PCoTTA: Continual Test-Time Adaptation for Multi-Task Point Cloud Understanding

Jincen Jiang*
Bournemouth University
jiangj@bournemouth.ac.uk

Qianyu Zhou*
Shanghai Jiao Tong University
zhouqianyu@sjtu.edu.cn

Yuhang Li
Shanghai University
yuhangli@shu.edu.cn

Xinkui Zhao†
Zhejiang University
zhaoxinkui@zju.edu.cn

Meili Wang
Northwest A&F University
wml@nwsuaf.edu.cn

Lizhuang Ma
Shanghai Jiao Tong University
lzma@sjtu.edu.cn

Jian Chang
Bournemouth University
jchang@bournemouth.ac.uk

Jian Jun Zhang
Bournemouth University
jzhang@bournemouth.ac.uk

Xuequan Lu†
La Trobe University
b.lu@latrobe.edu.au

Abstract

In this paper, we present PCoTTA, an innovative, pioneering framework for Continual Test-Time Adaptation (CoTTA) in multi-task point cloud understanding, enhancing the model's transferability towards the continually changing target domain. We introduce a multi-task setting for PCoTTA, which is practical and realistic, handling multiple tasks within one unified model during the continual adaptation. Our PCoTTA involves three key components: automatic prototype mixture (APM), Gaussian Splatted feature shifting (GSFS), and contrastive prototype repulsion (CPR). Firstly, APM is designed to automatically mix the source prototypes with the learnable prototypes with a similarity balancing factor, avoiding catastrophic forgetting. Then, GSFS dynamically shifts the testing sample toward the source domain, mitigating error accumulation in an online manner. In addition, CPR is proposed to pull the nearest learnable prototype close to the testing feature and push it away from other prototypes, making each prototype distinguishable during the adaptation. Experimental comparisons lead to a new benchmark, demonstrating PCoTTA's superiority in boosting the model's transferability towards the continually changing target domain. *Our source code is available at*: https://github.com/Jinec98/PCoTTA.

1 Introduction

Recent advancements in 3D point cloud understanding have marked a significant leap in the field of computer vision [20, 38, 52, 41] and 3D processing [9, 8, 12, 29]. Current methods [34, 47] primarily concentrate on training and testing on a single domain [33, 19]. Nevertheless, they encounter noticeable performance drops on other target data. Different datasets have domain gaps, also known as domain shifts. For instance, models trained on meticulously structured synthetic data, such as ModelNet40 [51], may encounter difficulties in adapting to intricate and noisy real-world data, such as ScanObjectNN [42].

*Equal contributions.
†Corresponding authors.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

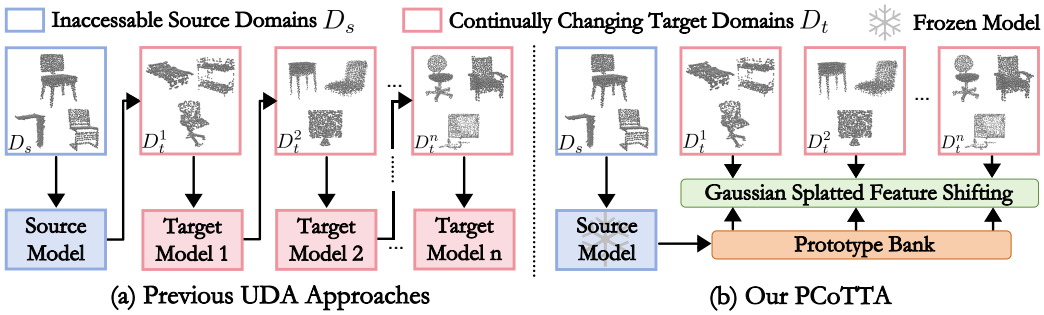

Figure 1: (a) Previous UDA approaches on point cloud suffer from catastrophic forgetting and error accumulation toward the continually changing target domains. (b) In contrast, we present an innovative framework PCoTTA to address these issues, enhancing the model's transferability.

To mitigate domain shifts, recent researchers have introduced Unsupervised Domain Adaptation (UDA) techniques [64, 27, 63, 65, 62] into point cloud understanding. Some studies synthesize diverse training data [56, 44, 61], and others leverage adversarial learning [36, 60, 26], pseudo labeling [53, 50, 54, 18, 37], consistency learning [44, 48, 30, 49], feature disentanglement [22] or self-supervised learning [1, 39, 66, 24] to align the latent features across different domains. Nonetheless, these methods still face challenges especially when the target domain is streaming online and the whole training set of the target domain is inaccessible. As such, Test-Time Adaptation is introduced into point cloud [21, 16, 17] where the model can adapt to target distributions in an online manner at test-time without requiring any prior knowledge of the whole target domain. However, these methods may still fail when the target domain is continually changing, referred to as Continual Test-Time Adaptation (CoTTA), and such an open problem is rarely explored in point cloud understanding contexts.

On the one hand, due to the lack of specific designs for 3D data, current CoTTA methods [45, 10, 40, 31, 2, 46, 13, 32, 58] that are designed for 2D images are inapplicable to 3D point cloud tasks or exhibit less desired performance. On the other hand, few works like MM-CCTA [3] target the CoTTA problem in 3D point cloud tasks. Although MM-CTTA [3] designs a Continual Cross-Modal Adaptive Clustering (CoMAC) approach for 3D semantic segmentation, it suffers from two primary limitations: (1) it is specifically designed for one task only, and cannot handle other point cloud tasks such as point cloud reconstruction, denoising, and registration. Redesigning and retraining a CoTTA method for each task is cost-expensive. (2) The adapted model would inevitably forget the previously learned data (catastrophic forgetting) and accumulate the model errors (error accumulation) during the continual adaptation, limiting the model's transferability toward the target domains.

Motivated by the above analysis, we present PCoTTA, an innovative, pioneering framework for Continual Test-Time Adaptation (CoTTA) in multi-task point cloud understanding, enhancing the model's transferability towards the continually changing target domain. Also, we introduce a multi-task setting for PCoTTA, which is practical and realistic, handling multiple tasks within one unified model during the adaptation. In particular, given an off-the-shelf model pre-trained on the source domains, our PCoTTA aims to bridge the gap between the source and continually changing target domains by dynamically scheduling the shifting amplitude at test time.

Our PCoTTA mainly consists of three novel modules. Firstly, to prevent catastrophic forgetting, we propose an automatic prototype mixture (APM) strategy that automatically mixes the source prototypes with the learnable target prototypes based on the automatic similarity balancing factor (ASBF), which avoids straying too far from its original source model. Secondly, to mitigate error accumulation, we present Gaussian Splatted feature shifting (GSFS) that dynamically shifts the testing sample toward the source domain based on the distance between the testing features and the shared prototype bank. In addition, we also introduce Gaussian weighted graph attention to further adaptively schedule the shifting amplitude in a learnable manner at test time. Our insight is to highlight the similarity between the target sample and its similar prototypes and suppress the dissimilar weights. It therefore mitigates the risk of catastrophic forgetting. Finally, we devise the contrastive prototype repulsion (CPR) to pull the nearest learnable prototype close to the testing feature and push it away from other prototypes, making learnable prototypes more distinguishable. Furthermore,

we present a new benchmark. We meticulously select a total of $30,954$ point cloud samples from 4 datasets, including 2 synthetic datasets (ModelNet40 [51] and ShapeNet [5]) and 2 real-world datasets (ScanNet [7] and ScanObjectNN [42]), encompassing 7 same object categories, and generate corresponding ground truth for 3 different tasks (reconstruction, denoising, and registration). Our main contributions are three-fold:

- We present PCoTTA, an innovative, pioneering, and unified framework for Continual Test-Time Adaptation (CoTTA) in multi-task point cloud understanding, enhancing the model's transferability towards the continually changing target domain. We introduce a multi-task setting with a new benchmark for PCoTTA, which is practical and realistic in the real world.
- We devise three innovative modules for PCoTTA, *i.e.,* automatic prototype mixture (APM), Gaussian Splatted feature shifting (GSFS), and contrastive prototype repulsion (CPR) strategies, where APM avoids straying too far from its original source model, mitigating the risk of catastrophic forgetting, and GSFS dynamically shifts the testing sample toward the source model, alleviating error accumulation, and CPR pulls the nearest learnable prototype close to the testing feature and pushes it away from other prototypes.
- Extensive experimental results with analysis demonstrate the effectiveness and superiority of our presented method, surpassing the state-of-the-art approaches by a large margin.

2 Related Work

Point Cloud Understanding. Pioneering works such as PointNet [34] and PointNet++ [35] process point clouds directly, with PointNet [34] utilizing pooling operations for spatial encodings and PointNet++ [35] employing hierarchical processing for capturing local structures at various scales. DGCNN [47] updates the graph in feature space to capture dynamic local semantic features, while PCT [15] addresses global context and dependencies within point clouds using order-invariant attention mechanisms. Recent methods like Point-BERT [57] and Point-MAE [33] have introduced Masked Point Modeling (MPM) for reconstructing obscured point clouds. Point-BERT [57] employs a BERT-style pre-training strategy for improving performance in subsequent tasks, while Point-MAE [33] uses masked autoencoders for self-supervised learning, enabling comprehensive representations without labeled data. PIC [11] explores the In-Context Learning (ICL) paradigm to enhance 3D point cloud understanding, showcasing the model's potential in multi-task learning. Despite their gratifying progress, they only consider a single data domain and suffer from performance degradation in target domains. Thus, we study continual test-time adaptation for point cloud tasks.

