

Test-Time Training for LiDAR Semantic Segmentation under Corruption via Geometric Inlier Discrimination

Hyeonseong Kim* Hyun-Kurl Jang* Kuk-Jin Yoon

Visual Intelligence Lab., KAIST

{brian617, jhg0001, kjyoon}@kaist.ac.kr

Abstract

LiDAR semantic segmentation must remain robust under various sensor and environmental corruptions to be reliable in safety-critical applications. Existing test-time adaptation methods, including approaches based on pseudo-labels and normalization statistics, have shown promising results but can still struggle under severe distribution shifts. To complement these approaches, we propose a geometry-aware test-time training framework that leverages an auxiliary self-supervised objective. Our method is based on geometric inlier discrimination (GeoID), which injects synthetic off-manifold points into the input and trains the model to distinguish geometry-consistent inliers from synthetically displaced outliers, enabling adaptation on unlabeled test data. To further stabilize this process under real corruptions, we introduce bidirectional unreliable point filtering (BiUPF), which uses inlier scores from the source-trained model to filter out unreliable regions on both original and synthetic points, focusing updates on high-confidence samples. Experiments on two large-scale corruption benchmarks, SemanticKITTI-C and nuScenes-C, show that our method consistently outperforms strong test-time adaptation baselines and improves robustness across diverse LiDAR corruptions. Our code is available at <https://github.com/hskim617/GeoID>.

1. Introduction

LiDAR semantic segmentation (LSS) predicts a semantic label for each point in a 3D point cloud captured by LiDAR and is vital for understanding the surrounding environment in applications such as autonomous driving and robotics. With recent advances in deep learning, data-driven supervised LSS methods [1, 7–10, 12, 13, 21, 23, 27, 32, 33, 41, 42, 45, 49, 52] have demonstrated remarkable performance. These approaches achieve strong accuracy when the test data follows the same distribution as the training

*Equal contribution.

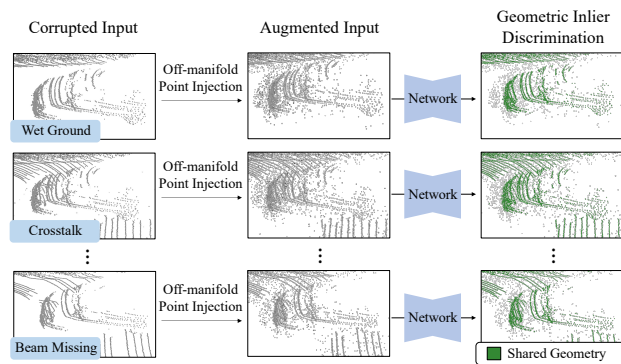


Figure 1. Overview of the proposed method. At test time, we adapt the model using a geometric inlier discrimination task, which serves as self-supervision for test-time training. Synthetic off-manifold points are injected into the raw test input, and the model is trained to distinguish geometry-consistent inliers from synthetically displaced off-manifold outliers, learning to capture the shared geometric structure under corruptions.

data, but in practice this assumption is often violated due to sensor failures and changing environmental conditions. Such data corruptions have been shown to significantly degrade the performance of LSS methods, weakening their robustness [22, 47]. In dynamic environments such as autonomous driving, where conditions constantly evolve and are inherently unpredictable, adapting to distribution shifts in an unsupervised manner at test time is therefore critical.

Test-time training (TTT) adapts a model to the test distribution by utilizing unlabeled test data online during testing. There have been several studies on TTT in the 2D image domain, focusing on regularizing the model by using the entropy of predictions in the test data as the objective function [25, 43, 48], updating batch normalization statistics with test data [28, 30, 38], or solving self-supervised tasks on the test data to provide supervision for adaptation [26, 40]. However, directly applying these methodologies to 3D point clouds has proven to be less effective [29, 35]. In recent work on LiDAR semantic segmentation, several studies have explored online TTT to adapt to distribution shifts in LiDAR point clouds [35, 51, 53].

These methods either aim to obtain more accurate pseudo-labels for self-training or to stabilize adaptation by fusing normalization statistics. However, self-training methods can be limited in effectiveness due to inaccurate model predictions, especially in severely corrupted environments or under substantial domain shifts. This highlights the need for an effective self-supervised objective for test-time training in LiDAR semantic segmentation.

In this paper, we introduce a test-time training framework for LiDAR semantic segmentation under corruption that exploits an auxiliary self-supervised objective. Our key idea is to identify and leverage geometry-consistent inliers that remain close to the underlying scene manifold at test time. To this end, we propose *geometric inlier discrimination* (GeoID), which trains the model to distinguish geometry-consistent inliers from synthetically displaced off-manifold outliers. Concretely, we construct synthetic off-manifold points by applying axis-aligned spatial displacements to the input point cloud and form a binary discrimination task between original and off-manifold points, as illustrated in Fig. 1. In the source domain, the model is jointly trained for semantic segmentation and GeoID. At test time, it is adapted on unlabeled target data using only the GeoID loss, which encourages the encoder to capture the geometric structure of the target scenes even under corruption. To further stabilize this self-supervision under real corruptions, we propose *bidirectional unreliable point filtering* (BiUPF), which uses inlier scores from the source-trained model to estimate point-wise reliability. Then, it filters out ambiguous regions on both original and synthetic points, retaining only high-confidence inliers and outliers for GeoID updates. This focuses adaptation on points that clearly follow or violate the scene geometry, preventing harmful updates. Overall, our framework enables geometry-aware test-time training that is robust to a wide range of LiDAR corruptions, and consistently outperforms strong TTA baselines on two large-scale corruption benchmarks, SemanticKITTI-C and nuScenes-C.

Our main contributions are as follows: (1) We propose a geometry-aware test-time training framework for LiDAR semantic segmentation under corruption, based on *geometric inlier discrimination* (GeoID) as an auxiliary self-supervised objective. (2) We introduce *bidirectional unreliable point filtering* (BiUPF), which filters out unreliable points on both original and synthetic inputs to stabilize GeoID supervision and prevent harmful updates. (3) We show that our method consistently outperforms strong test-time adaptation baselines on SemanticKITTI-C and nuScenes-C across diverse LiDAR corruptions.

