IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

GeoAdapt: Self-Supervised Test-Time Adaptation in LiDAR Place Recognition Using Geometric Priors

Joshua Knights[1,2], Stephen Hausler[1], Sridha Sridharan[2], Clinton Fookes[2], Peyman Moghadam[1,2]

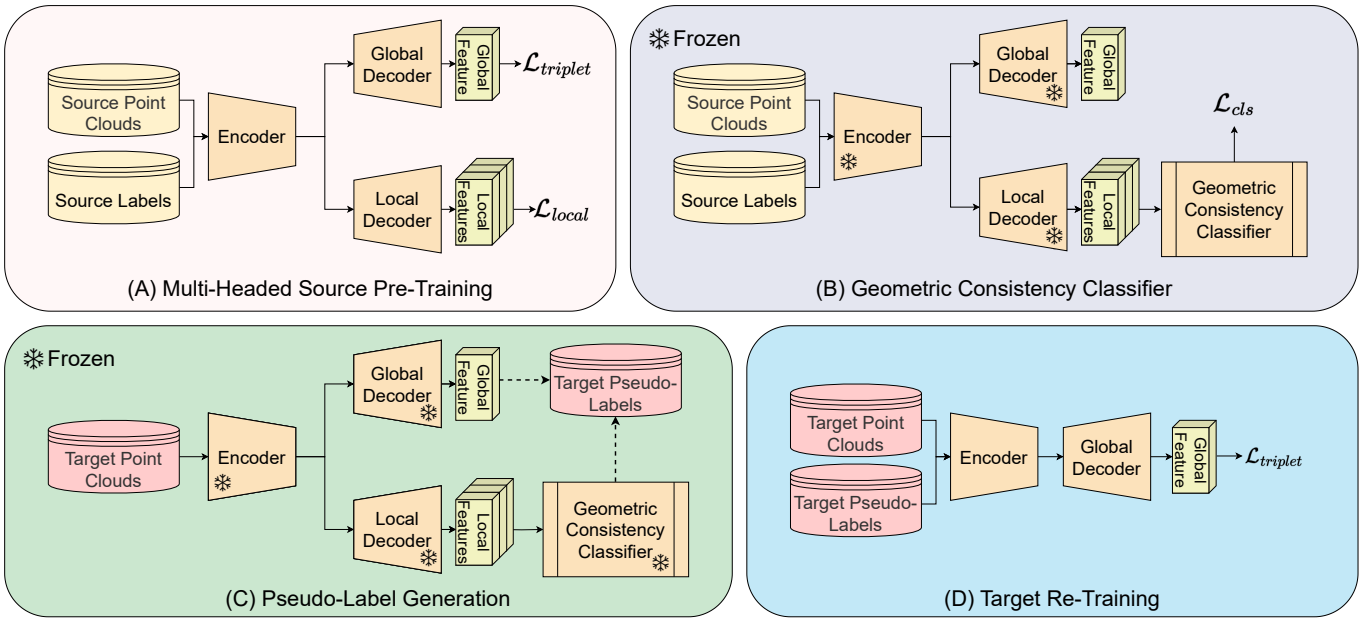


Fig. 1: Overview of *GeoAdapt*, our proposed Test-Time Adaptation approach. *GeoAdapt* adapts a model pre-trained on a source domain (A) by using a novel auxiliary classification head, which uses geometric consistency as a prior to classify a pair of clouds as positively or negatively associated (B). We then use the auxiliary head to generate pseudo-labels for pairs of point clouds on the target domain, using the L_2 similarity of the global features to guide which pairs to generate pseudo-labels for (C). These pseudo-labels are used to adapt the model's parameters to the target domain to improve test-time performance (D).

***Abstract*—LiDAR place recognition approaches based on deep learning suffer from significant performance degradation when there is a shift between the distribution of training and test datasets, often requiring re-training the networks to achieve peak performance. However, obtaining accurate ground truth data for new training data can be prohibitively expensive, especially in complex or GPS-deprived environments. To address this issue we propose *GeoAdapt*, which introduces a novel auxiliary classification head to generate pseudo-labels for re-training on unseen environments in a self-supervised manner. *GeoAdapt* uses geometric consistency as a prior to improve the robustness of our generated pseudo-labels against domain shift, improving the performance and reliability of our Test-Time Adaptation approach. Comprehensive experiments show that *GeoAdapt* significantly boosts place recognition performance across moderate to severe domain shifts, and is competitive with fully supervised test-time adaptation approaches. Our code is available at https://github.com/csiro-robotics/GeoAdapt.**

***Index Terms*—Place recognition, self-supervised, test-time adaptation.**

Manuscript received: August, 2, 2023; Revised October, 31, 2023; Accepted November, 20, 2023.

This paper was recommended for publication by Editor Cesar Cadena Lerma upon evaluation of the Associate Editor and Reviewers' comments.

[1] CSIRO Robotics and Autonomous Systems, DATA61, CSIRO, Australia. E-mails: firstname.lastname@csiro.au

[2] Signal Processing, AI and Vision Technologies (SAIVT), Queensland University of Technology (QUT), Brisbane, Australia. E-mails: {s.sridharan, c.fookes, peyman.moghadam}@qut.edu.au

Digital Object Identifier (DOI): see top of this page.

I. Introduction

PLACE recognition, the task of recognizing previously visited locations when traversing an environment, is a vital component of many applications in embodied intelligence and is essential for reliable loop closures in simultaneous localization and mapping (SLAM) or global relocalisation when deployed in a GPS-denied environment. The current state-of-the-art for place recognition is dominated by deep learning-based methods [1]–[5], which achieve remarkable performance when training and test data share a similar distribution. However, the performance of these learning-based approaches degrades significantly in the presence of domain shifts between the training and test data. While using a probabilistic approach can be sufficient to overcome this gap in the case of minor domain shifts [6], variations in the sensor, adversarial weather conditions, or major environmental shifts (*e.g.*, urban to natural) can pose significant challenges to the generalisability of LiDAR place recognition models, impacting their ability to be deployed over a wider range of platforms and environments [7].

