .4(a-b)). Parts1 contains 10K models with different structures and styles from Parts0. Secondly, we use 3D scanner to capture point clouds of real-world objects, and obtain 5K samples by random augmentation, denoted as ReaL (Fig .4(e-f)), where point clouds may have large noises and distortions, used for evaluating real-world adaptability of the method. In addition, we select building dataset Scene [7], which contains several indoor and outdoor point clouds to verify the model's cross-scene ability. We use a ratio of 4:1 to split the dataset for training and testing.

The Evaluation Metrics include total semantic segmentation accuracy, IoU of primitive semantic and instance segmentation, denoted as ACC, $SIoU$ and $mIoU$. Since most of the primitives in the Scene dataset are plane (floor, wall), the semantic metrics are not taken into consideration.

2) Implementation Details: The model trained on source dataset are preserved before adaption. For each target object, 7000 points are randomly sampled and normalized as input. In the adaption phase, each point cloud is randomly rotated 3 times around the z-axis for voting pseudo-labels. The Adam optimizer with learning rate of 0.0002 and batch size of 2 is used. The weights in Eq. (14) are initialized as 1,0.1,1 and 0.1, η_h and η_l are set to 0.9 and 0.6. All the experiments are conducted with one NVIDIA GeForce RTX3090 GPU.

B. Comparison Experiments

For comparison, some latest related methods [11], [31]–[35], [37] are re-implemented, which share the same network structure and training settings as our method, thus have the same inference time. Since they do not perform domain adaptation for instance segmentation, original triple loss is used. The quantitative results are presented in Table I.

Firstly, as observed from the results of Parts0 → Parts1 and Parts1 → Parts0, our method demonstrates superior cross-structure adaptation ability. Our method combines the strengths of SuperLine [37] and SHOT [32], and thus outperforms them comprehensively in all metrics. Secondly, due to the significant deformations and normal noises, applying the source model directly to ReaL yields limited effectiveness. It is obvious that our method achieves an increase of over 10% in semantic metrics and 18% $mIoU$, surpassing other methods by a considerable margin, illustrating the stronger adaptability to real point clouds. Furthermore, it is evident that in addressing cross-scene domain adaptation tasks, our method provides a highest increase in $mIoU$, surpassing other methods by over 10% on average.

The visualization results are depicted in Fig. 4. In cross-structure setting (Fig. 4 (a-d)), our method achieves more accurate segmentation boundaries. In case of Fig. 4 (e-f), the noise in the actual collected point cloud results in severe over-segmentation, whereas the results after adaption of our method are smoother. In addition, the source model may fail in the Scene, as illustrated in Fig. 4 (g-h). By incorporating the geometric properties, our method can produce more reasonable segmentation results compared to GPLU.

Overall, our method exhibits the strongest overall performance in cross-structure, sim-to-real, and cross-scene primitive segmentation tasks. In particular, it provides a substantial improvement in primitive instance segmentation. Besides, the real-world application of our method in industrial appearance

TABLE I

COMPARSION EXPERIMENTS

Method	Domain Shift									
	Parts0 → Parts1			Parts1 → Parts0			Parts0 → ReaL			Parts0 → Scene
	ACC	*SIoU*	*mIoU*	*ACC*	*SIoU*	*mIoU*	*ACC*	*SIoU*	*mIoU*	*mIoU*
No Adaption	83.7	68.2	60.1	69.1	60.4	62.3	65.0	59.0	58.5	39.1
AdaEnt [31]	84.4	70.5	63.6	70.0	60.7	64.1	71.0	63.8	63.5	48.2
SuperLine [37]	86.6	71.7	65.2	70.1	62.3	65.2	70.1	63.3	64.5	45.7
SHOT [32]	86.3	72.8	63.8	71.5	62.7	64.5	71.8	64.7	62.9	47.1
TT-SFUDA [34]	86.9	73.0	62.8	72.6	64.5	67.3	72.3	64.9	64.7	51.2
DPL [33]	87.5	73.5	63.7	69.9	61.8	66.4	72.0	65.2	63.1	49.9
U-SFAN [11]	87.4	73.8	64.2	72.8	65.1	68.7	73.5	66.2	67.5	52.8
GPLU [35]	88.7	74.1	65.2	73.8	65.3	67.5	72.8	65.4	66.9	54.9
Ours	**90.3**	**76.2**	**68.3**	**74.5**	**66.7**	**71.2**	**75.9**	**69.8**	**76.8**	**65.3**

TABLE II

ABLATION STUDY

CPLV	**GSPM**	**GSGA**	**SAIU**	*SIoU* of different types					*SIoU*	*ACC*	*mIoU*
				Plane	Cone	Cylinder	Sphere	Complex			
				81.4	54.3	83.2	47.9	51.0	68.2	83.7	60.1
✓				85.3	55.6	84.0	48.0	50.9	73.0	87.2	63.8
✓	✓			86.3	62.7	86.9	48.4	50.5	74.7	89.0	65.5
✓	✓	✓		88.9	66.5	87.4	48.3	52.0	76.0	90.1	65.4
✓	✓	✓	✓	88.7	66.8	87.4	47.6	51.1	76.2	90.3	68.3
	✓	✓	✓	87.2	68.3	85.2	47.3	51.0	75.2	88.1	67.3
✓		✓	✓	85.6	59.4	86.3	47.8	51.2	74.3	88.4	68.0
✓	✓		✓	87.3	61.7	87.4	47.9	50.3	74.9	89.7	68.2

inspection is given in the supplementary video.