Continual Test-Time Adaptation. This task aims to adapt the pre-trained model toward the continually changing environments at test time. CoTTA [45] employs a weighted augmentation-averaged mean teacher framework to address this issue. [14] capitalizes on the temporal correlations within streamed input data through reservoir sampling and instance-aware batch normalization. [13, 55] introduce domain-specific prompts and domain-agnostic prompts to preserve both domain-specific and domain-shared knowledge, respectively. Meanwhile, EATA [32] focuses on adapting non-redundant samples to facilitate efficient updates. Another work RMT [10] uses a mean teacher setup with symmetric cross-entropy and contrastive learning. More recently, MM-CTTA [3] designs a Continual Cross-Modal Adaptive Clustering (CoMAC) approach for 3D semantic segmentation. Despite these methods showing promising potential in 3D data, they mainly suffer from two limitations: Firstly, they are specifically designed for one task only, and they cannot handle other point cloud tasks like those in PIC [11]. Secondly, the model would inevitably forget the previously learned knowledge (catastrophic forgetting) and accumulate prediction errors (error accumulation) during the continual adaptation, leading to undesirable results. In contrast, we present a unified model, PCoTTA, for continual test-time adaptation of multi-task point cloud understanding.

3 Method

We present a novel framework, namely PCoTTA, for Continual Test-Time Adaptation in point cloud understanding tasks with the practical multi-task and multi-domain setting. As depicted in Figure 2, we propose an innovative approach to effectively address the challenges of continuously changing target data in test time within a unified model. In particular, our PCoTTA consists of three novel components: Automatic Prototype Mixture (APM) to mitigate catastrophic forgetting, Gaussian

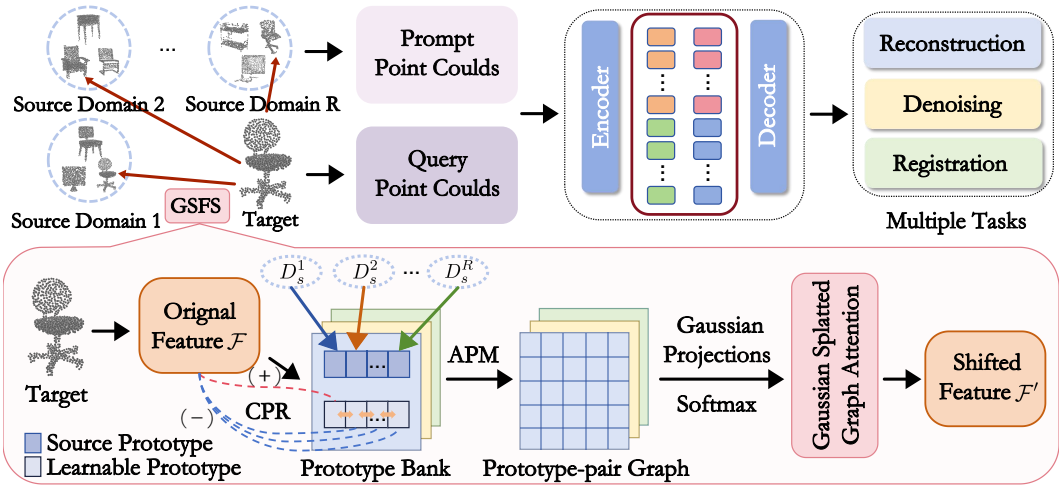

Figure 2: Our PCoTTA. It addresses continually changing targets by using their nearest source sample as a prompt for multi-task learning within a unified model. We introduce Gaussian Splatted Feature Shifting (GSFS) to align unknown targets with sources, improving transferability. Source prototypes from different domains and learnable prototypes form a prototype bank. The Automatic Prototype Mixture (APM) pairs these prototypes based on the similarity to the target, preventing catastrophic forgetting. We project these prototypes as Gaussian distributions onto the feature plane, with larger weights assigned to more relevant ones. Our graph attention updates these weights dynamically to mitigate error accumulation. Additionally, our Contrastive Prototype Repulsion (CPR) ensures that learnable prototypes are distinguishable for different targets, enhancing adaptability.

Splatted Feature Shifting (GSFS) to alleviate error accumulation, and Contrastive Prototype Repulsion (CPR) to make learnable prototypes distinctive across continually changing target domains.

3.1 Point Cloud Continual Test-Time Adaptation

Problem Formulation. In this work, we study a practical setting of continual test-time adaptation for multi-task point cloud understanding. Suppose we have R source domains $D_s = \{D_s^1, D_s^2, \ldots, D_s^R\}$, our PCoTTA employs the input point clouds $\{I_q, I_p\}$ (along with their targets $\{T_q^k, T_p^k\}$, where k represents the task index) from two different sources $\{D_s^i, D_s^j\} \in D_s, (i \neq j)$ to form the context pairs, facilitating the model with a comprehensive representation that effectively generalizes across all source domains. In the pre-training phase, each input sample comprises two context pairs: the input point cloud pair (query and prompt) and their corresponding target pair addressing the same task. During the test time, our PCoTTA strives to align streamed target data $I_t \in D_t$ (where $D_t = \{D_t^1 \cup D_t^2 \cup \ldots\}$ denotes the set of continuously varying target domains) towards sources that possess correlative features to the off-the-shelf pre-trained model.

Multi-task Learning Objective. We follow PIC [11] for three point cloud understanding tasks: (1) Reconstruction, which focuses on generating a dense point cloud from the sparse input; (2) Denoising, aiming at eliminating noise or outliers from the input point cloud; (3) Registration, dedicated to restoring the original orientation of a randomly rotated point cloud. Please note these three tasks might be slightly different from conventional definitions. They are used as they can be handled similarly given current point learning can predict point positions directly. This makes them 'unified' with position output and a single loss. We employ the MPM framework to generate query results across multiple downstream tasks, with a unified objective and a unified model. Let $\Phi(\cdot)$ denote the model shared across all domains and all tasks, and predicted masked patches P can be depicted as:

$$\{I_q, I_p\} \rightarrow P = \Phi(\mathcal{F}(I_q) \oplus \mathcal{F}(T_q^k) \oplus \mathcal{F}(I_p) \oplus \mathcal{F}(T_p^k), \mathcal{M}), \tag{1}$$

where $\mathcal{F}(\cdot)$ represents the feature encoder that produces patch-wise features, *i.e.,* the tokens, from point cloud feature, and $\mathcal{M}$ denotes the masked token utilized to replace the masked patches in the inputs. During the pre-training stage, $\mathcal{M}$ is derived from the random masking among query and prompt point clouds; whereas at test time, $\mathcal{M}$ exclusively masks the query target to generate the

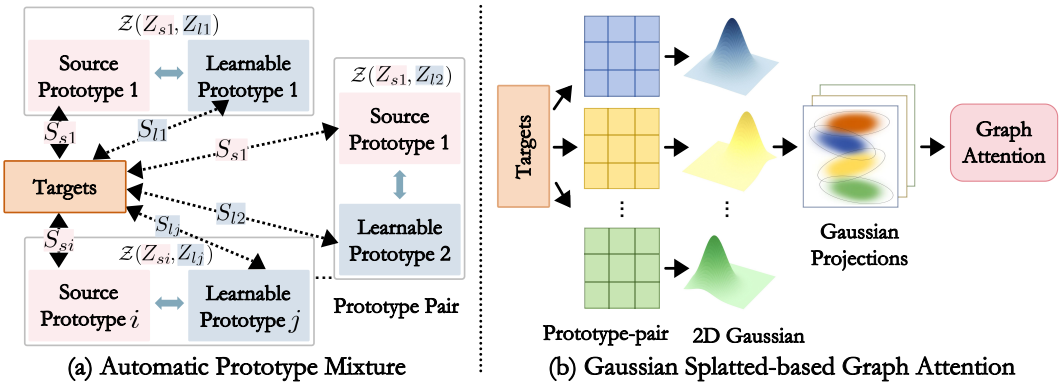

Figure 3: (a) Automatic Prototype Mixture (APM) considers both source and learnable prototypes with their similarities to the target, mitigating catastrophic forgetting by preserving source information. (b) Gaussian Spaltted-based Graph Attention enables dynamic updating weights among all prototype-pair nodes based on the Gaussian projections splatted onto the feature plane.

task-specific query output. The Chamfer Distance (CD) is used as the loss, measuring the similarity between the predicted masked patch P and its corresponding ground truth G:

$$\mathcal{L}_{cd} = \frac{1}{|P|}\sum_{x\in P}\min_{y\in G}\|x-y\|_2^2 + \frac{1}{|G|}\sum_{y\in G}\min_{x\in P}\|y-x\|_2^2. \tag{2}$$

3.2 Automatic Prototype Mixture

The empirical evidence perceived by the human visual system illustrates that when people are not certain about the identity of an object, they would seek to find a distinct object from other domains that share high semantic similarity with the current object in the target domain. Motivated by this, we propose Automatic Prototype Mixture (APM) that adapts to continuously changing target data by aligning it with model-familiarized prototypes of source domains at test time.