A straightforward solution to addressing this problem is to refine a model by re-training using additional labelled samples on test environment. However, ground truth for training place recognition models is generally obtained using GPS or SLAM, leading to difficulties when re-training on environments which are either GPS-denied or sufficiently large

and complex enough to introduce failures into the SLAM solution. In addition, the source data initially used to train the model may not be available at test-time due to privacy or accessibility reasons. Ideally, the model should be adapted to the test data in a self-supervised manner, using only the pre-trained model parameters from the target domain. This setting is known as Test-Time Adaptation (TTA).

Test-Time Adaptation describes the task of refining a model pre-trained on a *source* dataset using an unlabelled *target* dataset, accommodating for the distribution shift between the source and target domains. Many existing TTA approaches [8]–[11] employ pseudo-labelling, where predictions from the source pre-trained model are used to guide re-training on the target dataset. However, pseudo-labels generated by the pre-trained model usually contain significant noise due to the domain gap between the source and target datasets, resulting in poor adaptation when they are used naively for re-training. In addition, TTA approaches for tasks such as segmentation and object detection are generally inapplicable to LiDAR place recognition due to fundamental differences in the nature of the ground truth required; for tasks in the former category, the ground truth denotes properties of the individual training samples such as semantic class or object locations, while for LiDAR place recognition the ground truth instead denotes the relationship between point clouds (*i.e.*, are they *positively* or *negatively* associated).

In this work, we propose a novel TTA approach for LiDAR place recognition which we call *GeoAdapt*, shown in Figure 1. *GeoAdapt* formulates the task of pseudo-label generation for place recognition as a classification problem, introducing a novel auxiliary classification head which classifies pairs of point clouds as positively or negatively associated in a registration-free manner for re-training without access to the target ground-truth. Specifically, the auxiliary classification head uses local feature matching to propose inlier correspondences between a pair of input point clouds and estimates the confidence of these correspondences using the geometric consistency of the input point clouds as a prior. We observe that positively associated point clouds are often geometrically consistent, while it is highly unlikely that the correspondences proposed for negatively associated point clouds will exhibit such geometrically consistent behaviour. By classifying the predicted confidence of our proposed inliers we significantly reduce the likelihood of noise in our generated pseudo-labels, resulting in reliable TTA without requiring any access to data or supervision from the source and target domains respectively.

We comprehensively benchmark *GeoAdapt* using five large-scale, public LiDAR datasets (two source and three target datasets), and demonstrate that our proposed approach consistently improves place recognition performance on out-of-distribution target domains. Notably, we show that our approach is competitive with re-training using ground truth supervision on the target domain, demonstrating the strength of using geometric consistency as the basis of self-supervision when generating labels for training. Our contributions are as follows:

- We introduce *GeoAdapt*, a novel Test-Time Adaptation approach which allows for re-training a model on unseen target environments without access to any ground truth supervision. Our approach employs a novel auxiliary classification head which uses geometric consistency as a prior to generate pseudo-labels in a registration-free manner which are robust against domain shifts, improving the effectiveness of our adaptation to target environments.
- We validate our method using five large-scale, public LiDAR datasets (2 source and 3 target datasets), and show our approach improves performance on unseen target domains by a large margin and achieves competitive performance with fully supervised adaptation.

II. Related Work

A. LiDAR Place Recognition

LiDAR Place Recognition (LPR) is formulated as a retrieval problem, where different methods encode a point cloud into a compact global descriptor which can be used to query a database of previously visited places. Existing approaches can be broadly categorised into handcrafted [12], [13], hybrid [14], [15], or fully end-to-end deep learning [1]–[5] approaches. Deep learning approaches in particular demonstrate remarkable performance when the training and test data share a similar distribution, but can have their generalisability on unseen data suffer significantly in the presence of a domain shift [16]. A key commonality of deep learning based approaches is the use of metric loss functions such as the triplet [1], [2], quadruplet [3], [17] or contrastive [4] loss, which are reliant on the selection of hard positive or negative examples using ground truth sensor pose information to form training tuples. Obtaining accurate sensor pose information is often done using either GPS or SLAM, with both approaches introducing challenges. Target environments may be GPS-denied, and even when available, GPS-based ground truth can result in false or noisy positive and negative examples when sensor occlusion or field of view are not taken into consideration [18]. Using SLAM to generate the ground truth suffers from the drawback of requiring accurate loop closures, which necessitates high-performing place recognition solutions beforehand.

B. Test-Time Adaptation

Test-Time Adaptation (TTA) aims to adapt a model pre-trained on a labelled source domain to improve performance on an unseen and unlabelled target domain, where there is a significant domain shift between the source and target. Related tasks include domain generalization [19], [20] which aims to train models using only the source data to perform broadly well across a variety of conditions, and online TTA [9], [21], which aims to adapt the model online from a stream of data during evaluation. A common approach [8]–[11] to TTA is to leverage predictions from the pre-trained model to predict pseudo-labels on the target domain, which are used for re-training the model to close the domain gap. These pseudo-labels are generally noisy as a result of the domain gap between source and target domains, with recent approaches using auxiliary heads [10] or uncertainty estimation [11] to refine the pseudo-labels for re-training. In this work, we propose an auxiliary classification head which employs geometric

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

priors relating to the geometric consistency of matching point clouds to produce robust and reliable pseudo-labels, allowing for effective target adaptation even under moderate to severe domain gaps.