2. Related Work

Unsupervised Domain Adaptation for LiDAR Semantic Segmentation. LiDAR semantic segmentation is a core

component of scene understanding for robotics and autonomous driving. Extensive research has explored how to efficiently process LiDAR data, including point-based [13, 24, 33, 42], voxel-based [8, 41, 52], and projection-based methods [1, 6, 21, 45, 49], each tailored to the sparse and irregular nature of 3D measurements. Despite their success, adapting such models to out-of-distribution (OOD) data remains challenging, as changes in sensors, environments, and weather can significantly degrade performance.

Unsupervised domain adaptation (UDA) has been actively studied to mitigate this issue by bridging the distribution gap between a labeled source domain and an unlabeled target domain. Feature-alignment methods [4, 46] reduce the discrepancy between source and target representations, while adversarial training [16, 34] introduces a domain discriminator to encourage domain-invariant features. More recent approaches further enhance robustness by explicitly learning domain-invariant representations [18, 19, 36, 37, 50]. Although UDA has shown promising results by leveraging unlabeled target data during training, real-world deployments often impose stricter constraints: the source data may be unavailable due to privacy or storage limitations, and target data may not be accessible before deployment. These constraints highlight the need for source-free, test-time adaptation approaches that can robustly handle distribution shifts using only target domain observations at inference.

Test-Time Training. Test-time training (TTT) refers to adapting a model, initially trained on source data, to target data when access to the source data is unavailable during testing. Several prominent TTT techniques have been proposed in the image domain, including entropy minimization and batch normalization adaptation. [31, 43] use the entropy minimization objective on model predictions to perform adaptation to the target domain. [25] extends entropy minimization by incorporating a diversity objective, transferring the hypothesis learned from the source data to the target distribution. [17, 28, 30, 38] adjust the model batch normalization statistics for adaptation. An alternative approach employs self-supervision for TTT, initially introduced through rotation prediction [40] as a self-supervised task, which has since expanded to include various auxiliary tasks [2, 11, 26]. Moreover, recent studies have validated the effectiveness of TTT in various 3D vision tasks, including point cloud classification [29], 3D shape completion [39], and semantic scene completion [15]. In the LiDAR domain, TTT methods have also been explored. For example, [35] leverages pseudo-labels to facilitate synthetic-to-real adaptation, while [53] learns prototypes in the target domain and uses them to refine pseudo-labels. Batch normalization adaptation has also been explored for LiDAR [51]. However, existing LiDAR-focused TTT methods can suffer from erroneous optimiza-

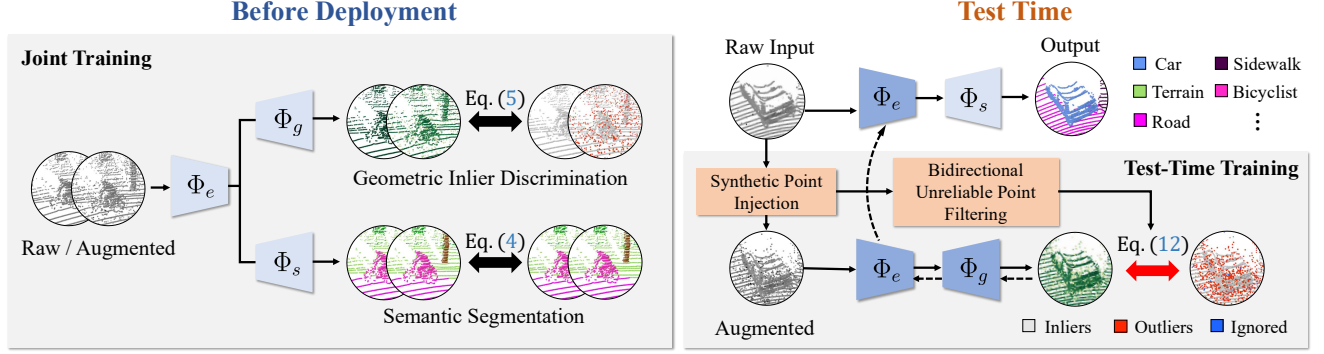


Figure 2. Overview of our GeoID-based test-time training framework for LiDAR semantic segmentation. During source training, the encoder is jointly optimized with a segmentation head and a GeoID head using raw and synthetic point cloud. At test time, GeoID together with bidirectional unreliable point filtering (BiUPF) selects reliable inliers and outliers to drive self-supervised adaptation.

tion caused by inaccurate model predictions or unstable adaptation under severe distribution shifts, particularly in heavily corrupted environments. To address these challenges, we propose a novel TTT framework for LSS that incorporates a self-supervised task. Our work demonstrates the advantages of using geometric inlier discrimination as an auxiliary task for TTT through experiments on datasets corrupted by adverse weather or sensor failure.

3. Method

In this work, we propose a test-time training (TTT) framework for LiDAR semantic segmentation. Given a model $\mathcal{F}$ trained on a source domain $\mathcal{D}_{\text{src}} \sim P_{\text{src}}$ and deployed on a target domain $\mathcal{D}_{\text{tgt}} \sim P_{\text{tgt}}$, the goal of TTT is to adapt model to P_{tgt} using only unlabeled test samples, thereby improving robustness under domain shift without requiring semantic supervision.

In practice, various LiDAR corruptions at test time (e.g., adverse weather, sensor failures) distort the observed point cloud and push measurements away from the clean scene geometry learned in the source domain. Nevertheless, the underlying 3D structure of the environment (roads, buildings, vehicles) remains largely shared across domains. Our key idea is to identify and leverage geometry-consistent inliers that remain close to the true scene manifold, while down-weighting corrupted, off-manifold measurements at test time. To this end, we introduce *geometric inlier discrimination (GeoID)* as an auxiliary self-supervised objective (Sec. 3.1), which trains the model to distinguish geometry-consistent inliers from synthetically displaced off-manifold outliers. Our overall framework is illustrated in Fig. 2. During source training, we jointly optimize the semantic segmentation head and the GeoID branch with a shared encoder (Sec. 3.2). At test time, we update the encoder and the GeoID decoder using inlier discrimination loss (Sec. 3.3). To further stabilize this self-supervision, we discard low-confidence predictions on both original and synthetic points and update the model only with

high-confidence inlier and outlier samples.