Source Prototypes Estimation. Our insight lies in that source prototypes can potentially represent source domains' feature distribution. Pulling the target data toward source prototypes within the feature space can effectively narrow the domain gap, bolstering the model's transferability. Accordingly, the source prototypes $Z_s^i (i \in [1, R])$ can be determined by computing the average of all tokens produced by the MPM framework across all data within the sources:

$$Z_s^i = \frac{1}{N_{D_s^i}}\sum_{n=1}^{N_{D_s^i}}\mathcal{F}(I_n), \quad Z_s \in \mathbb{R}^{R\times K\times M\times C}, \tag{3}$$

where $N_{D_s^i}$ denotes the sample number in domain D_s^i, K represents the tasks number, and M indicates the tokens number in each sample. After pre-training on the multi-task and multi-domain setting, we save all source prototypes Z_s derived from the model at the last epoch, considering them as the shared common knowledge available to the target data during the test time.

Prototype Bank. We propose a novel prototype bank that stores not only the source prototypes Z_s but also a series of *learnable prototypes* $Z_l \in \mathbb{R}^{S\times K\times M\times C}$, where S indicates the number of all potential target domains D_t. The learnable prototypes Z_l aim to extract the current domain knowledge, thereby paving the way for handling subsequent unknown test data. We achieve the test-time adaptation of target tokens through the mixture of the paired prototypes in the prototype bank, selectively updating only the learnable prototypes while maintaining the source ones, thus mitigating the risk of catastrophic forgetting of the source domain knowledge due to the over-reliance on the adaptively learned information.

Prototype-pair Node Mixture. The source prototypes Z_s along with the learnable prototypes Z_l in the prototype bank are paired to form prototype-pair nodes. As illustrated in Figure 3(a), the tokens from each test data $\mathcal{F}(I_t)$ serve as the central node in a graph structure, adjacent to all prototype-pair nodes. We propose the Automatic Prototype Mixture (APM) module, designed to merge source and

learnable prototypes within each node by considering their token-wise feature distances with the test data, *i.e.,* the dot product between two feature vectors.

Firstly, we need to repeat the test data tokens $\mathcal{F}(I_t)$ to align with the total number of prototypes:

$$\mathcal{R}(I_t) = [\underbrace{\mathcal{F}(I_t)\ \mathcal{F}(I_t) \dots \mathcal{F}(I_t)}_{\text{repeat } x \text{ times}}], \tag{4}$$

where x equals R or S. Then, the similarity $\mathcal{S}_s$ between source prototypes Z_s and the test data I_t is:

$$\mathcal{S}_s = \frac{1}{M} \sum Diag(Norm(Z_s) \cdot Norm(\mathcal{R}(I_t)^T)) \in \mathbb{R}^{R \times K}, \tag{5}$$

where $Diag(\cdot)$ indicates creating a diagonal matrix, $Norm(\cdot)$ denotes normalization along the last dimension (*i.e.,* the feature channel), and $(\cdot)^T$ represents transposition specifically applied to the last two dimensions. Likewise, the similarity $\mathcal{S}_l \in \mathbb{R}^{S \times K}$ between the test data and the learnable prototypes can also be determined.

We further propose the Automatic Similarity Balancing Factor (ASBF) to measure the impact of the source and learnable prototypes toward the test data through the similarities $\mathcal{S}_s$ and $\mathcal{S}_l$, automatically prioritizing the prototypes and assigning greater weight to more similar components. The mixed prototypes (*i.e.,* the prototype-pair nodes) Z_m can be defined as:

$$Z_m = \mathcal{Z}(Z_s, Z_l) = \frac{\mathcal{S}_s}{\mathcal{S}_s + \mathcal{S}_l} \cdot Z_s + \frac{\mathcal{S}_l}{\mathcal{S}_s + \mathcal{S}_l} \cdot Z_l \in \mathbb{R}^{R \times S \times K}. \tag{6}$$

APM effectively considers the two types of prototypes while ensuring that the engagement with the original pre-trained model and source prototype is maintained, preventing catastrophic forgetting.

3.3 Gaussian Splatted Feature Shifting

Our PCoTTA considers all nodes but applies dynamically updated weights to each edge, enabling distinguishing the feature shifting in the continual test-time adaptation. To this end, we propose the Gaussian Splatted Feature Shifting (GSFS), preventing error accumulation in an online manner.

Gaussian Splatted-based Graph Attention. Our key insight is that prototypes within a node (*i.e.,* the source-learnable prototypes pair) can mutually constrain each other and collaboratively determine the weight of the edge connected to this node. We interpret the similarities between the test data and these two types of prototypes as Gaussian projections onto a plane, with the source and learnable prototypes corresponding to two orthogonal axes, respectively. As illustrated in Figure 3(b), the projections of all nodes on the feature plane are treated as a blend of Gaussians, where nodes with stronger correlations to the test data (*i.e.,* higher similarities) are assigned larger weights. In this manner, all prototype-pair nodes are seamlessly integrated into the feature adaptation process. We compute the attention coefficient of each node as follows:

$$\mathcal{E}(\mathcal{S}_s, \mathcal{S}_l) = \omega - \mathcal{G}(\mathcal{S}_s, \mathcal{S}_l) = \omega - \frac{1}{2\pi\sigma_{\mathcal{S}_s}\sigma_{\mathcal{S}_l}} e^{-\frac{1}{2}\left(\frac{(\mathcal{S}_s - \mu_{\mathcal{S}_s})^2}{\sigma^2_{\mathcal{S}_s}} + \frac{(\mathcal{S}_l - \mu_{\mathcal{S}_l})^2}{\sigma^2_{\mathcal{S}_l}}\right)}, \tag{7}$$

where $\sigma_{\mathcal{S}_s}, \sigma_{\mathcal{S}_l}$ represent the variances of $\mathcal{S}_s$ and $\mathcal{S}_l$, respectively, and $\mu_{\mathcal{S}_s}, \mu_{\mathcal{S}_l}$ denote their mean values. Note that the Gaussian function is inversely correlated with the similarity. We introduce a parameter ω, set slightly above the maximum similarity observed, to ensure that more similar prototypes have a stronger influence.

Attention-based Feature Shifting. The attention coefficient $\mathcal{E}^{i,j}$ reflects the relative importance of the source Z_s^i and the learned Z_l^j prototypes. To ensure comparability across all connected nodes, we normalize coefficients using the Softmax function. Furthermore, we adopt a learnable shared attention module to dynamically update edge weights as follows:

$$\mathcal{W}_{i,j} = \Psi_\theta(Softmax(\mathcal{E}^{i,j})) = \Psi_\theta(\frac{e^{\mathcal{E}^{i,j}}}{\sum_{m=1}^{N_{Z_m}} e^{\mathcal{E}^{i,m}}}), \tag{8}$$

where Ψ_θ denotes a series of Convolution Layers parameterized by θ, and N_{Z_m} indicates the total number of the mixed prototypes (equals $R \times S$), indexed by m. Thereby, we can merge all prototype-pair nodes with the central node, *i.e.,* the test data features, using the adaptive edge weights:

$$\mathcal{F}'(I_t) = \frac{1}{R \times S} \sum_{i,j=0}^{R,S} ((1 - \mathcal{W}_{i,j}) \cdot \mathcal{F}(I_t) + \mathcal{W}_{i,j} \cdot Z_m^{i,j}). \tag{9}$$

The proposed GSFS dynamically updates the contribution from each node in the graph with Gaussian Splatted-based graph attention, effectively assigning distinctive weights of feature shifting according to each node's relevance to the test data. This enables the test data to effectively align with task-beneficial domains, significantly diminishing the potential for error accumulation in the model.

3.4 Contrastive Prototype Repulsion

The learnable prototypes within the prototype bank strive to capture the domain-specific knowledge of the current test data. Instead of predicting domain pseudo-labels to all test data, a common practice in prior techniques [45, 13], our method pulls the most similar learnable prototype closer to the test data while pushing it away from the others, thereby implicitly learning the distinctive features from different samples. To this end, we introduce Contrastive Prototype Repulsion (CPR) that effectively refines the learnable prototypes in the prototype bank, ensuring their distinctiveness and preventing domain-flattening from iterative learning and settling at sub-optimal points. We form a positive pair between the test data $\mathcal{F}'(I_t)$ and their nearest learnable prototype Z_l^t, and the rest serve as negative pairs. Our CPR optimization objective can be expressed as:

$$\mathcal{L}_{pr} = -\frac{1}{S}\sum_{Z_i \in S} \log\left(\frac{e^{\mathcal{F}'(I_t)\cdot Z_l^t/\tau}}{e^{\mathcal{F}'(I_t)\cdot Z_l^t/\tau} + \sum_{k\neq t} e^{\mathcal{F}'(I_t)\cdot Z_l^k/\tau}}\right), \tag{10}$$

where τ is the temperature parameter, set to 0.07 by default. Therefore, the overall loss function of our PCoTTA in the test time adaptation can be defined as follows:

$$\mathcal{L} = \mathcal{L}_{cd} + \alpha \cdot \mathcal{L}_{pr}, \tag{11}$$

where α is the weighting factor that balances the two loss terms.