Works exploring domain adaptation and generalization in 3D perception have focused primarily on the tasks of semantic segmentation [9], [19], [20], [22], [23] and object detection [24]–[27], with a particular focus on the domain gaps induced by either sim-to-real [28], [29] or changing LiDAR sensors [23] between the source and target. However, the problem of adaptation for LiDAR place recognition has gone largely unexplored in the literature to date. InCloud [16] re-trains the model on new domains to address the challenge of continual learning, but requires access to the target ground truth to do so. Continual SLAM [30] adapts the depth and odometry prediction of a SLAM system deployed in an unfamiliar environment, but freezes its loop closure (*i.e.*, place recognition) components during deployment. SGV [31] and Uncertainty-LPR [32] improve test-time performance by re-ranking the top retrieved candidates and filtering out uncertain examples respectively, but will still struggle when the top retrieved candidates are not well calibrated on the target domains. The closest existing work to ours is [33], which performs domain adaptation for visual place recognition but relies on SLAM to identify positive and negative examples for training. Specifically, they used pose graph optimization to calculate which pairs of images are positives or negatives, based on an initial set of potential place recognition matching pairs. However, as they note in their work, pose graph optimization can fail even with just a single incorrect loop closure. Therefore, during severe domain shifts, it is more likely that SLAM-based TTA will fail to find accurate positive and negative examples. Our approach does not suffer from this limitation, and can successfully adapt to severe domain shifts such as from urban to natural environments. Our proposed approach formulates pseudo-label generation for place recognition as a classification task and avoids using SLAM or any other pose estimation techniques as a prior, allowing for simple and effective adaptation of a model between source and target domains.

III. Problem Formulation

Presume we have a place recognition model pre-trained on a *source* dataset $\mathcal{S} = \{(\mathcal{P}_1, t_1), (\mathcal{P}_2, t_2), \cdots (\mathcal{P}_N, t_N)\}$, where $\mathcal{P} \in \mathbb{R}^{n\times 3}$ denotes a point cloud and t denotes the ground truth sensor pose for that point cloud. Now also presume we have a second *target* dataset $\mathcal{T} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots \mathcal{P}_N\}$ that contains point clouds with no accompanying ground truth. When there are significant shifts in the distribution of $\mathcal{S}$ and $\mathcal{T}$ due to differences in sensor, environment, or other factors, the model pre-trained on $\mathcal{S}$ will suffer from degraded place recognition performance on $\mathcal{T}$ as a result of the aforementioned domain shift. The aim of TTA is to adapt the parameters of a model pre-trained on $\mathcal{S}$ to improve performance on $\mathcal{T}$ without having any access to source data or ground truth supervision at test time.

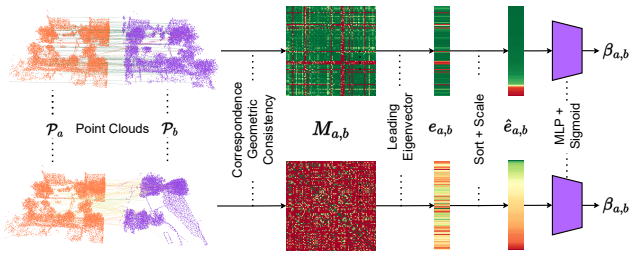


Fig. 2: Geometric Consistency Classifier. Positively associated point clouds (top) produce correspondences which are highly geometrically consistent, while negatively associated point clouds (bottom) produce correspondences with little to no consistency. The geometric consistency matrix $M_{a,b}$ is used to estimate the inlier confidence, and then to predict the pair-likelihood score for a given pair of input point clouds.

IV. GeoAdapt

In this section we propose *GeoAdapt*, a TTA approach for LiDAR place recognition which introduces a novel auxiliary classification head which outputs a pair-likelihood estimation for any given pair of input point clouds. TTA using *GeoAdapt* is split into four different steps. First, we pre-train a model on $\mathcal{S}$ to produce both a global descriptor and dense local features from different encoders. Secondly, we freeze the model and train our proposed Geometric Consistency Classifier using the frozen local features, which uses geometric consistency as a prior in order to produce pair-likelihood scores which are highly robust against domain shift. Thirdly, we use the pair-likelihood scores produced by the auxiliary classifier to generate pseudo-labels for $\mathcal{T}$, and finally we use these pseudo-labels to re-train the model on $\mathcal{T}$ without requiring source data or target supervision. Figure 1 provides an overview of our model design and adaptation pipeline.

A. Multi-Headed Source Pre-Training

The backbone of our approach is a multi-headed model architecture consisting of a global feature head and an auxiliary classification head. The global feature head consists of a shared encoder and a shallow decoder which outputs a global feature vector for an input point cloud, while the auxiliary classification head consists of the shared encoder, a local feature decoder which outputs dense point features for the input point cloud, and our proposed Geometric Consistency Classifier. Before training the Geometric Consistency Classifier, we pre-train the shared encoder and both the local and global feature decoders on the source dataset. To train the global feature we use the triplet loss which is defined as:

$$\mathcal{L}_{triplet} = \left[\|g_{anc} - g_{pos}\|_2 - \|g_{anc} - g_{neg}\|_2 + \delta\right]_+, \quad (1)$$

where $g_{anc}, g_{pos}, g_{neg} \in \mathbb{R}^{256}$ are the global feature associated with anchor, positive and negative point clouds $\mathcal{P}_{anc}, \mathcal{P}_{pos}, \mathcal{P}_{neg}$, respectively, $[\cdot]_+$ denotes the hinge loss and δ is a margin hyperparameter. In order to select our positive and negative point clouds we rely on the label $y_{a,b}$ which indicates if two point clouds $\mathcal{P}_a, \mathcal{P}_b$ are positive associated,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

negatively associated, or neither. $y_{a,b}$ is generated using the ground truth sensor pose such that:

$$y_{a,b} = \begin{cases} \text{Positive} & s\left(t_a, t_b\right) \leq T_{pos} \\ \text{Negative} & s\left(t_a, t_b\right) \geq T_{neg} \\ \text{Neither} & \text{Otherwise} \end{cases}, \tag{2}$$

where $s\left(t_a, t_b\right)$ denotes the distance between the ground truth sensor poses for $\mathcal{P}_a, \mathcal{P}_b$ and T_{pos}, T_{neg} denote positive and negative distance thresholds in the world.