3.1. Geometric Inlier Discrimination

We introduce GeoID, a self-supervised task that explicitly leverages LiDAR scene geometry. Given an input LiDAR point cloud, we construct a synthetic outlier distribution Q that intentionally breaks local geometric regularity and use it to supervise a binary discrimination between *geometry-consistent inliers* and *off-manifold outliers*.

We denote the raw point cloud by $\mathcal{X} = \{\mathbf{p}_i\}_{i=1}^N$, where each point is represented as

$$\mathbf{p}_i = (\mathbf{x}_i, r_i) \in \mathbb{R}^4, \quad (1)$$

with $\mathbf{x}_i \in \mathbb{R}^3$ the 3D coordinate and $r_i \in \mathbb{R}$ the LiDAR intensity. We construct off-manifold outliers by applying axis-aligned spatial displacements to the point coordinates along a randomly selected axis:

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \mathbf{e}_a u, \quad u \sim \mathcal{U}([- \delta, \delta]), \quad (2)$$

where $\mathbf{e}_a$ is one of the three unit vectors aligned with the x , y , or z axis. For each displaced coordinate $\tilde{\mathbf{x}}_i$, we assign an intensity by copying from its nearest real point,

$$\tilde{r}_i = r_{j(i)}, \quad j(i) \in \arg \min_k \|\tilde{\mathbf{x}}_i - \mathbf{x}_k\|_2, \quad (3)$$

and define the synthetic point $\tilde{\mathbf{p}}_i = (\tilde{\mathbf{x}}_i, \tilde{r}_i)$. Let $\tilde{\mathcal{X}} = \{\tilde{\mathbf{p}}_i\}_{i=1}^N$ denote the set of synthetic points. We then form an augmented point cloud $\tilde{\mathcal{X}} = \mathcal{X} \cup \tilde{\mathcal{X}} = \{\tilde{\mathbf{p}}_i\}_{i=1}^{2N}$ by injecting the synthetic points into the original point cloud. We denote by P the real inlier distribution over $\mathcal{X}$ and by Q the synthetic, geometrically implausible outlier distribution induced by Eq. (2).

Finally, we assign binary labels $z_i \in \{0, 1\}$, with $z_i = 1$ for original points $\mathbf{p}_i \in \mathcal{X}$ and $z_i = 0$ for synthetic points $\tilde{\mathbf{p}}_i \in \tilde{\mathcal{X}}$. These self-supervised labels define the GeoID objective and are used to train the inlier-outlier discriminator both during source training and at test time for adaptation.

3.2. Joint Training

We jointly optimize the semantic segmentation decoder Φ_s and the GeoID decoder Φ_g with a shared encoder Φ_e on source data. This multi-task setup encourages Φ_e to capture features that are both semantically discriminative and sensitive to geometric plausibility.

Given a labeled point cloud $\mathcal{X} = \{\mathbf{p}_i\}_{i=1}^N$ with labels $\mathcal{Y} = \{y_i\}_{i=1}^N$, we generate synthetic point cloud $\tilde{\mathcal{X}}$ and form the augmented point cloud $\hat{\mathcal{X}}$. Synthetic points in $\tilde{\mathcal{X}}$ are excluded from the segmentation loss via an ignore mask. Let $p(y | \mathbf{p}) = \text{softmax}(\Phi_s(\Phi_e(\mathbf{p})))_y$. Then, we minimize the cross-entropy on raw points $\mathcal{X}$ for the semantic segmentation:

$$\mathcal{L}_{\text{seg}} = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{p}_i \in \mathcal{X}} \left[-\log p(y_i | \mathbf{p}_i) \right]. \quad (4)$$

The GeoID decoder produces inlier scores $c(\hat{\mathbf{p}}_i) = \sigma(\Phi_g(\Phi_e(\hat{\mathbf{p}}_i))) \in [0, 1]$, where $\sigma(\cdot)$ denotes the sigmoid function, and is trained with the binary cross-entropy loss:

$$\mathcal{L}_{\text{geo}} = \frac{1}{|\hat{\mathcal{X}}|} \sum_{\hat{\mathbf{p}}_i \in \hat{\mathcal{X}}} \left[-z_i \log c(\hat{\mathbf{p}}_i) - (1 - z_i) \log (1 - c(\hat{\mathbf{p}}_i)) \right]. \quad (5)$$

The overall joint training objective simply sums the segmentation loss and the GeoID loss:

$$\mathcal{L}_{\text{train}} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{geo}}. \quad (6)$$

Through this joint optimization, Φ_e learns to segment the real inlier distribution P while simultaneously separating geometry-consistent regions likely under P from off-manifold regions likely under Q , providing a suitable initialization for test-time training with GeoID.

3.3. Test-Time Training

At test time, LiDAR points arrive sequentially, and the model adapts using only the GeoID objective. For each test sample $\mathcal{X}$, we construct B independent synthetic outlier sets $\{\tilde{\mathcal{X}}_b\}_{b=1}^B$ as in Eq. (2) and form a mini-batch of augmented point clouds

$$\{\hat{\mathcal{X}}_b\}_{b=1}^B = \{\mathcal{X} \cup \tilde{\mathcal{X}}_b\}_{b=1}^B, \quad (7)$$

where B is the mini-batch size.

Bidirectional Unreliable Point Filtering. In real-world corruptions at test time, some LiDAR measurements are artifacts, and conversely, a displaced synthetic point can occasionally lie on a geometric surface. Both situations indicate regions where the real-synthetic labeling can be unreliable due to geometric ambiguity.

To avoid unstable updates from such cases, we introduce *bidirectional unreliable point filtering (BiUPF)* shown in Fig. 3. We keep the source encoder and GeoID decoder,

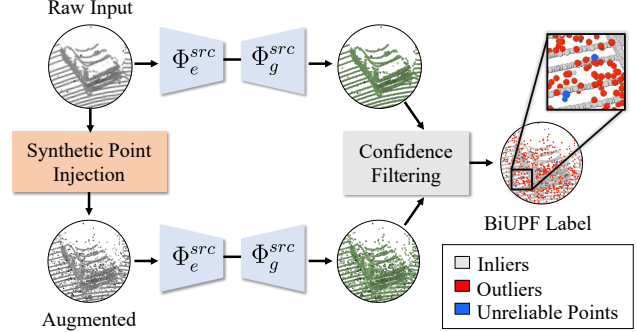


Figure 3. Overview of BiUPF. We use inlier scores from the source model to filter out unreliable points, retaining only high-confidence inliers and outliers for GeoID supervision.