4 Experiments

4.1 Experimental Setting

Implementation Details. We implement our method using PyTorch and perform experiments on two NVIDIA A40 GPUs. Following PIC [11], we set the training batch size to 128 and utilize the AdamW optimizer [28]. The learning rate is set to 0.001, with a cosine learning scheduler and a weight decay of 0.05. All models are trained for 300 epochs during the pertaining stage, and we train the pre-trained model for 3 epochs on the source domains to initialize our prototype bank. At testing time, we continuously adapt test samples to the source pre-trained model and validate the anti-forgetting capability of our method across multiple rounds. Each point cloud is sampled to $1,024$ points and then split into 64 patches, with each patch consisting of 32 points. Within the MPM framework, the mask ratio is set to 0.7, consistent with prior studies [57, 33].

New Benchmark. We meticulously curate and select data from 4 distinct datasets (2 synthetic and 2 real-world datasets), containing 7 identical object categories. Subsequently, we generate corresponding ground truth based on 3 different tasks. The synthetic datasets include ModelNet40 [51] and ShapeNet [5]. ModelNet40 consists of $3,713$ samples for training and 686 for testing, while ShapeNet comprises $15,001$ training samples and $2,145$ testing samples. We also consider real-world data: ScanNet [7] and ScanObjectNN [42]. ScanNet provides annotations for individual objects in real 3D scans, and we choose $5,763$ samples for training and $1,677$ for testing. ScanObjectNN includes $1,577$ training samples and 392 testing samples. In all experiments, we employ ScanNet [7] and ShapeNet [5] as the source domains and evaluate the transferability of our method on the other two target domains, *i.e.,* ModelNet40 [51] and ScanObjectNN [42] with 3 repeated times by default.

4.2 Main Results

Table 1 shows the comparison results of our PCoTTA against other methods across tasks of reconstruction, denoising, and registration in the introduced setting. Our method consistently outperforms others by a large margin, demonstrating superior adaptability in a multi-domain multi-task setting. Conventional methods such as PointNet [34], DGCNN [47], and PCT [15] often struggle with unseen data, leading to significant performance drops. Augmentation-based methods like Pointmixup [6] and PointCutMix [59], though adapted for multi-domain learning, exhibit limited performance in

Table 1: Comparisons with the state-of-the-art approaches on the CoTTA setting. We report the Chamfer Distance (CD, $\times 10^{-3}$) for different tasks. The lower CD denotes the better performance.

Time		t⟶																	
Rounds		1						2						3					
Target Domains		ModelNet40			ScanObjectNN			ModelNet40			ScanObjectNN			ModelNet40			ScanObjectNN		
Methods	Setting	Rec.	Den.	Reg.	Rec.	Den.	Reg.	Rec.	Den.	Reg.	Rec.	Den.	Reg.	Rec.	Den.	Reg.	Rec.	Den.	Reg.
PointNet [34]	Task-specific Models	38.2	38.1	40.4	39.3	39.5	41.5	37.7	38.4	40.7	39.0	39.8	42.0	38.2	38.1	40.9	39.2	39.5	42.2
DGCNN [47]		36.0	33.7	36.0	37.3	35.6	37.6	35.3	32.7	34.1	36.6	34.6	36.0	36.1	32.6	34.7	37.1	34.4	36.5
PCT [15]		29.7	29.6	30.6	30.2	30.3	31.5	29.6	29.8	30.6	30.2	30.5	31.8	30.8	29.5	30.8	31.5	30.1	31.8
Pointmixup [6]		37.3	36.8	38.5	38.4	37.0	40.3	37.0	36.5	37.9	38.9	36.7	40.1	37.8	36.8	38.1	38.5	36.9	40.7
PointCutMix [59]		41.5	40.1	38.2	43.3	44.7	40.5	41.1	40.4	38.5	42.9	44.1	40.7	40.8	40.0	39.2	43.1	44.5	40.2
PointNet [34]	Multi-task Models	38.3	38.8	41.4	39.5	40.4	43.0	38.0	38.5	41.3	39.3	40.2	42.8	38.4	38.6	42.1	39.6	40.4	43.3
DGCNN [47]		37.0	33.5	36.0	38.1	35.2	37.7	36.9	33.2	36.0	38.1	35.5	37.7	36.9	33.2	36.5	38.0	35.2	37.8
PCT [15]		29.6	30.2	32.5	30.4	30.9	33.7	29.9	30.4	32.4	30.7	30.9	33.5	29.8	30.0	31.9	30.5	30.8	33.1
Pointmixup [6]		37.8	41.5	39.2	44.6	45.1	40.7	38.3	40.9	39.1	43.4	44.2	41.6	38.2	41.3	39.2	44.1	44.8	40.9
PointCutMix [59]		42.3	44.1	39.9	45.4	47.3	43.8	41.9	43.2	40.1	45.2	46.8	42.3	42.1	43.7	40.4	45.2	47.1	42.9
Baseline [11]	ICL Models	79.7	126.3	106.3	82.3	129.5	113.4	86.5	127.7	106.8	83.0	124.7	110.2	78.9	123.6	110.6	84.5	125.9	112.4
PIC [11]		69.2	64.7	58.4	72.5	77.4	62.8	72.0	65.4	60.3	71.8	79.5	60.3	70.2	60.8	54.9	71.8	78.3	60.6
AdaBN [23]	CoTTA Models	58.7	52.1	37.7	64.1	76.8	57.2	58.9	51.5	37.2	64.1	74.2	53.9	56.8	50.3	35.5	62.1	71.7	51.1
TENT [43]		57.9	50.6	36.8	64.8	76.4	55.0	57.8	50.0	36.7	64.7	73.5	51.1	55.2	48.4	35.0	62.1	69.2	49.7
CoTTA [45]		58.3	50.1	36.4	62.5	73.6	50.3	56.7	49.0	34.4	60.4	71.8	49.1	55.2	46.9	34.3	59.6	66.3	48.5
ViDA [25]		52.4	47.2	35.1	58.2	69.8	47.5	51.6	46.9	34.3	57.6	67.2	45.5	51.3	46.2	32.8	54.4	63.1	42.8
RMT [10]		31.2	44.0	34.3	47.4	59.6	39.9	30.6	43.5	33.9	45.6	53.0	35.8	30.4	42.7	33.8	45.9	51.1	36.4
SANTA [4]		32.3	42.1	37.8	44.9	55.2	38.6	31.7	41.9	37.4	42.0	53.4	35.6	30.1	41.6	36.4	40.6	52.9	34.7
Our **PCoTTA**		**6.3**	**21.4**	**15.4**	**8.9**	**28.3**	**20.7**	**5.5**	**19.9**	**14.6**	**8.5**	**26.9**	**19.6**	**5.4**	**18.6**	**12.1**	**8.2**	**25.2**	**19.3**

multi-task generalization. Despite incorporating task-specific heads, these methods still fall short compared to our unified model, which excels across all tasks due to our Automatic Prototype Mixture (APM) and Gaussian Splatted Feature Shifting (GSFS) modules. While PIC [11] performs well in multi-task scenarios, its transferability is limited, often failing with changing target data. Our PCoTTA addresses this by aligning the target data with source prototypes and dynamically updating learnable prototypes at test time, effectively narrowing the domain gap. Our method demonstrates strong continuous online learning abilities, improving results across multiple validations and showing resilience against catastrophic forgetting and error accumulation.

We compare our PCoTTA with advanced CoTTA methods like AdaBN [23], TENT [43], CoTTA [45], ViDA [25], RMT [10], and SANTA [4]. To ensure fairness, we also update the LayerNorm parameters for AdaBN, TENT, and SANTA. While these methods handle continuously changing targets, they struggle with multi-task aspect in our challenging setting. Even when equipped with multi-task capabilities, these methods still underperform compared to our PCoTTA. Our success is attributed to three main factors: (1) Usually, these methods heavily rely on the student-teacher architecture to realize consistency regularization. As a result, they would inevitably introduce pseudo-label noise, leading to error accumulation. Although they use symmetric cross-entropy or other techniques to alleviate the pseudo-label noise, such problems still exist and cannot be fundamentally addressed. In contrast, our PCoTTA framework does not use any online or offline pseudo-labeling techniques, which inherently avoids the risk of error accumulation. (2) These methods are specifically designed for CoTTA in 2D images and perform well on 2D images. However, compared to 2D images, 3D point cloud data is disordered, unstructured, and sparsely distributed, making these 2D image-based CoTTA methods less effective or even inapplicable. Our method involves specific designs for 3D point cloud data, *e.g.,* Gaussian Splatted-based Graph Attention for comprehensive, patch similarity-based adaptation, well-suited for 3D data, and achieves better performances than these methods. (3) These methods often focus on single tasks and all lack specialized design in multi-task learning, which may lead to gradient conflicts in the optimization process of continual test-time adaptation. Instead, our PCoTTA devises task-specific prototype banks where individual source-learnable prototype pairs are used for different adaptations in each task, thus favoring the multi-task learning in our setting.