For our local features $l \in \mathbb{R}^{|\mathcal{P}| \times 16}$, we follow [3] and use a combination of ground truth pose and ICP [34] to align two positively associated point clouds $\mathcal{P}_a$, $\mathcal{P}_b$ and retrieve a set of point correspondences $C^{a \leftrightarrow b}$, where each correspondence is denoted as $c_i \in C^{a \leftrightarrow b} = \left\{\left(x_a^i, l_a^i\right), \left(x_b^i, l_b^i\right)\right\}$ with $x_a^i, l_a^i, x_b^i, l_b^i$ denoting the 3D point co-ordinate and local feature from $\mathcal{P}_a$,$\mathcal{P}_b$ linked by correspondence c_i. The local features are then trained using the hardest contrastive loss:

$$\begin{aligned} \mathcal{L}_{local} = \sum_{c_i \in C^{a \leftrightarrow b}} \Bigg\{ & \left[\left\| l_a^i - l_b^i \right\|_2^2 - m_p \right]_+ / \left| C^{a \leftrightarrow b} \right| \\ & + \lambda_n \left[m_n - \min_{k \in \mathcal{M}} \left\| l_a^i - l_b^k \right\|_2^2 \right]_+ / \left| C^{a \leftrightarrow b} \right| \\ & + \lambda_n \left[m_n - \min_{k \in \mathcal{M}} \left\| l_b^i - l_a^k \right\|_2^2 \right]_+ / \left| C^{a \leftrightarrow b} \right| \Bigg\}, \end{aligned} \tag{3}$$

where $\mathcal{M}$ is a random subset of features used for hard negative mining, hyperparameters m_p, m_n are scalar margins and λ_n is a scalar weight. This gives us a combined training loss for the global and local features of:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{triplet} + \mathcal{L}_{local}, \tag{4}$$

where $\mathcal{L}_{triplet}$ and $\mathcal{L}_{local}$ are the losses to train the local and global features respectively.

B. Geometric Consistency Classifier

The second component of our auxiliary classification head is our proposed Geometric Consistency Classifier (GCC), which we train on top of the frozen local feature extractor outlined in the previous section. While local point features are no less susceptible to domain shift than global point cloud embeddings, by leveraging the geometric consistency of inlier correspondences as a prior our proposed GCC is able to produce reliable and accurate pseudo-labels for target adaptation in a manner which is robust against the deleterious impact of domain shift between the source and target datasets.

Given two point clouds $\mathcal{P}_a$, $\mathcal{P}_b$ we extract local features l_a, l_b before using a nearest neighbour search on the local features to produce a set of proposed point correspondences between the two point clouds $\hat{C}^{a \leftrightarrow b}$. Next, we construct a geometric consistency matrix $M_{a,b}$ for which each entry $m_{i,j}$ measures the pairwise length consistency between correspondences c_i, c_j in $\hat{C}^{a \leftrightarrow b}$, which is defined as:

$$m_{i,j} = \left[1 - \frac{d_{i,j}^2}{d_{thr}^2} \right]_+, d_{i,j} = \left| \left\| x_a^i - x_a^j \right\|_2 - \left\| x_b^i - x_b^j \right\|_2 \right|, \tag{5}$$

where d_{thr}^2 is a hyperparameter which controls sensitivity to length difference. As seen in Figure 2 when two point clouds are positively correlated $M_{a,b}$ is dominated by a large cluster of geometrically consistent inliers, while no such cluster exists for negatively associated point clouds. Inspired by [35] we can consider the leading eigenvector $e_{a,b} \in \mathbb{R}^{\left|\hat{C}^{a \leftrightarrow b}\right|}$ of $M_{a,b}$ to be the association of each correspondence $c_i \in \hat{C}^{a \leftrightarrow b}$ with the main cluster of $M_{a,b}$, and therefore a good estimation for inlier probability. We observe in Figure 2 that not only is the distribution of values in $e_{a,b}$ clearly distinct between positively and negatively associated point clouds, but also that the chances of two negatively associated point clouds forming a large spatially consistent cluster in order to mimic the distribution of a positive match is extraordinarily small. Therefore, we feed the inlier probability into an MLP to produce a pair-likelihood score as follows:

$$\beta_{a,b} = \sigma\left(f\left(\hat{e}_{a,b}; \theta\right)\right), \tag{6}$$

where $\beta_{a,b}$ is the pair-likelihood score for point clouds $\mathcal{P}_a$, $\mathcal{P}_b$, $f\left(\cdot; \theta\right)$ is an MLP parameterised by θ, σ is the sigmoid function, and $\hat{e}_{a,b}$ is $e_{a,b}$ with values sorted and scaled into the range $(0, 1)$. Unlike common registration methods such as ICP [34] the GCC does not require any initial alignment in order to reliably associate the input clouds, leading to more reliable detection of orthogonal or reverse revisits which provide valuable hard positive examples to the pseudo-label set. To train the GCC we use the binary cross-entropy loss:

$$\mathcal{L}_{cls} = -\left(y_{a,b} \cdot \log \beta_{a,b} + \left(1 - y_{a,b}\right) \log\left(1 - \beta_{a,b}\right)\right), \tag{7}$$

where $y_{a,b}$ is 1 for a positive pair and 0 otherwise. The GCC is trained on $\mathcal{S}$ where there is full access to the ground-truth supervision, but requires no re-training or ground truth when used for pseudo-label generation.