Φ_e^{src} and Φ_g^{src} , frozen after joint training and use them to compute an inlier score for each point:

$$c^{\text{src}}(\mathbf{p}) = \sigma(\Phi_g^{\text{src}}(\Phi_e^{\text{src}}(\mathbf{p}))) \in [0, 1]. \quad (8)$$

We then filter out low-reliability points by using the inlier score $c^{\text{src}}(\mathbf{p})$ as a reliability measure. On the raw input,

$$\mathcal{I}_{\text{in}} = \{i \mid \mathbf{p}_i \in \mathcal{X}, c^{\text{src}}(\mathbf{p}_i) \geq \tau_r\}, \quad (9)$$

and on the synthetic set $\tilde{\mathcal{X}} = \bigcup_{b=1}^B \tilde{\mathcal{X}}_b$,

$$\mathcal{I}_{\text{out}} = \{j \mid \tilde{\mathbf{p}}_j \in \tilde{\mathcal{X}}, c^{\text{src}}(\tilde{\mathbf{p}}_j) \leq 1 - \tau_r\}, \quad (10)$$

where τ_r denotes the reliability threshold. We then define the reliable index set as $\mathcal{I} = \mathcal{I}_{\text{in}} \cup \mathcal{I}_{\text{out}}$. Computing the GeoID objective only over $\mathcal{I}$ focuses gradients on points that clearly exhibit or clearly violate geometric regularity, stabilizing self-supervision and steering the encoder toward geometry-aware features during adaptation.

Test-Time Training. Given the reliable set $\mathcal{I}$ from BiUPF, we perform a self-supervised update for each test sample using only the GeoID task. We assign binary targets $z_i = 1$ for $i \in \mathcal{I}_{\text{in}}$ and $z_i = 0$ for $i \in \mathcal{I}_{\text{out}}$. We define the GeoID inlier score produced by the current test-time model with parameters θ as

$$c_{\theta}(\mathbf{p}) = \sigma(\Phi_g(\Phi_e(\mathbf{p}))). \quad (11)$$

The per-sample objective is the binary cross-entropy over the retained points:

$$\mathcal{L}_{\text{test}} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left[-z_i \log c_{\theta}(\hat{\mathbf{p}}_i) - (1 - z_i) \log (1 - c_{\theta}(\hat{\mathbf{p}}_i)) \right]. \quad (12)$$

We update the shared encoder Φ_e together with the GeoID decoder Φ_g . This adaptation refines Φ_e so that geometry-consistent structures in the target point cloud are emphasized and mapped to features that remain compatible with the segmentation decoder Φ_s .

3.4. Theoretical Insight and Intuition

We explain why the proposed geometric inlier discrimination task with test-time training provides a geometry-aware self-supervised signal that improves LiDAR semantic segmentation under corruption.

Setup. Let P_{src} and P_{tgt} be the source and target LiDAR point distributions, and Q_{src} , Q_{tgt} the corresponding synthetic outlier distributions. The GeoID inlier scores $c(\mathbf{x}) = \sigma(\Phi_g(\Phi_e(\mathbf{x})))$ is trained to predict 1 for real points and 0 for synthetic points with Eq. (5). During TTT on the target domain, Φ_e and Φ_g are updated using unlabeled test samples and their synthetic counterparts.

Assumptions. We rely on three mild, realistic assumptions: (i) *Geometry-consistent inliers*: real LiDAR points are more likely to form coherent geometric patterns (surfaces, object shapes, locally consistent structures) than arbitrary off-structure configurations. (ii) *Geometry-breaking synthetic noise*: our displacement-based outlier construction more frequently produces geometrically implausible (off-manifold) points than real observations. (iii) *Shared geometry-driven semantics*: typical shifts (sensor, sparsity, weather) alter density and appearance, but the semantic meaning of geometric patterns (e.g., planar roads, vertical facades) is largely preserved across domains.

Geometric inlier discrimination. For a fixed (P, Q) , the Bayes-optimal auxiliary classifier is

$$c^*(\mathbf{x}) = \frac{P(\mathbf{x})}{P(\mathbf{x}) + Q(\mathbf{x})}. \quad (13)$$

Under (i) and (ii), geometry-consistent points have $P(\mathbf{x}) \gg Q(\mathbf{x})$ so $c^*(\mathbf{x}) \approx 1$, while geometrically implausible points have $Q(\mathbf{x}) \gg P(\mathbf{x})$ so $c^*(\mathbf{x}) \approx 0$. Thus, the optimal solution behaves like an indicator of geometry-consistent inliers, without requiring an explicit parametric model of surfaces. Minimizing $\mathcal{L}_{\text{geo}}$ therefore encourages the shared encoder Φ_e to separate geometry-consistent and off-manifold patterns in feature space.

Why geometry over corruption-specific cues. At test time, supervision comes only from the real versus synthetic labels. Our synthetic points are designed to break geometric plausibility rather than imitate specific corruptions. Corruption-specific artifacts (e.g., snow, sensor noise, sparsity) arise from the scene and the sensing process and thus can occur in both real and synthetic points. Because these cues are not reliably aligned with the real-synthetic labels, they are suboptimal for minimizing $\mathcal{L}_{\text{geo}}$. In contrast, geometric consistency is intentionally aligned: coherent structures are more often real, displaced structures more often synthetic. Hence, optimization naturally favors geometry-consistent features over corruption-specific statistics.

Role of bidirectional unreliable point filtering. BiUPF further sharpens this effect by discarding low-reliability predictions from both sides. Original points predicted as

noise and synthetic points predicted as real are excluded from the GeoID loss. This restricts self-supervision to high-confidence real points that exhibit stable geometry and high-confidence synthetic points that clearly violate it, reinforcing the interpretation of the GeoID task as learning a robust inlier indicator.

Test-time training effect. Under distribution shift, Φ_e 's target features drift, creating a decision boundary mismatch. TTT with the GeoID loss updates Φ_e so that target points forming coherent geometric patterns are mapped to inlier-like features, while synthetic and clearly off-manifold patterns are pushed away. This reconstruction of a geometry-consistent inlier region in the target feature space reduces the mismatch with the source-trained Φ_s under assumption (iii), providing a geometry-aware, label-free regularization that improves robustness across diverse corruptions.