4.3 Ablation Studies

Effect of Each Component. Table 2 shows the effects of different components. Compared with the baseline, Model A simply shifts target features by equally fusing with every source-learnable prototype pair (APM), demonstrating that our prototype bank effectively enriches the source information for targets, thereby improving the model's transferability. By adding GSFS, we achieve better performance. This is because Model B uses the similarity between targets and prototypes as weights

during aggregation, and meanwhile, our attention mechanism in GSFS also enables dynamic updating of these weights, offering greater weights to prototypes closer to the current sample. Finally, adding CPR (Ours) enables the prototype bank's learnable prototypes to be more distinct, achieving the best performance. These improvements confirm that these individual components are complementary and together they significantly promote the performance.

Quantity of Learnable Prototypes. We conducted an additional ablation study on the number of learnable prototypes, as shown in Table 3, and the results indicate marginal changes. Additionally, we show the case with no learnable prototypes (*i.e.,* quantity 0), where our method degrades to aligning the target feature by solely considering source prototypes' similarities. While this case achieves some degree of test-time adaptation, its performance is less decent than our PCoTTA.

Table 2: Ablation studies on our proposed three modules. We report the CD ($\times 10^{-3}$) for three different tasks on ModelNet40.

Models	APM	GSFS	CPR	Rec.	Den.	Reg.
Baseline				78.9	123.6	110.6
A	✓			23.5	37.4	30.1
B	✓	✓		14.6	31.2	27.2
Ours	✓	✓	✓	**5.4**	**18.6**	**12.1**

Table 3: Ablation studies on the quantity of learnable prototypes.

Models	Quantity	Rec.	Den.	Reg.
I	0	15.5	32.4	30.7
II	1	8.2	24.9	17.4
Ours	**2**	**5.4**	**18.6**	**12.1**
III	3	6.8	19.5	14.8
IV	4	6.5	20.3	14.0

Cross Validation. Table 4 shows our model's effectiveness in bridging the domain gap from the synthetic to real scan data. Consistently, our method surpasses CoTTA [45] in all tasks, demonstrating the superiority of our method. Remarkably, our model also performs better than CoTTA [45] when pre-trained on the two real scan datasets which involve background interference and missing parts. This underscores our method's strong transferability between various domains.

Efficiency Analysis. We present an analysis of model parameters and running time in Table 5. The results show that our method achieves fast inference on target data, and our model has the fewest parameters compared to other CTTA methods. As such, this shows potential for many real-world applications, *e.g.,* autonomous driving and virtual reality, since our PCoTTA is an end-to-end test-time adaptation method without relying on a teacher-student model or pseudo-labeling technique, it is more efficient and suitable for real-time deployment.

Table 4: Cross validation with synthetic data: ShapeNet (SP), ModelNet40 (MN), and real scan data: ScanNet (SN), ScanObjectNN (SO).

Methods	Sources → Targets	Rec.	Den.	Reg.
CoTTA [45]	SP + MN → SN + SO	63.6	70.4	57.2
Ours		**12.7**	**30.7**	**23.2**
CoTTA [45]	SN + SO → SP + MN	58.8	50.3	39.6
Ours		**10.4**	**26.1**	**17.4**

Table 5: Comparison of model efficiency. We report the Runtime (s), Flops (G), and Parameters (M) as metrics.

Methods	Run.	Flop.	Para.
CoTTA [45]	4.96	24.26	86.72
ViDA [25]	5.14	19.99	100.98
Ours	**0.06**	**12.11**	**28.91**

4.4 Visualization and Analysis

Visualization of Different Tasks. Figure 4 illustrates the qualitative results in the last round of our PCoTTA model. From the figure, we have two observations. Firstly, our proposed PCoTTA manages to generate quality predictions in the continually changing target domain by leveraging the proposed distinctive prototype bank, minimizing the discrepancies between source and target domains. Secondly, without retraining a CoTTA method for each task, our proposed PCoTTA is able to successfully handle multiple tasks such as point cloud reconstruction, denoising, and registration and multiple domains with a unified model, demonstrating strong practicability and transferability in the real world. We provide more visual comparisons with state-of-the-art methods in Appendix A.4.

T-SNE Feature Visualization. To understand how our PCoTTA aligns the domains, we visualize the feature distributions of the source and target domains via t-SNE. We display the latent features of the point cloud reconstruction task in Figure 5. From the figure, we make the following observations: The baseline model means directly deploying the source pre-trained model in the continually changing domains, resulting in an unsatisfactory alignment. Although CoTTA [45] aligns the source and

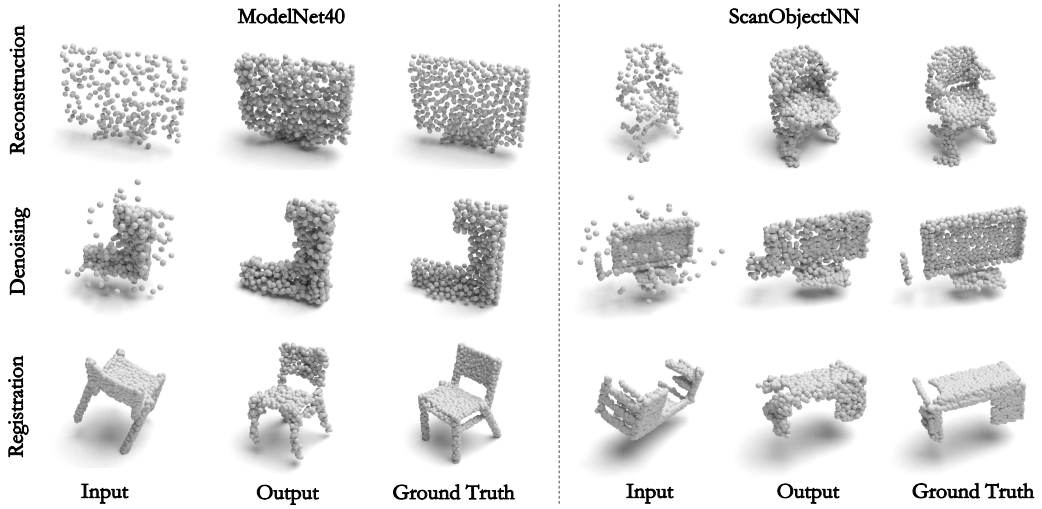

Figure 4: Visualization of our PCoTTA's prediction and their ground truths under 3 different tasks.

target domains to some extent, there still exists some cases of miss-alignment or over-alignment. For example, some samples are either not aligned with the cluster or over-clustered. In contrast, our PCoTTA achieves a better and more even feature alignment across domains, demonstrating its superiority in narrowing domain shifts in continually changing environments.

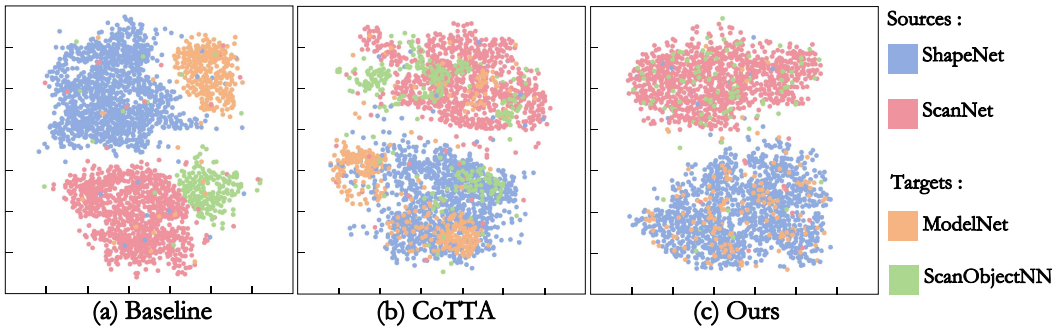

Figure 5: T-SNE visualization of the source and target features.

5 Conclusion

In this paper, we present an innovative, pioneering, and unified framework, namely PCoTTA for Continual Test-Time Adaptation in multi-task point cloud understanding, boosting the model's transferability towards the continually changing target domains. Our approach effectively mitigates catastrophic forgetting and error accumulation issues through the three novel modules: automatic prototype mixture (APM), Gaussian Splatted feature shifting (GSFS), and contrastive prototype repulsion (CPR). These three components make our model more adaptable and robust across continually changing domains by aligning the targets towards all source domains. Furthermore, we present a new benchmark in terms of the practical Continual Test-Time Adaptation for multi-task point cloud understanding. Comprehensive experiments show our PCoTTA's superior performance, proving its efficacy in significantly improving the model's transferability across various domains. We believe our work will inspire a new direction and interesting ideas in the community, in terms of Continual Test-Time Adaptation for multi-task point cloud understanding.

Acknowledgments

Jincen Jiang is supported by the China Scholarship Council (Grand Number 202306300023), and the Research and Development Fund of Bournemouth University.