C. Pseudo-Label Generation

Now with both the global and auxiliary classification heads fully trained on $\mathcal{S}$, we can generate pseudo-labels for adaptation to $\mathcal{T}$. For each target point cloud $\mathcal{P}_a \in \mathcal{T}$ we extract global and local features g_a, l_a. Next for each point cloud, we use the L_2 distance between global descriptors to retrieve the top-K nearest neighbours in the feature space as a list of potential positive and negative candidate point clouds, $\{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_K\}$. We then generate pseudo-labels for each point cloud and its list of candidates as follows:

$$y'_{a,b} = \begin{cases} \text{Positive} & \beta_{a,b} \geq \alpha_{pos} \\ \text{Negative} & \beta_{a,b} \leq \alpha_{neg} \\ \text{Neither} & \text{Otherwise} \end{cases}, \tag{8}$$

where $\alpha_{pos}, \alpha_{neg}$ are scalar thresholds used to define whether the pair-likelihood score indicates a positive or negative example, and $b \in \{1, 2, \cdots K\}$ is the index of the candidate point cloud. The advantage of implementing these confidence thresholds is that underneath a severe domain shift, the pre-trained model features often may not be informative enough to reliably identify how correlated two input point clouds are.

TABLE I: Details of source and target training and evaluation sets. * indicates shared training set for Wild-Places [36].

Dataset	Setting	Sensor	Number of Scans
Source:			(train)
MulRan [37]	Urban	OS1-64	29,112
KITTI [38]	Urban	HDL-64E	23,201
Target (Moderate):			(train / query / database)
ALITA [39]	Urban	VLP-16	15,439 / 451 / 576
NCLT [40]	Campus	HDL-32E	20,503 / 4,486 / 26,945
Target (Severe):			
Wild-Places (V) [36]	Natural	VLP-16	19,096* / 6,395 / 23,472
Wild-Places (K) [36]	Natural	VLP-16	19,096* / 9,642 / 39,847

The confidence thresholds allows us to identify when the GCC is uncertain about a pair of point clouds and exclude them from the set of pseudo-labels used for re-training, improving the reliability of our test-time adaptation.

D. Target Re-Training

When adapting the model to the target dataset we discard the local feature decoder and GCC, re-training only the encoder and global feature decoder. The model is fine-tuned (from source pre-trained weights) with Equation 1, using the generated pseudo-labels $y'_{a,b}$ in place of ground-truth supervision. If for a point cloud $\mathcal{P}_a \in \mathcal{T}$ either no positive or no negative examples were found in the previous step, then that point cloud is excluded as an anchor during re-training.

V. Experimental Settings

A. Training and Implementation Details

We train our approach on 4 NVIDIA Tesla P100-16GB GPUs, and follow the data augmentation strategy proposed by [1]. We set the hyperparameters $\alpha_{pos}, \alpha_{neg}$ to 0.95 and 0.2 respectively for all experiments based on our ablations in Section VI-C, and the number of considered candidates K to 50. We use a sparse-convolutional architecture with skip connections for our encoder and decoders, and GeM [41] pooling in our global feature head. For the source datasets we train the model from scratch for 80 epochs with a learning rate of 1e-3 decayed by a factor of 10 at 30 and 60 epochs, and the GCC is trained for 5 epochs using a learning rate of 0.01 decayed with a cosine scheduler. When re-training we fine-tune the pre-trained model for 40 epochs at a learning rate of 1e-4, decayed by a factor of 10 at 25 epochs.

B. Datasets

We evaluate our proposed approach using two source datasets and three target datasets. **Source:** We use the MulRan [37] and KITTI Odometry [38] datasets for source pre-training, both of which were collected using vehicle-mounted sensors in urban environments. We follow [3] and use a combination of MulRan sequences 01 and 02 from the DCC environment and sequences 01 and 03 from the Riverside environment for training, and use all 11 KITTI sequences when training. **Target:** We further subdivide our target datasets into two categories: those experiencing a 'moderate' domain shift, representing urban-to-urban or urban-to-campus TTA, and those experiencing a 'severe' domain shift, representing the much more challenging urban-to-natural TTA scenario. For the 'moderate' datasets we use the ALITA [39] and NCLT [40] datasets. For ALITA we use the training, query and database splits proposed by the authors of the dataset, while for NCLT we use sequences (2012-01-08, 2012-01-15) for training and sequences (2012-11-04, 2012-11-16, 2012-11-17, 2012-12-01, 2013-02-23, 2013-04-05) for evaluation, following the same query/database split used in [15]. For the 'severe' target datasets we use the recently introduced Wild-Places [36] benchmark dataset, a natural place recognition dataset recorded across two large-scale natural environments. We use Wild-Places (V) and Wild-Places (K) to refer to the *Venman* and *Karawatha* environments in the Wild-Places dataset respectively, and use the *inter-sequence* training and evaluation setup established by the authors. Table I outlines the environment, sensors and training splits for each of the source and target datasets used in our experiments.

C. Evaluation Metrics

For evaluation, we calculate the L_2 distance between the global embedding for a query point cloud and point clouds from a database consisting of different traversals of the same region from the database. We report mean Recall@N (R@N) for N = 1, 5, 1%, considering a query point cloud to be successfully relocalised if one of the top-N retrieved database candidates is within 5m for ALITA [39] and within 3m for all other datasets. We use SOURCE→TARGET to denote the source and target datasets for a given result.