C. Ablation Study

A set of ablation experiments are conducted to verify roles of the proposed Consistency-based Pseudo-Label Voting strategy (CPLV), the Geometric Saliency-guided Prototype Matching module (GSPM), the Geometric Saliency-guided Graph Aggregation module (GSGA) and the Semantic-aware Instance Relationship Updating strategy (SAIU). The results are reported in Table II, 1) by adding pseudo-label voting strategy, all metrics are improved, indicating that the voting strategy can provide more robust pseudo-labels. 2) The addition of the Geometric Saliency-guided Prototype Matching module significantly improves the performance of regular primitives, as geometric saliency provides more discriminative information. 3) The Graph Aggregation module further improves the segmentation results of complex types, which typically locate in mid-confidence region. 4) Finally, the Semantic-aware Instance Relationship Updating strategy leverages the hierarchical relationships between semantics and instances, leading to a significant enhancement in primitive instance segmentation performance.

We also present t-SNE visualization results of features learned through our method and source model only, as shown in the Fig. 5. Intuitively, the target feature distribution of the source model is messy, with overlapping of different instances in the feature space. Our method helps to separate the different instances in target feature space while making features within the same primitive more compact.

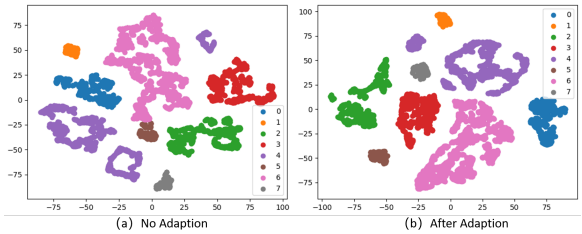

Fig. 5. t-SNE visualizations of primitive segmentation on target domain. Different colors represent features of different primitives.

V. CONCLUSION

This paper introduces a novel SF-UDA method for point cloud primitive segmentation, addressing domain shift and privacy concerns in industrial scenarios. The method utilizes a multi-confidence framework, incorporating transformation consistency, task confidence, and geometric saliency to improve generalization across diverse industrial datasets. This approach shows promise for enhancing efficiency and reliability in industrial automation, with potential for further refinement and real-world applications, advancing point cloud segmentation in practical contexts. Future work will continue to improve the robustness of the domain adaptation.

REFERENCES

[1] E. Ahmed, A. Saint, A. Shabayek *et al.*, "Deep learning advances on different 3d data representations: A survey." *arXiv:1808.01462*, 2018.

[2] D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three-dimensional shapes," *Proc. Roy. Soc. London. Series. B. Biol. Sciences*, vol. 200, no. 1140, pp. 269–294, 1978.

[3] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," in *Computer Graphics Forum*, vol. 26, no. 2, 2007, pp. 214–226.

[4] S. Xia, D. Chen, R. Wang *et al.*, "Geometric primitives in lidar point clouds: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 685–707, 2020.

[5] F. Lafarge and P. Alliez, "Surface reconstruction through point set structuring," in *Computer Graphics Forum*, vol. 32, no. 2pt2, 2013, pp. 225–234.

[6] Z. Kang and Z. Li, "Primitive fitting based on the efficient multibaysac algorithm," *PloS one*, vol. 10, no. 3, pp. 1–21, 2015.

[7] J. Huang, Y. Zhang, and M. Sun, "Primitivenet: Primitive instance segmentation with local primitive embedding under adversarial metric," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 15 343–15 353.

[8] J. Wang, H. Zhu, H. Guo *et al.*, "Few-shot point cloud semantic segmentation via contrastive self-supervision and multi-resolution attention," in *Proc. IEEE Int. Conf. Robot. and Autom. (ICRA)*. IEEE, 2023, pp. 2811–2817.

[9] Y. Fang, P.-T. Yap, W. Lin *et al.*, "Source-free unsupervised domain adaptation: A survey," *arXiv preprint arXiv:2301.00265*, 2022.

[10] C. Saltori, S. Lathuiliére, N. Sebe *et al.*, "Sf-uda 3d: Source-free unsupervised domain adaptation for lidar-based 3d object detection," in *2020 Int. Conf. 3D Vis. (3DV)*. IEEE, 2020, pp. 771–780.