References

[1] Idan Achituve, Haggai Maron, and Gal Chechik. Self-supervised learning for domain adaptation on point clouds. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 123–133, 2021. 2

[2] Dhanajit Brahma and Piyush Rai. A probabilistic framework for lifelong test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3582–3591, 2023. 2

[3] Haozhi Cao, Yuecong Xu, Jianfei Yang, Pengyu Yin, Shenghai Yuan, and Lihua Xie. Multi-modal continual test-time adaptation for 3d semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18809–18819, 2023. 2, 3

[4] Goirik Chakrabarty, Manogna Sreenivas, and Soma Biswas. Santa: Source anchoring network and target alignment for continual test time adaptation. *Transactions on Machine Learning Research*, 2023. 8, 15

[5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3, 7

[6] Yunlu Chen, Vincent Tao Hu, Efstratios Gavves, Thomas Mensink, Pascal Mettes, Pengwan Yang, and Cees GM Snoek. Pointmixup: Augmentation for point clouds. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 330–345. Springer, 2020. 7, 8, 15

[7] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5828–5839, 2017. 3, 7

[8] Dasith de Silva Edirimuni, Xuequan Lu, Gang Li, Lei Wei, Antonio Robles-Kelly, and Hongdong Li. Straightpcf: Straight point cloud filtering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20721–20730, 2024. 1

[9] Dasith de Silva Edirimuni, Xuequan Lu, Zhiwen Shao, Gang Li, Antonio Robles-Kelly, and Ying He. Iterativepfn: True iterative point cloud filtering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13530–13539, 2023. 1

[10] Mario Döbler, Robert A Marsden, and Bin Yang. Robust mean teacher for continual and gradual test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7704–7714, 2023. 2, 3, 8, 15

[11] Zhongbin Fang, Xiangtai Li, Xia Li, Joachim M Buhmann, Chen Change Loy, and Mengyuan Liu. Explore in-context learning for 3d point cloud understanding. *Advances in Neural Information Processing Systems*, 36, 2024. 3, 4, 7, 8, 15, 16, 17

[12] Sheldon Fung, Xuequan Lu, D Edirimuni, Wei Pan, Xiao Liu, and Hongdong Li. Semreg: Semantics constrained point cloud registration. In *Proceedings of the Europeon Conference on Computer Vision*, 2024. 1

[13] Yulu Gan, Yan Bai, Yihang Lou, Xianzheng Ma, Renrui Zhang, Nian Shi, and Lin Luo. Decorate the newcomers: Visual domain prompt for continual test time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7595–7603, 2023. 2, 3, 7

[14] Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. Note: Robust continual test-time adaptation against temporal correlation. *Advances in Neural Information Processing Systems*, 35:27253–27266, 2022. 3

[15] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7:187–199, 2021. 3, 7, 8, 15

[16] Ahmed Hatem, Yiming Qian, and Yang Wang. Point-tta: Test-time adaptation for point cloud registration using multitask meta-auxiliary learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16494–16504, 2023. 2

[17] Ahmed Hatem, Yiming Qian, and Yang Wang. Test-time adaptation for point cloud upsampling using meta-learning. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1284–1291. IEEE, 2023. 2

[18] Qianjiang Hu, Daizong Liu, and Wei Hu. Density-insensitive unsupervised domain adaption on 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17556–17566, 2023. 2

[19] Jincen Jiang, Xuequan Lu, Lizhi Zhao, Richard Dazaley, and Meili Wang. Masked autoencoders in 3d point cloud representation learning. *IEEE Transactions on Multimedia*, 2023. 1

[20] Jincen Jiang, Lizhi Zhao, Xuequan Lu, Wei Hu, Imran Razzak, and Meili Wang. Dhgcn: Dynamic hop graph convolution network for self-supervised point cloud learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 12883–12891, 2024. 1

[21] Jincen Jiang, Qianyu Zhou, Yuhang Li, Xuequan Lu, Meili Wang, Lizhuang Ma, Jian Chang, and Jian Jun Zhang. Dg-pic: Domain generalized point-in-context learning for point cloud understanding. In *European Conference on Computer Vision*, pages 455–474. Springer, 2025. 2

[22] Peng Jiang and Srikanth Saripalli. Lidarnet: A boundary-aware domain adaptation model for point cloud semantic segmentation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2457–2464. IEEE, 2021. 2

[23] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016. 8, 15, 16

[24] Hanxue Liang, Hehe Fan, Zhiwen Fan, Yi Wang, Tianlong Chen, Yu Cheng, and Zhangyang Wang. Point cloud domain adaptation via masked local 3d structure prediction. In *European Conference on Computer Vision*, pages 156–172. Springer, 2022. 2

[25] Jiaming Liu, Senqiao Yang, Peidong Jia, Renrui Zhang, Ming Lu, Yandong Guo, Wei Xue, and Shanghang Zhang. Vida: Homeostatic visual domain adapter for continual test time adaptation. *arXiv preprint arXiv:2306.04344*, 2023. 8, 9, 15

[26] Wei Liu, Zhiming Luo, Yuanzheng Cai, Ying Yu, Yang Ke, José Marcato Junior, Wesley Nunes Gonçalves, and Jonathan Li. Adversarial unsupervised domain adaptation for 3d semantic segmentation with multi-modal learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176:211–221, 2021. 2

[27] Shaocong Long, Qianyu Zhou, Xiangtai Li, Xuequan Lu, Chenhao Ying, Yuan Luo, Lizhuang Ma, and Shuicheng Yan. Dgmamba: Domain generalization via generalized state space model. In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM)*, 2024. 2

[28] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 7

[29] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021. 1

[30] Zhipeng Luo, Zhongang Cai, Changqing Zhou, Gongjie Zhang, Haiyu Zhao, Shuai Yi, Shijian Lu, Hongsheng Li, Shanghang Zhang, and Ziwei Liu. Unsupervised domain adaptive 3d detection with multi-level consistency. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8866–8875, 2021. 2

[31] Fahim Faisal Niloy, Sk Miraj Ahmed, Dripta S Raychaudhuri, Samet Oymak, and Amit K Roy-Chowdhury. Effective restoration of source knowledge in continual test time adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2091–2100, 2024. 2

[32] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International Conference on Machine Learning*, pages 16888–16905. PMLR, 2022. 2, 3

[33] Yatian Pang, Wenxiao Wang, Francis EH Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *European Conference on Computer Vision*, pages 604–621. Springer, 2022. 1, 3, 7

[34] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017. 1, 3, 7, 8, 15

[35] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017. 3

[36] Can Qin, Haoxuan You, Lichen Wang, C-C Jay Kuo, and Yun Fu. Pointdan: A multi-scale 3d domain adaption network for point cloud representation. *Advances in Neural Information Processing Systems*, 32, 2019. 2

[37] Amirreza Shaban, JoonHo Lee, Sanghun Jung, Xiangyun Meng, and Byron Boots. Lidar-uda: Self-ensembling through time for unsupervised lidar domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19784–19794, 2023. 2

[38] Di Shao, Xuequan Lu, Weijia Wang, Xiao Liu, and Ajmal Saeed Mian. Trici: Triple cross-intra branch contrastive learning for point cloud analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 1

[39] Yuefan Shen, Yanchao Yang, Mi Yan, He Wang, Youyi Zheng, and Leonidas J Guibas. Domain adaptation on point clouds via geometry-aware implicits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7223–7232, 2022. 2

[40] Damian Sójka, Sebastian Cygert, Bartłomiej Twardowski, and Tomasz Trzciński. Ar-tta: A simple method for real-world continual test-time adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3491–3495, 2023. 2

[41] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019. 1

[42] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1588–1597, 2019. 1, 3, 7

[43] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. 8, 15, 16

[44] Feiyu Wang, Wen Li, and Dong Xu. Cross-dataset point cloud recognition using deep-shallow domain adaptation network. *IEEE Transactions on Image Processing*, 30:7364–7377, 2021. 2

[45] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7201–7211, 2022. 2, 3, 7, 8, 9, 15, 16

[46] Yanshuo Wang, Jie Hong, Ali Cheraghian, Shafin Rahman, David Ahmedt-Aristizabal, Lars Petersson, and Mehrtash Harandi. Continual test-time domain adaptation via dynamic sample selection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1701–1710, 2024. 2

[47] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5):1–12, 2019. 1, 3, 7, 8, 15

[48] Yan Wang, Junbo Yin, Wei Li, Pascal Frossard, Ruigang Yang, and Jianbing Shen. Ssda3d: Semi-supervised domain adaptation for 3d object detection from point cloud. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2707–2715, 2023. 2

[49] Yi Wei, Zibu Wei, Yongming Rao, Jiaxin Li, Jie Zhou, and Jiwen Lu. Lidar distillation: Bridging the beam-induced domain gap for 3d object detection. In *European Conference on Computer Vision*, pages 179–195. Springer, 2022. 2

[50] Yushuang Wu, Zizheng Yan, Ce Chen, Lai Wei, Xiao Li, Guanbin Li, Yihao Li, Shuguang Cui, and Xiaoguang Han. Scoda: Domain adaptive shape completion for real scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17630–17641, 2023. 2

[51] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015. 1, 3, 7

[52] Mutian Xu, Runyu Ding, Hengshuang Zhao, and Xiaojuan Qi. Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3173–3182, 2021. 1

[53] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. St3d: Self-training for unsupervised domain adaptation on 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10368–10378, 2021. 2

[54] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. St3d++: Denoised self-training for unsupervised domain adaptation on 3d object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6354–6371, 2022. 2

[55] Senqiao Yang, Jiarui Wu, Jiaming Liu, Xiaoqi Li, Qizhe Zhang, Mingjie Pan, Yulu Gan, Zehui Chen, and Shanghang Zhang. Exploring sparse visual prompt for domain adaptive dense prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16334–16342, 2024. 3