VI. Results

A. Comparison to State-of-the-Art

Tables II and III compare the performance of *GeoAdapt* to existing state-of-the-art approaches for place recognition for 'moderate' and 'severe' domain shifts respectively. We compare against one handcrafted approach, ScanContext [12], and two state-of-the-art learning-based approaches, MinkLoc3Dv2 [2] and LoGG3D-Net [3]. As previously mentioned TTA approaches for tasks such as segmentation or object detection are generally not applicable to LiDAR place recognition due to fundamental differences in the ground truth required for training, and to the best of our knowledge no other approach for Test-Time Adaptation on LiDAR place recognition exists in the literature. Therefore we also compare against SGV [31], which performs re-ranking on the target to improve performance at test time without requiring re-training or access to target supervision.

We observe that the exclusively source pre-trained methods report significantly diminished results when evaluated on out-of-domain data, achieving a maximum R@1 of only 66.64% and 9.36% on the 'moderate' and 'severe' target datasets respectively. By adapting the model to the target data distribution *GeoAdapt* is able to maintain high performance on the target datasets without having required access to any ground truth information during adaptation, reporting a R@1 performance of 83.99% (+17.35%) and 50.56% (+41.2%) on the 'moderate' and 'severe' datasets. These results illustrate the impact that the domain shifts induced by a change in environment and LiDAR sensor can have on model generalisability, and the effectiveness of *GeoAdapt* in addressing this challenge.

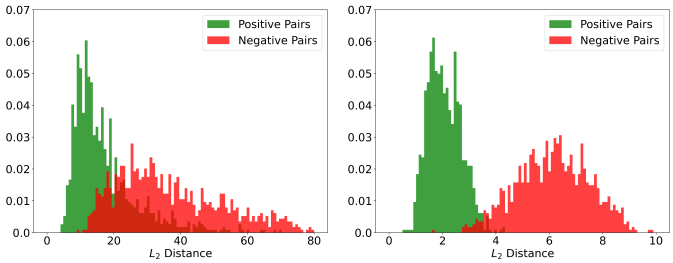
IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

TABLE II: Test-Time Adaptation performance for LiDAR place recognition under "moderate" domain shifts. Rows above and below the dashed line represent source-only and target-adapted approaches respectively. Bold and underlined values are first and second place respectively for a given column.

Method	KITTI→ALITA			KITTI→NCLT			MulRan→ALITA			MulRan→NCLT			Average		
	R@1	R@5	R@1%	R@1	R@5	R@1%	R@1	R@5	R@1%	R@1	R@5	R@1%	R@1	R@5	R@1%
ScanContext [12]	56.03	77.07	59.59	64.74	69.54	74.88	56.03	77.07	59.59	64.74	69.54	74.88	60.39	73.31	67.24
MinkLoc3Dv2 [2]	72.74	92.75	77.44	54.58	71.38	84.42	70.21	91.05	75.08	69.01	86.39	**95.70**	66.64	85.39	83.16
LoGG3D-Net [3]	59.68	82.76	64.72	41.80	58.43	73.37	71.19	92.57	75.26	33.28	47.12	58.95	52.02	70.41	68.98
LoGG3D-Net+SGV [31]	88.32	91.97	88.64	57.82	67.36	73.37	93.24	96.83	93.9	50.48	56.48	58.95	72.47	78.16	78.72
GeoAdapt	**96.95**	**100.0**	**98.25**	69.45	85.55	**94.63**	**97.56**	**100.0**	**98.69**	71.98	88.03	95.68	83.99	93.40	**96.81**
GeoAdapt+SGV	96.91	100.0	98.22	**82.83**	**91.41**	**94.63**	97.13	99.83	98.60	**83.82**	**91.86**	95.68	**90.17**	**95.78**	96.78

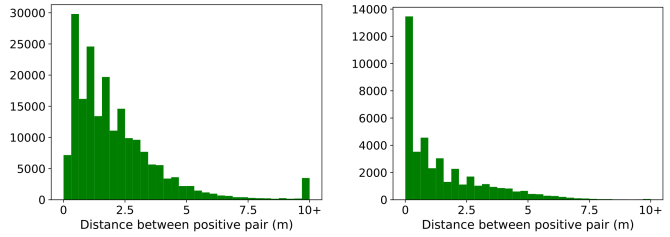
TABLE III: Test-Time Adaptation performance for LiDAR place recognition under "severe" domain shifts.

Method	KITTI→Wild-Places(V)			KITTI→Wild-Places(K)			MulRan→Wild-Places(V)			MulRan→Wild-Places(K)			Average		
	R@1	R@5	R@1%	R@1	R@5	R@1%	R@1	R@5	R@1%	R@1	R@5	R@1%	R@1	R@5	R@1%
ScanContext [12]	33.98	48.79	61.56	38.44	53.56	66.16	33.98	48.79	61.56	38.44	53.56	66.16	36.21	51.18	63.86
MinkLoc3Dv2 [2]	5.40	17.15	41.82	6.49	18.79	42.53	11.18	26.47	50.01	14.35	33.32	58.98	9.36	23.93	48.36
LoGG3D-Net [3]	6.00	15.13	32.54	8.55	19.43	37.26	3.91	11.69	27.97	5.16	14.07	30.43	5.91	15.08	32.05
LoGG3D-Net+SGV [31]	18.84	23.89	32.54	21.85	28.34	37.26	14.8	20.94	27.97	15.56	22.8	30.43	17.76	23.99	32.05
GeoAdapt	60.48	84.93	**96.30**	47.95	75.23	**93.02**	51.99	77.26	**91.82**	41.45	69.66	**89.54**	50.56	76.77	**92.67**
GeoAdapt+SGV	**74.12**	**90.74**	96.30	**58.68**	**81.09**	93.02	**70.34**	**86.96**	91.82	**56.39**	**78.04**	89.54	**64.88**	**84.21**	92.67



(a) Before *GeoAdapt* (b) After *GeoAdapt*

Fig. 3: Comparison of L_2 distance between positive and negative pairs of global embeddings for the source pre-trained model (a) and *GeoAdapt* (b). *GeoAdapt* significantly improves contrast between positive and negative pairs on the target domain after re-training, leading to more reliable revisit detection and retrieval. Results reported on KITTI→Wild-Places (K).