Intuitive view across corruption types. The above perspective can be intuitively interpreted across common LiDAR corruptions.

- *Sensor change / sparsity*: Different sensors or ranges reduce density, but large-scale shapes (roads, facades, vehicles) remain. GeoID pushes the encoder to focus on such geometry-consistent patterns rather than absolute density, so sparse yet coherent structures are still treated as inliers.
- *Weather / artifacts*: Snow, fog, and spurious returns create clutter. Since synthetic points are designed to break geometric regularity and BiUPF removes ambiguous points, the adaptation signal emphasizes stable structures over clutter-like artifacts.
- *Dropout / occlusion*: Even with missing regions, remaining points locally preserve consistent geometry. GeoID reinforces these residual coherent regions as inliers, forming a cleaner inlier manifold from incomplete observations.

Across these cases, the GeoID with BiUPF consistently amplifies geometry-consistent patterns and suppresses off-manifold behavior in the target representation, thereby reducing mismatch with the segmentation decoder.

4. Experiments

4.1. Experiment Protocol

4.1.1. Evaluation Dataset.

We use two large-scale LiDAR semantic segmentation datasets, SemanticKITTI and nuScenes-lidarseg, as our source datasets and evaluate their corrupted counterparts, which introduce various real-world corruptions.

SemanticKITTI [3] consists of 19,130 training samples and 4,071 evaluation samples from sequence 08, captured by a 64-beam LiDAR and annotated with 19 semantic classes.

nuScenes-lidarseg [5] comprises 28,130 training samples and 6,019 test samples, captured by a 32-beam LiDAR and

Table 1. Comparison results against test-time adaptation methods, reporting mIoU (%) for each corruption type in SemanticKITTI-C. The best result is highlighted in **bold**.

Method	Fog	Wet Ground	Snow	Motion Blur	Beam Missing	Crosstalk	Incomplete Echo	Cross Sensor	Mean
Source-only	33.20	45.53	32.04	54.09	51.05	24.46	50.54	47.76	42.33
TTT [40]	36.79	47.84	33.69	53.23	51.20	27.10	50.81	47.99	43.58
TENT [43]	35.80	38.37	35.04	52.19	48.70	29.37	48.60	44.55	41.58
DUA [28]	32.46	39.13	35.09	52.73	49.52	28.58	49.39	45.68	41.57
GIPSO [35]	31.64	37.47	34.87	51.09	47.68	31.83	47.38	43.80	40.72
HGL [53]	35.92	37.84	35.62	52.05	48.42	28.51	48.10	44.75	41.40
GeoID (Ours)	40.14	47.66	40.63	54.59	52.10	40.82	50.79	48.97	46.96

annotated with 16 semantic classes.

Corruption dataset [22] consists of SemanticKITTI-C and nuScenes-C, which apply eight types of corruptions, including adverse weather, disturbances, and sensor failures, to the original datasets. We follow the same split configuration as the original datasets. Each corruption dataset provides three severity levels, and we report the average over all levels for our main evaluation.

4.1.2. Baseline

Most existing TTA methods for LSS primarily target synthetic-to-real adaptation and use MinkowskiNet [9] as their backbone. Following this setting, we evaluate corruption robustness on real-world datasets using MinkowskiNet, and compare our approach against these LiDAR TTA methods as well as several representative TTA methods originally proposed for 2D vision. In addition, we evaluate our method on a recent corruption-robustness benchmark built on top of the Cylinder3D [52] backbone.

Source-only models are trained only on semantic labels and evaluated on test samples using original LiDAR data.

TTT [40] adapts the model parameters using an auxiliary self-supervised task, specifically rotation prediction. Originally designed for the image domain, the implementation was modified for LiDAR by predicting rotations of 0, 30, 60, and 90 degrees. The rotation prediction branch shares the same encoder as the main segmentation branch.

TENT [43] updates batch normalization statistics and transformation parameters using an entropy minimization objective, thus managing the distribution gap between the source and target domains.

DUA [28] adapts the statistics of the batch normalization layer during the online evaluation of target data.

GIPSO [35] generates pseudo labels using the frozen source model and propagates them based on geometric features from an auxiliary branch. The generated pseudo labels are then used to adapt the target model online.

HGL [53] generates local geometric pseudo-labels using a KNN-based clustering strategy, and prototypes are employed to finetune the pseudo-labels.

4.1.3. Implementation Details

We train the model from scratch on both source datasets with a learning rate of 0.01 and a batch size of 4. For all training, including both the baselines and our method, we use data augmentations such as random rotation around the z-axis, random translation, scaling, and random flipping along the x, y, or x-y plane. However, for TTT [40], we use quantized rotation angles instead of fully random rotations for the rotation prediction task. For a fair comparison, we keep the architectural details identical to those of the corresponding baselines. During test-time training, we use a learning rate of 0.001 for both evaluation datasets. We employ the Adam optimizer [20] for both joint training and test-time training, and use a mini-batch size of $B = 2$ during test-time training. Additional implementation details are provided in the *supple*.

4.2. Results

Table 1 shows the results for all corruption types on SemanticKITTI-C. Our method significantly outperforms all baselines across most corruption types. This demonstrates that the proposed GeoID provides effective self-supervision during test-time training, allowing the model to successfully adapt to corrupted LiDAR data. Methods such as TENT [43], which introduces entropy-based regularization, and DUA [28], which updates batch normalization statistics, show poor performance, suggesting that methods originally designed for the 2D image domain are not effective when directly applied to LiDAR point clouds. Surprisingly, GIPSO [35] and HGL [53] that were proposed specifically for test-time adaptation in LiDAR semantic segmentation also perform poorly except for a few corruption types. We suspect that self-training methods relying on pseudo-labels struggle to adapt under severe distribution shifts, as inaccurate pseudo-labels hinder effective adaptation.

In contrast, TTT [40] slightly outperforms the source-only model overall, indicating that the rotation prediction task helps the model adapt by capturing information across the entire LiDAR scene. However, the large performance gap between our method and TTT [40] suggests that the rotation prediction task is less suitable for test-time training of LSS models than the proposed GeoID task. Overall,

Table 2. Comparison results against test-time adaptation methods, reporting mIoU (%) for each corruption type in nuScenes-C. The best result is highlighted in **bold**.