[56] Li Yi, Boqing Gong, and Thomas Funkhouser. Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15363–15373, 2021. 2

[57] Xumin Yu, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, and Jiwen Lu. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19313–19322, 2022. 3, 7

[58] Zhiqi Yu, Jingjing Li, Zhekai Du, Fengling Li, Lei Zhu, and Yang Yang. Noise-robust continual test-time domain adaptation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 2654–2662, 2023. 2

[59] Jinlai Zhang, Lyujie Chen, Bo Ouyang, Binbin Liu, Jihong Zhu, Yujin Chen, Yanmei Meng, and Danfeng Wu. Pointcutmix: Regularization strategy for point cloud classification. *Neurocomputing*, 505:58–67, 2022. 7, 8, 15

[60] Weichen Zhang, Wen Li, and Dong Xu. Srdan: Scale-aware and range-aware domain adaptation network for cross-dataset 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6769–6779, 2021. 2

[61] Sicheng Zhao, Yezhen Wang, Bo Li, Bichen Wu, Yang Gao, Pengfei Xu, Trevor Darrell, and Kurt Keutzer. epointda: An end-to-end simulation-to-real domain adaptation framework for lidar point cloud segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3500–3509, 2021. 2

[62] Qianyu Zhou, Zhengyang Feng, Qiqi Gu, Jiangmiao Pang, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Context-aware mixup for domain adaptive semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 33(2):804–817, 2023. 2

[63] Qianyu Zhou, Qiqi Gu, Jiangmiao Pang, Xuequan Lu, and Lizhuang Ma. Self-adversarial disentangling for specific domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(7):8954–8968, 2023. 2

[64] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Shouhong Ding, and Lizhuang Ma. Test-time domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[65] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Ran Yi, Shouhong Ding, and Lizhuang Ma. Instance-aware domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20453–20463, 2023. 2

[66] Longkun Zou, Hui Tang, Ke Chen, and Kui Jia. Geometry-aware self-training for unsupervised domain adaptation on object point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6403–6412, 2021. 2

A Appendix

Overview. The supplementary includes the following sections:

- **A.1.** Comparison Methods and Reproduction Details.
- **A.2.** Discussion on the Comparison Studies.
- **A.3.** More Ablation Study.
- **A.4.** More Visualization Results.
- **A.5.** Limitations.
- **A.6.** Societal Impacts.

A.1 Comparison Methods and Reproduction Details

Given it is a new setting, we reproduce some state-of-the-art point cloud learning methods and domain adaptation techniques based on the following schemes: *(1) Conventional Methods*. We choose 5 representative methods in point cloud learning, *i.e.,* PointNet [34], DGCNN [47], PCT [15], Pointmixup [6], and PointCutMix [59], reproducing them for multi-task multi-domain learning. In this setting, these methods share a backbone network while incorporating independent task-specific heads designed for different tasks. To ensure a fair comparison, we also devise a multi-task learning framework for these compared methods utilizing a shared backbone network and head to simultaneously learn all 3 tasks in a unified model. Aligning with our method, we replicate the augmentation-based methods Pointmixup and PointCutMix as a domain adaptation model, where each sample is mixed with another sample from a randomly selected source domain. The rest methods are trained directly on multiple different sources. *(2) In-Context Learning (ICL) Methods*. We select PIC [11] which handles multi-task point cloud learning but lacks domain adaptation capability. In our benchmark, we treat all sources as an expanded dataset for PIC training, allowing it to integrate multi-domain information. During the testing stage, we transfer the trained model to the target domain for inference, thus enabling multi-domain learning. Following PIC, we evaluate its baseline that utilizes the prompt target point cloud for prediction. *(3) Continual Test-Time Adaptation (CoTTA) Methods*. We implement several advanced CoTTA methods, utilizing PIC as the backbone network, including AdaBN [23], TENT [43], CoTTA [45], ViDA [25], RMT [10], and SANTA [4]. For teacher-student model-based 2D CoTTA methods, we replace test-time augmentations for 2D images like resizing and flipping with typical 3D data augmentation techniques, such as rotation and scaling. We follow the official settings of forward times for teacher model training. Furthermore, given the Chamfer Distance (CD) loss is used in our point cloud understanding tasks (essentially regression tasks), we optimize the teacher-student model using the typical CD loss as a consistency loss instead of a cross-entropy loss which is typically used in classification tasks, to keep consistency. To ensure fairness in reproduction, we update the LayerNorm parameters for Transformer-based models, like AdaBN, TENT, and SANTA. All methods (including ours) conduct 3 independent evaluation rounds, with samples shuffled randomly in each round.

A.2 Discussion on the Comparison Studies

In Table 1, we present an comprehensive comparison of our PCoTTA and other methods on a series of tasks within our newly established benchmark, including reconstruction, denoising, and registration. For CoTTA methods, following CoTTA [45], we adopt multiple continual adaptation rounds. Regarding other methods, we conduct three individual evaluations for comparison. Notably, our method consistently surpasses all others by a large margin, demonstrating remarkable performance across various tasks among multiple domains. Our unified model adeptly bridges the gaps between the source and target domains at test time even when the targets are continually changing, showing strong adaptation capabilities in this challenging multi-domain multi-task setting.

Comparison to Conventional Methods. Conventional point cloud learning methods like PointNet [34], DGCNN [47], and PCT [15], often face challenges in generalizing to unseen data, leading to significant performance drops. On the other hand, we enable augmentation-based methods, such as Pointmixup [6] and PointCutMix [59], multi-domain learning by mixing samples from different sources. However, they still exhibit limited performance in multi-task generalization. Notably, despite incorporating individual heads for these methods to handle diverse tasks in a task-specific scheme,

they still fall short compared to our method which excels across all tasks. Meanwhile, they also demonstrate inferior performance with a shared network in a multi-task scheme, whereas our PCoTTA is a fully unified model that effectively bridges domain gaps in the multi-task and multi-domain setting. This success is largely due to our proposed APM to estimate the prototype bank generalized across sources and targets, and GSFS to adaptively align the testing sample toward source domains.

Comparison to ICL Methods. ICL methods such as PIC [11] excel in multi-task scenarios using a unified model, but their generalization across various domains is limited. They struggle with unseen data, often failing to perform the specified task well using the prompts provided in existing source domains. In contrast, our PCoTTA effectively tackles this challenge by aligning test data features with familiar source prototypes and dynamically updated learnable prototypes. By utilizing the most similar source sample as the prompt pair, our method effectively narrows the gap between source and target domains, enabling our unified model to enjoy multi-domain generalizability. In addition, to simulate a continuously changing target domain, we performed three independent validations of the compared methods (each time the test target domain would be shuffled), showing a fluctuating performance. Conversely, our method demonstrates strong continuous online learning abilities to attain progressively improved results, thereby verifying the capability of our method in alleviating catastrophic forgetting and error accumulation.

Comparison to CoTTA Methods. Continuous test-time adaptation methods like AdaBN [23], TENT [43], and CoTTA [45] can handle continuously changing unseen targets due to their generalized models. Nonetheless, they struggle with multiple tasks, especially in our challenging setting. To ensure a fair comparison, we adjusted these methods with using PIC as their backbone, equipping them with multi-tasking and multi-domain learning capabilities. As shown in Table 1, our method still outperforms them significantly, consistently demonstrating the ability of our method to continue learning across multiple validation rounds. This success highlights the effectiveness of our feature shifting module (GSFS), which uses Gaussian Splatted Graph Attention to dynamically align unseen samples closer to both the source and learnable prototypes with adaptive weights, enabling our model to handle unfamiliar targets effectively.

A.3 More Ablation Study

we present additional ablation studies in Table A, evaluating the use of CPR and GSFS individually. These results clearly demonstrate the incremental benefits of each component and their combined effect on improving performance. source prototypes are the key and indispensable information that we exploit to devise our methodology. Without it, our method is incomplete in addressing the test-time feature shifting during the continual adaptation and would lead to less decent results. We have also analyzed the cases without source prototypes in our ablation studies, as shown by models B and C in Table A. In this case, our method relies solely on learnable prototypes (i.e., incomplete framework), achieving a certain degree of adaptation. Although this reduces our method's effectiveness, it still outperforms CoTTA [45] Specifically, CoTTA achieves 58.3, 56.7, and 55.2 during the 3 different rounds, while our PCoTTA achieves 36.8, 36.2, and 35.7, demonstrating superiority and the state-of-the-art performance in continual test-time adaptation for 3D point cloud.

Table A: Ablation studies on the individual use of our three proposed modules in PCoTTA.

Models	APM	GSFS	CPR	Rec.	Den.	Reg.
Baseline				78.9	123.6	110.6
A	✓			23.5	37.4	30.1
B		✓		35.7	41.3	39.8
C			✓	46.2	55.6	51.5
Ours	✓	✓	✓	**5.4**	**18.6**	**12.1**

A.4 More Visualization Results

Visual Analysis across Continuous Rounds. we provide T-SNE visualizations for 3 independent validation rounds in Figure A and task-specific visualizations in Figure B. Our method remains stable across continuous rounds, demonstrating that our proposed APM and GSFS effectively mitigate catastrophic forgetting by explicitly leveraging constant source prototypes and source domain representations, thereby avoiding over-reliance on adaptively learned information.