(a) KITTI→NCLT (b) KITTI→Wild-Places (K)

Fig. 4: Histogram of distances between positive pairs of point clouds selected by *GeoAdapt*. *GeoAdapt* generates pseudo-labels based on geometric consistency rather than a scalar distance threshold, resulting in a more flexible positive selection for target re-training.

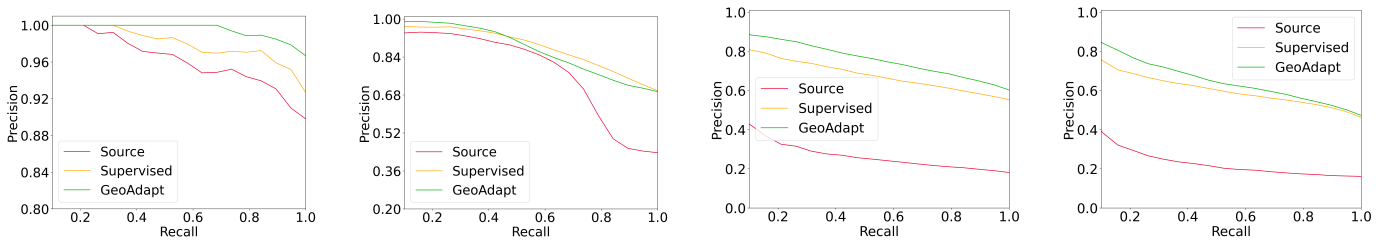
Comparing to SGV [31], we observe that when applied to LoGG3D-Net [3] SGV improves performance by 20.45% and 11.85% on the 'moderate' and 'severe' target datasets respectively but is still out-performed by *GeoAdapt*. The upper bound of performance for re-ranking methods is limited by the quality of the proposed re-ranking candidates, which can be an issue when - as particularly apparent in Table III - domain shifts severely degrade the quality of the initial proposals. We observed that the highest performance overall comes from combining *GeoAdapt* with SGV, "bootstrapping" the model on the target domain to improve the quality of the re-ranking candidates and boosting the performance of *GeoAdapt* by an additional 6.18% and 14.32% on the 'moderate' and 'severe' target datasets respectively.

B. Comparison to Pre-Trained and Fully Supervised Models

For place recognition, the model's feature space should demonstrate strong separability between global embeddings which are positively and negatively associated. Domain shifts between the source and target data can severely degrade this separability, which in turn degrades our ability to use the global features to identify and retrieve place re-visits at evaluation. In Figure 3 we show that re-training with *GeoAdapt* significantly improves the separability of positive and negative examples on the target, leading to more reliable revisit detection and retrieval. Figure 5 presents precision-recall curves for a source pre-trained model, a model re-trained with *GeoAdapt*, and a model re-trained with the supervised ground truth on the target. We observe that *GeoAdapt* not only clearly outperforms the source pre-trained model, but also performs competitively with and sometimes out-performs the model re-trained using the supervised ground truth.

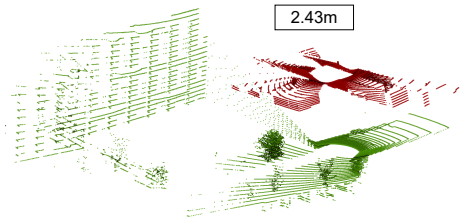
We present several explanations for this behaviour. Firstly, as has been raised in previous works [18] the ground truth for place recognition datasets often contain errors due to noise in the GPS or SLAM-derived odometry ground truth, leading to faulty selection of positive and negative pairs which detrimentally impact the model's performance when used for re-training. Secondly, as shown in Figure 6 using a hard scalar threshold to form pairs for training presents some drawbacks even assuming a perfect ground truth pose. Occlusion or other environmental factors can result in 'positive' pairs of nearby point clouds which have little to no shared information (Figure 6a), and highly correlated point clouds separated by a distance just slightly higher than the scalar threshold can be excluded from training despite having a great deal of shared information (Figure 6b).

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

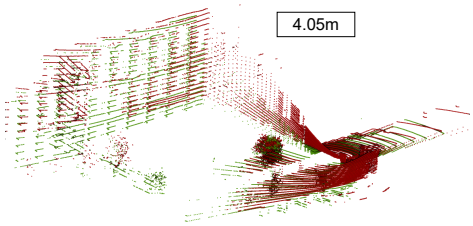


(a) KITTI→ALITA (b) KITTI→NCLT (c) KITTI→Wild-Places (V) (d) KITTI→Wild-Places (K)

Fig. 5: Precision-Recall curves for models pre-trained on source data, re-trained with supervised ground truth, and re-trained with *GeoAdapt*. *GeoAdapt* outperforms not only the source pre-training but also in some cases ground truth supervision.



(a) Positive pair selected by supervised ground truth



(b) Positive pair selected by *GeoAdapt*

Fig. 6: Examples of positive examples for supervised re-training and *GeoAdapt*. *GeoAdapt* removes faulty positive pairs included by the supervised ground truth, and includes positives the supervised ground truth would otherwise exclude.