Method	Fog	Wet Ground	Snow	Motion Blur	Beam Missing	Crosstalk	Incomplete Echo	Cross Sensor	Mean
Source-only	47.23	69.12	37.69	67.20	63.07	19.71	59.62	46.38	51.25
TTT [40]	46.11	66.91	45.03	65.17	61.89	16.84	59.14	47.16	51.03
TENT [43]	33.94	63.77	40.29	64.10	59.34	32.40	56.30	42.52	49.08
DUA [28]	34.83	68.34	42.82	67.27	62.72	32.90	58.85	45.95	51.71
GIPSO [35]	35.74	65.49	39.35	64.12	59.98	28.88	56.94	43.20	49.21
HGL [53]	33.81	64.05	39.55	63.76	59.39	31.62	56.16	42.46	48.85
GeoID (Ours)	57.40	69.27	53.91	69.12	63.49	30.84	59.72	50.11	56.73

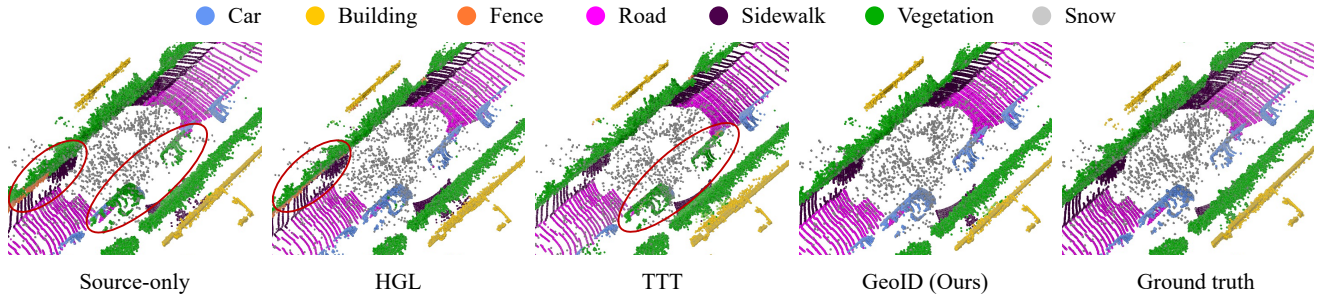


Figure 4. Qualitative results on SemanticKITTI-C. The region highlighted in the red circle indicates incorrect predictions.

these results demonstrate that our method is highly effective at adapting to the target distribution by identifying geometrically consistent inliers and successfully handling a wide range of corruption scenarios in LiDAR point clouds.

Table 2 shows the results for all corruption types on nuScenes-C. Our method outperforms the baselines on most corruption types and performs comparably to the best method in the remaining cases. This highlights the strong generalization capability of the proposed approach. Methods originally designed for the 2D image domain [40, 43] and LiDAR-domain adaptation methods [35, 53] exhibit severe failures to adapt in some corruption cases. DUA [28] achieves slightly better overall performance than the source-only model, but our method still maintains a substantial performance margin over all baselines, further validating the suitability of the GeoID task for test-time training across two large-scale benchmarks.

Table 3 reports a comparison on a recent test-time adaptation benchmark [51] for LiDAR corruption robustness built on the Cylinder3D backbone. Even with the Cylinder3D backbone, our method achieves the best overall performance and consistently outperforms existing TTA baselines. This indicates that our method is not tied to a specific architecture, but generalizes well across different LiDAR segmentation backbones. A detailed breakdown of performance for each corruption type is provided in the *supple*.

In Fig. 4, we visualize qualitative semantic segmentation results on SemanticKITTI-C. Under snow corruption, both the source-only model and TTT [40] misclassify the car as vegetation, while HGL [53] misclassifies vegetation

Table 3. Comparison results between GeoID and other test-time adaptation methods on SemanticKITTI-C using the Cylinder3D [52] segmentation model.

Source-only	MixNorm [14]	TENT [43]	CoTTA [44]	NormFusion [51]	GeoID (Ours)
52.43	53.83	52.86	54.11	55.25	55.46

as a fence. In contrast, our GeoID correctly predicts the labels for these points.

4.3. Analysis

We first conduct an ablation study to analyze the effects of GeoID and BiUPF during test-time training, and then evaluate how our method performs under different update strides for sparse adaptation and under continually changing corruption-induced distribution shifts. All experiments in this analysis are conducted on SemanticKITTI-C at the moderate corruption severity level. Additional results on clean target datasets, the correlation between GeoID and semantic segmentation, and sensitivity analyses of the reliability threshold τ_r , maximum displacement δ , and perturbation direction, are provided in the *supple*.

Ablation Study. To assess the contribution of each component of our method, we perform an ablation study on GeoID and BiUPF at test-time training. The results are summarized in Tab. 4. When GeoID is applied together with BiUPF, the performance consistently improves over the pre-adaptation model across all corruption types, yielding an average gain

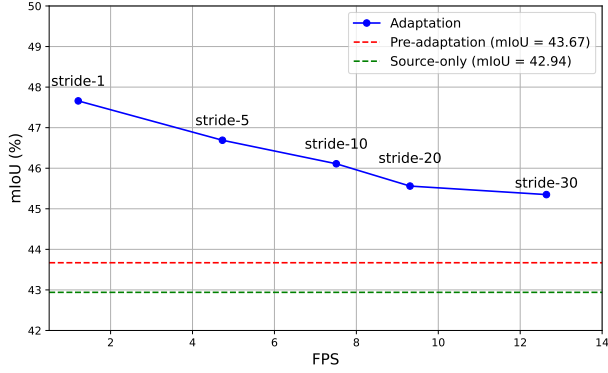


Figure 5. Effect of update stride on FPS and adaptation performance in test-time training. With a stride of 30, our method achieves over 12 FPS while preserving its adaptation effect.

Table 4. Ablation study of GeoID and BiUPF within the test-time training framework on SemanticKITTI-C.