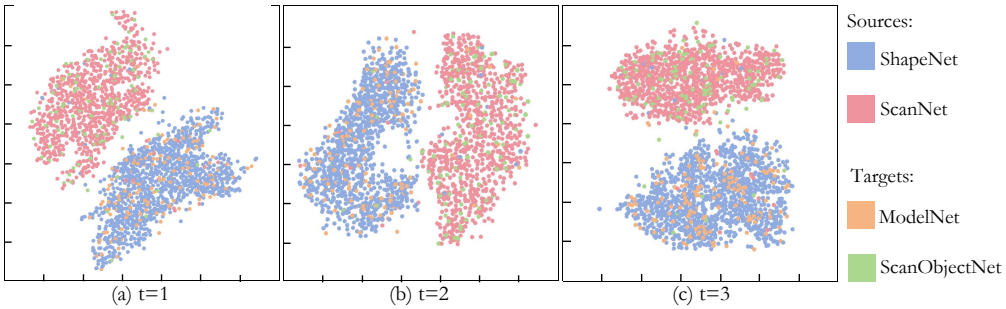

Figure A: T-SNE visualization of three individual evaluation rounds.

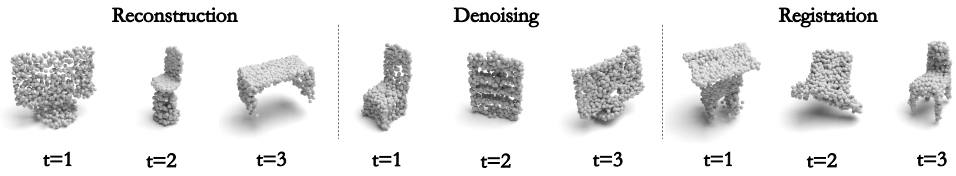

Figure B: Task-specific visualization of three individual evaluation rounds.

More Visual Comparisons. Figure C visually compares qualitative results from the final iteration, showcasing our method's effectiveness in adapting to continually changing targets. From the figure, it is obvious that our PCoTTA excels at producing high-quality predictions across multiple tasks, even as the target domain changes continuously. This success is attributed to our three innovative modules: APM, GSFS, and CPR, which effectively minimize the discrepancies between the source and target domains, thereby enhancing the overall prediction quality. Moreover, Our PCoTTA can handle multiple tasks simultaneously without needing to retrain the off-the-shelf model for each specific task. Our unified model demonstrates its versatility and efficiency by adeptly managing various tasks, including point cloud reconstruction, denoising, and registration. It proves the capability underscores the model's strong practicability and transferability, crucial for real-world applications.

A.5 Limitations

Though our PoCoTTA can handle multi-task point cloud understanding via a unified model, balancing different objectives for largely varied data across multiple tasks poses significant difficulties. For instance, while our method is effective for denoising, it is still not fully optimized for this task, especially compared to specialized denoising methods. In addition, we follow PIC [11] for the three tasks, however, PIC's task names may be slightly misused and different from the original problems, *e.g.,* its registration is to restore the original point cloud from a rotated one, not transformation between two point clouds. They are used as they can be handled similarly given current point learning can predict point positions directly. This makes them 'unified' with position output (x, y, z) and a single loss. We do believe it is a promising future direction of designing more advanced 'unified' ways for other tasks.

A.6 Societal Impacts

Positive Societal Impacts. Our approach reduces the reliance on labeled data for training machine learning models by leveraging a unified model capable of handling various tasks. This cost-saving measure not only benefits businesses by reducing the financial burden associated with data labeling but also democratizes access to AI technologies by lowering the barrier to entry for organizations with limited resources. Furthermore, by providing a unified model that can address multiple tasks, our approach streamlines the development and deployment process for AI systems. This rapid release of unified models enables quicker adoption of AI technologies across various domains, facilitating innovation and improving productivity in sectors such as healthcare, finance, and transportation.

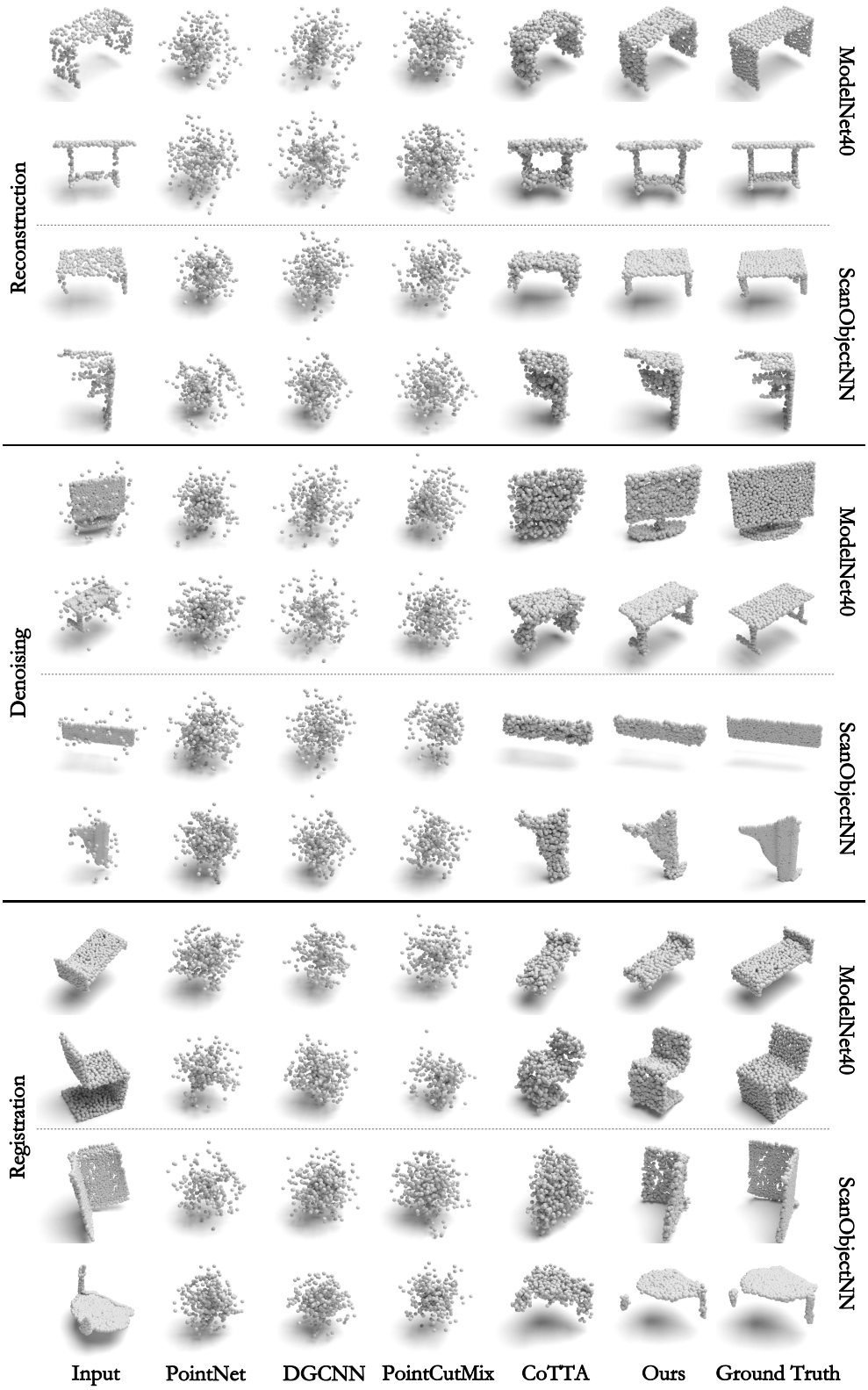

Figure C: Visualization of our PCoTTA and state-of-the-art methods under 3 different tasks.

Potential Negative Societal Impact. The automation of tasks previously performed by humans, facilitated by the rapid deployment of unified models, may lead to job displacement in certain sectors. This could result in economic hardships for individuals and communities reliant on these jobs, exacerbating income inequality and social unrest.

Mitigation strategies. Offering financial assistance, career counseling, and job placement services can help support workers affected by job displacement. Government agencies, non-profit organizations, and private sector employers can collaborate to provide comprehensive support to affected individuals and communities. Prioritizing ethical considerations in the development and deployment of AI technologies can also help mitigate potential negative impacts on society.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: See the abstract and the end of Section 1.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Appendix A.5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: See Section 3.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We disclose the experimental settings to reproduce the main experimental results in our paper in Section 4.1 and the settings of all compared methods in Appendix A.2. Additionally, we provide the code for our proposed method at: https://github.com/Jinec98/PCoTTA.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide our code and data at: https://github.com/Jinec98/PCoTTA.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We provide the optimization and train/test details of our proposed method in Section 4.1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Ours reports the results of multiple rounds of the experiment, reflecting the statistics of the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Section 4.1.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: See Appendix A.6.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Appendix A.6.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The assets used in the paper are properly credited, and we respect the license and terms of use of these assets throughout our research procedures.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: Our paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.