[11] S. Roy, M. Trapp, A. Pilzer *et al.*, "Uncertainty-guided source-free domain adaptation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2022, pp. 537–555.

[12] C. R. Qi, H. Su, K. Mo *et al.*, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2017, pp. 652–660.

[13] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5099–5108, 2017.

[14] Y. Wang, Y. Sun, Z. Liu *et al.*, "Dynamic graph cnn for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, 2019.

[15] P. Guerrero, Y. Kleiman, M. Ovsjanikov, and N. J. Mitra, "Pcpnet learning local shape properties from raw point clouds," in *Computer Graphics Forum*, vol. 37, no. 2, 2018, pp. 75–85.

[16] V. Ganapathi-Subramanian, O. Diamanti, S. Pirk *et al.*, "Parsing geometry using structure-aware shape templates," in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 672–681.

[17] G. Sharma, R. Goyal, D. Liu *et al.*, "Csgnet: Neural shape parser for constructive solid geometry," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2018, pp. 5515–5523.

[18] L. Li, M. Sung, A. Dubrovina *et al.*, "Supervised fitting of geometric primitives to 3d point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 2652–2660.

[19] G. Sharma, D. Liu, S. Maji *et al.*, "Parsenet: A parametric surface fitting network for 3d point clouds," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 261–276.

[20] E.-T. Lê, M. Sung, D. Ceylan *et al.*, "Cpfn: Cascaded primitive fitting networks for high-resolution point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 7457–7466.

[21] S. Yan, Z. Yang, C. Ma *et al.*, "Hpnet: Deep primitive segmentation using hybrid representations," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 2753–2762.

[22] W. Li, S. Xie, W. Min, Y. Jiang, C. Wang, and J. Li, "Spherical coordinate transformation-embedded deep network for primitive instance segmentation of point clouds," *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, p. 102983, 2022.

[23] X. Zhang, H. Zhang, J. Lu *et al.*, "Target-targeted domain adaptation for unsupervised semantic segmentation," in *Proc. IEEE Int. Conf. Robot. and Autom. (ICRA)*, 2021, pp. 13 560–13 566.

[24] Y. Shen, Y. Yang, M. Yan *et al.*, "Domain adaptation on point clouds via geometry-aware implicits," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2022, pp. 7223–7232.

[25] C. Qin, H. You, L. Wang *et al.*, "Pointdan: A multi-scale 3d domain adaption network for point cloud representation," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[26] H. Fan, X. Chang, W. Zhang *et al.*, "Self-supervised global-local structure modeling for point cloud domain adaptation with reliable voted pseudo labels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2022, pp. 6377–6386.

[27] P. Jiang and S. Saripalli, "Lidarnet: A boundary-aware domain adaptation model for point cloud semantic segmentation," in *Proc. IEEE Int. Conf. Robot. and Autom. (ICRA)*, 2021, pp. 2457–2464.

[28] Y. Bian, L. Hui, J. Qian *et al.*, "Unsupervised domain adaptation for point cloud semantic segmentation via graph matching," in *Proc. IEEE Int. Conf. Intell. Robot. and Syst. (IROS)*. IEEE, 2022, pp. 9899–9904.

[29] Z. Yuan, M. Cheng, W. Zeng *et al.*, "Prototype-guided multitask adversarial network for cross-domain lidar point clouds semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023.

[30] Y. Liu, W. Zhang, and J. Wang, "Source-free domain adaptation for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2021, pp. 1215–1224.

[31] M. Bateson, H. Kervadec, J. Dolz *et al.*, "Source-relaxed domain adaptation for image segmentation," in *Int. Conf. Med. Comput. and Comput. Asisted. Intervention.* Springer, 2020, pp. 490–499.

[32] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *Int. Conf. Mach. Learn. (ICML)*. PMLR, 2020, pp. 6028–6039.

[33] C. Chen, Q. Liu, Y. Jin *et al.*, "Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling," in *Int. Conf. Med. Comput. and Comput. Asisted. Intervention.* Springer, 2021, pp. 225–235.

[34] V. VS, J. M. J. Valanarasu, and V. M. Patel, "Target and task specific source-free domain adaptive image segmentation," *arXiv preprint arXiv:2203.15792*, 2022.

[35] M. Litrico, A. Del Bue, and P. Morerio, "Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2023, pp. 7640–7650.

[36] D. Hegde, V. Kilic, V. Sindagi *et al.*, "Source-free unsupervised domain adaptation for 3d object detection in adverse weather," in *Proc. IEEE Int. Conf. Robot. and Autom. (ICRA)*. IEEE, 2023, pp. 6973–6980.

[37] X. Zhao, S. Yang, T. Huang *et al.*, "Superline3d: Self-supervised line segmentation and description for lidar point cloud," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2022, pp. 263–279.

[38] S. Koch, A. Matveev, Z. Jiang *et al.*, "Abc: A big cad model dataset for geometric deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 9601–9611.