Method	Fog	Wet Ground	Snow	Motion Blur
Pre-adapt.	38.3	45.0	33.7	53.0
Adapt. (w/o BiUPF)	42.5 (+4.2)	45.4 (+0.4)	39.8 (+6.1)	54.5 (+1.5)
Adapt. (w/ BiUPF)	42.6 (+4.3)	46.9 (+1.9)	40.7 (+7.0)	54.6 (+1.6)
Method	Beam Missing	Crosstalk	Incomplete Echo	Cross Sensor
Pre-adapt.	51.7	25.5	51.3	50.9
Adapt. (w/o BiUPF)	50.5 (-1.2)	43.1 (+17.6)	50.5 (-0.8)	48.2 (-2.7)
Adapt. (w/ BiUPF)	52.7 (+1.0)	40.9 (+15.4)	51.7 (+0.4)	51.3 (+0.4)

of 4.0%. This empirically demonstrates that, under a reasonable assumption on the target distribution, geometric inlier discrimination on unlabeled target data can effectively adapt the model. In contrast, when BiUPF is disabled and only GeoID is applied, the performance slightly degrades for all corruptions. For several corruptions (beam missing, incomplete echo, and cross sensor), the performance even falls marginally below the pre-adaptation baseline. These observations indicate that removing unreliable points from the supervision signal via BiUPF is crucial for GeoID, preventing harmful updates and enabling stable adaptation under corrupted LiDAR inputs.

Update Stride in Test-Time Training. Most of the computational cost in test-time training arises from backpropagation. Therefore, when a high frame rate (FPS) is required, sparse adaptation with a larger update stride can improve overall FPS. We evaluate our method with various update strides, as shown in Fig. 5. On an NVIDIA A6000 GPU, our method runs at over 12 FPS, incurring only a 2.31% performance penalty when the model is updated every 30 samples, while maintaining an adaptation speed comparable to typical LiDAR frame rates. Even under this setting, our method still yields a 2.41% improvement over the source-only model, showing that it can sustain high FPS while preserving a reasonable level of adaptation.

Continual Test-Time Training. In real-world deploy-

Table 5. Continual test-time training results on SemanticKITTI-C. “Comb. 1-3” denote three random continual test streams constructed by combining eight different LiDAR corruptions.

Method	Comb. 1	Comb. 2	Comb. 3
Pre-adapt.	42.79	42.14	44.79
Adapt.	46.61	48.01	49.29

ments, sensor data are often affected by time-varying corruptions (e.g., noise, blur, and adverse weather). To evaluate the adaptation performance of our method in a continual test-time training scenario under such evolving corruptions, we construct a setting in which different corruption types appear in sequence. Specifically, we partition the test data stream into eight equal segments and randomly assign one of eight corruption types to each segment. We generate three such random combinations in total, with full details provided in the *supple*. Results on this benchmark are reported in Tab. 5. Across all combinations, our method consistently improves over the pre-adaptation performance, with an average gain of 4.37%. These results show that our method can successfully adapt even under challenging conditions with continuously evolving corruptions.

5. Conclusion

In this paper, we addressed test-time training for LiDAR semantic segmentation under corruption. We proposed a test-time training framework that leverages an auxiliary self-supervised task based on geometric inlier discrimination (GeoID). We demonstrated that GeoID provides strong self-supervision under distribution shifts by identifying geometrically consistent inliers in corrupted LiDAR point clouds. In addition, we introduced bidirectional unreliable point filtering to further improve adaptation by removing unreliable points from the supervision signal. Finally, we showed that our method supports efficient sparse adaptation and remains effective even when distribution shifts change continuously over time, leading to robust performance across diverse corruption scenarios.

Limitations and Future Works. Our current instantiation of off-manifold points relies on simple axis-aligned displacements, without explicitly modeling fine-grained surfaces. Developing more accurate surface models and principled off-manifold perturbations could lead to a more faithful geometry-consistent manifold for test-time adaptation. In addition, we do not explicitly detect real artifacts caused by corruptions, nor do we handle highly complex mixtures of corruptions, both of which are promising directions. Finally, our method assumes that different domains share similar underlying geometry; relaxing this assumption to handle cases where geometry itself changes remains an important avenue for future work.

Acknowledgments

This work was supported by the InnoCORE program of the Ministry of Science and ICT(N10250156); by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. RS-2024-00457882, AI Research Hub Project); and by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2026-25473963).

References

- [1] Angelika Ando, Spyros Gidaris, Andrei Bursuc, Gilles Puy, Alexandre Boulch, and Renaud Marlet. Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5240–5250, 2023. [1](#), [2](#)
- [2] Alexander Bartler, Andre Bühler, Felix Wiewel, Mario Döbler, and Bin Yang. Mt3: Meta test-time training for self-supervised test-time adaption. In *International Conference on Artificial Intelligence and Statistics*, pages 3080–3090. PMLR, 2022. [2](#)
- [3] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Juergen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9297–9307, 2019. [5](#)
- [4] Yikai Bian, Le Hui, Jianjun Qian, and Jin Xie. Unsupervised domain adaptation for point cloud semantic segmentation via graph matching. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9899–9904. IEEE, 2022. [2](#)
- [5] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multi-modal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. [5](#)
- [6] Ke Chen, Ryan Oldja, Nikolai Smolyanskiy, Stan Birchfield, Alexander Popov, David Wehr, Ibrahim Eden, and Joachim Pehserl. Mvlidarnet: Real-time multi-class scene understanding for autonomous driving using multiple views. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2288–2294. IEEE, 2020. [2](#)
- [7] Hui-Xian Cheng, Xian-Feng Han, and Guo-Qiang Xiao. Transrvnet: Lidar semantic segmentation with transformer. *IEEE Transactions on Intelligent Transportation Systems*, 24(6):5895–5907, 2023. [1](#)
- [8] Ran Cheng, Ryan Razani, Ehsan Taghavi, Enxu Li, and Bingbing Liu. 2-s3net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12547–12556, 2021. [2](#)
- [9] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019. [6](#)
- [10] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *International Symposium on Visual Computing*, pages 207–222. Springer, 2020. [1](#)
- [11] Yossi Gandelsman, Yu Sun, Xinlei Chen, and Alexei Efros. Test-time training with masked autoencoders. *Advances in Neural Information Processing Systems*, 35:29374–29385, 2022. [2](#)
- [12] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018. [1](#)
- [13] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11108–11117, 2020. [1](#), [2](#)
- [14] Xuefeng Hu, Gokhan Uzunbas, Sirius Chen, Rui Wang, Ashish Shah, Ram Nevatia, and Ser-Nam Lim. Mixnorm: Test-time adaptation through online normalization estimation. *arXiv preprint arXiv:2110.11478*, 2021. [7](#)
- [15] Hyun-Kurl Jang, Jihun Kim, Hyeokjun Kweon, and Kuk-Jin Yoon. Talos: Enhancing semantic scene completion via test-time adaptation on the line of sight. *Advances in Neural Information Processing Systems*, 37:74211–74232, 2024. [2](#)
- [16] 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](#)
- [17] Ansh Khurana, Sujoy Paul, Piyush Rai, Soma Biswas, and Gaurav Aggarwal. Sita: Single image test-time adaptation. *arXiv preprint arXiv:2112.02355*, 2021. [2](#)
- [18] Hyeonseong Kim, Yoonsu Kang, Changgyoon Oh, and Kuk-Jin Yoon. Single domain generalization for lidar semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17587–17598, 2023. [2](#)
- [19] Jaeyeul Kim, Jungwan Woo, Jeonghoon Kim, and Sunghoon Im. Rethinking lidar domain generalization: Single source as multiple density domains. In *European Conference on Computer Vision*, pages 310–327. Springer, 2024. [2](#)
- [20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [6](#)
- [21] Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. Rethinking range view representation for lidar segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 228–240, 2023. [1](#), [2](#)

- [22] Lingdong Kong, Youquan Liu, Xin Li, Runnan Chen, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Robo3d: Towards robust and reliable 3d perception against corruptions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19994–20006, 2023. 1, 6
- [23] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. Spherical transformer for lidar-based 3d recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17545–17555, 2023. 1
- [24] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018. 2
- [25] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020. 1, 2
- [26] Yuejiang Liu, Parth Kothari, Bastien Van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt++: When does self-supervised test-time training fail or thrive? *Advances in Neural Information Processing Systems*, 34: 21808–21820, 2021. 1, 2
- [27] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 4213–4220. IEEE, 2019. 1
- [28] M Jehanzeb Mirza, Jakub Micorek, Horst Possegger, and Horst Bischof. The norm must go on: Dynamic unsupervised domain adaptation by normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14765–14775, 2022. 1, 2, 6, 7
- [29] M Jehanzeb Mirza, Inkyu Shin, Wei Lin, Andreas Schriebl, Kunyang Sun, Jaesung Choe, Mateusz Kozinski, Horst Possegger, In So Kweon, Kuk-Jin Yoon, et al. Mate: Masked autoencoders are online 3d test-time learners. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16709–16718, 2023. 1, 2
- [30] Zachary Nado, Shreyas Padhy, D Sculley, Alexander D’Amour, Balaji Lakshminarayanan, and Jasper Snoek. Evaluating prediction-time batch normalization for robustness under covariate shift. *arXiv preprint arXiv:2006.10963*, 2020. 1, 2
- [31] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofu Chen, Shijian Zheng, Peilin Zhao, and Minghui Tan. Efficient test-time model adaptation without forgetting. In *International conference on machine learning*, pages 16888–16905. PMLR, 2022. 2
- [32] 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 conference on computer vision and pattern recognition*, pages 652–660, 2017. 1
- [33] 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. 1, 2
- [34] Mrigank Rochan, Shubhra Aich, Eduardo R Corral-Soto, Amir Nabatchian, and Bingbing Liu. Unsupervised domain adaptation in lidar semantic segmentation with self-supervision and gated adapters. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2649–2655. IEEE, 2022. 2
- [35] Cristiano Saltori, Evgeny Krivosheev, Stéphane Lathuilière, Nicu Sebe, Fabio Galasso, Giuseppe Fiameni, Elisa Ricci, and Fabio Poiesi. Gipso: Geometrically informed propagation for online adaptation in 3d lidar segmentation. In *European Conference on Computer Vision*, pages 567–585. Springer, 2022. 1, 2, 6, 7
- [36] Cristiano Saltori, Aljosa Osep, Elisa Ricci, and Laura Leal-Taixé. Walking your lidog: A journey through multiple domains for lidar semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 196–206, 2023. 2
- [37] Jules Sanchez, Jean-Emmanuel Deschaud, and François Goulette. Domain generalization of 3d semantic segmentation in autonomous driving. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 18077–18087, 2023. 2
- [38] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. *Advances in neural information processing systems*, 33:11539–11551, 2020. 1, 2
- [39] Michael Schopf-Kuester, Zorah Lähner, and Michael Moeller. 3d shape completion with test-time training. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024. 2
- [40] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020. 1, 2, 6, 7
- [41] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching efficient 3d architectures with sparse point-voxel convolution. In *European conference on computer vision*, pages 685–702. Springer, 2020. 1, 2
- [42] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotequi, 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, 2
- [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. 1, 2, 6, 7
- [44] 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. 7
- [45] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. SqueezeSeg: Convolutional neural nets with recurrent crf for

- real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1887–1893. IEEE, 2018. [1](#), [2](#)
- [46] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4376–4382. IEEE, 2019. [2](#)
- [47] Xu Yan, Chaoda Zheng, Ying Xue, Zhen Li, Shuguang Cui, and Dengxin Dai. Benchmarking the robustness of lidar semantic segmentation models. *International Journal of Computer Vision*, 132(7):2674–2697, 2024. [1](#)
- [48] Marvin Zhang, Sergey Levine, and Chelsea Finn. Memo: Test time robustness via adaptation and augmentation. *Advances in neural information processing systems*, 35:38629–38642, 2022. [1](#)
- [49] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zelong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9601–9610, 2020. [1](#), [2](#)
- [50] Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, and Dacheng Tao. Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14781–14791, 2024. [2](#)
- [51] Yunxiang Zhou, Xin Lou, Ying He, Wei Zhang, and Indriyati Atmosukarto. Normfusion: Real-time test-time adaptation for lidar segmentation under corruption. In *2025 IEEE Conference on Artificial Intelligence (CAI)*, pages 1029–1035. IEEE, 2025. [1](#), [2](#), [7](#)
- [52] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9939–9948, 2021. [1](#), [2](#), [6](#), [7](#)
- [53] Tianpei Zou, Sanqing Qu, Zhijun Li, Alois Knoll, Lianghua He, Guang Chen, and Changjun Jiang. Hgl: Hierarchical geometry learning for test-time adaptation in 3d point cloud segmentation. In *European Conference on Computer Vision*, pages 19–36. Springer, 2024. [1](#), [2](#), [6](#), [7](#)