Comparatively, in Figure 4 we show that the positives selected by *GeoAdapt* follow a dynamic threshold guided by the geometric similarity of the inputs, which is effective in tackling the edge cases shown in Figure 6. Our results suggest that the flexibility provided by using a learned approach to guide positive and negative selection can give an advantage over using an inflexible threshold on the target's supervised ground truth, though we leave further investigation of this phenomena for future work.

C. Ablation Studies

In this section we explore the impact of how certain design choices and hyperparameter selection impacts the performance of the proposed approach. Table IV looks at the impact of pre-processing the pair-likelihood vector $e_{a,b}$ before its use as input to the MLP. We observe that both sorting and scaling of $e_{a,b}$ are critical to the performance of the GCC, with Recall@1 performance dropping 23.54% on the KITTI→Wild-Places (K) setup in their absence. Figure 7 explores the impact of changing the value of hyperparameters $\alpha_{pos}, \alpha_{neg}$ on the KITTI→Wild-Places (K) scenario. We observe that while employing a stricter value of α_{pos} has a notably positive impact on adaptation performance, no strong trend is present for changing the value of α_{neg}. This result suggests that false

TABLE IV: Ablation study on the effects of scaling and sorting the inlier confidence scores. Results are reported on KITTI→Wild-Places (K).

Scaling	Sorting	R@1	R@5	R@1%
-	-	24.41	49.14	72.21
✓	-	38.36	66.32	88.29
-	✓	45.19	72.68	92.2
✓	✓	**47.95**	**75.23**	**93.02**

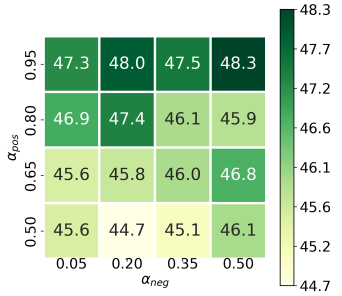


Fig. 7: Impact of α_{pos}, α_{neg} (Recall@1) on KITTI→Wild-Places (K).

positive examples have a significantly more deleterious impact on adaptation than false negatives, emphasising the importance of filtering them out of the set of pseudo-labels for the target data. We found that values for α_{pos} and α_{neg} selected through this ablation produce consistently strong results across all experimental setups as demonstrated by the results in Tables II and III, such that choosing new thresholds for new target datasets is not necessary.

VII. Conclusion

In this paper we address the task of Test-Time Adaptation for LiDAR place recognition, adapting a source pre-trained model in a self-supervised manner to improve performance on an unlabelled target environment. We propose a novel approach *GeoAdapt*, which re-trains the model using robust pseudo-labels generated by an auxiliary classification head. We demonstrate that geometric consistency as a prior when generating pseudo-labels results in strong and reliable adaptation to target environments underneath even severe domain shifts, and even performs competitively with fully supervised re-training on the target domain. The ability of *GeoAdapt* to outperform supervised re-training highlights several shortcomings in how ground truth is currently generated for place recognition datasets, and raises interesting avenues for future work investigating how geometric consistency can be used as

a prior to guide both supervised and self-supervised learning in LiDAR place recognition. We believe that this work opens several avenues for future exploration of TTA for localisation tasks. These avenues include targeting specific domain shifts induced by factors such as changing sensor or weather effects, investigating how *GeoAdapt* could be used to perform large-scale data annotation for city or larger scale place recognition, extending the work to TTA for 6-DoF metric localisation and addressing the additional challenges posed by self-supervised adaptation in the continual [16] or online [30] settings.

ACKNOWLEDGEMENTS

This work was partially funded by CSIRO's Machine Learning and Artificial Intelligence Future Science Platform (MLAI FSP). The work was supported in part by an Australian Research Council (ARC) Discovery Program Grant No: DP200101942.

REFERENCES

[1] Komorowski, Jacek, "MinkLoc3D: Point Cloud Based Large-Scale Place Recognition," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021.

[2] Jacek Komorowski, "Improving point cloud based place recognition with ranking-based loss and large batch training," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 3699–3705.

[3] K. Vidanapathirana, M. Ramezani *et al.*, "LoGG3D-Net: Locally Guided Global Descriptor Learning for 3D Place Recognition," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[4] L. Wiesmann, L. Nunes *et al.*, "KPPR: Exploiting momentum contrast for point cloud-based place recognition," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 592–599, 2022.

[5] L. Hui, H. Yang *et al.*, "Pyramid Point Cloud Transformer for Large-Scale Place Recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6098–6107.

[6] M. U. M. Bhutta, Y. Sun *et al.*, "Why-so-deep: Towards boosting previously trained models for visual place recognition," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1824–1831, 2022.

[7] M. Ramezani, E. Griffiths *et al.*, "Deep Robust Multi-Robot Relocalisation in Natural Environments," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.

[8] J. Liang, D. Hu *et al.*, "Do We Really Need to Access the Source Data? Source Hypothesis Transfer for Unsupervised Domain Adaptation," in *International Conference on Machine Learning*, 2020.

[9] I. Shin, Y.-H. Tsai *et al.*, "MM-TTA: Multi-Modal Test-Time Adaptation for 3D Semantic Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 16 928–16 937.

[10] Y. Zhang, S. Borse *et al.*, "AuxAdapt: Stable and Efficient Test-Time Adaptation for Temporally Consistent Video Semantic Segmentation," *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2633–2642, 2021.

[11] M. Litrico, A. Del Bue *et al.*, "Guiding Pseudo-labels with Uncertainty Estimation for Source-free Unsupervised Domain Adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[12] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.

[13] G. Kim, S. Choi *et al.*, "Scan Context++: Structural Place Recognition Robust to Rotation and Lateral Variations in Urban Environments," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1856–1874, 2022.

[14] K. Vidanapathirana, P. Moghadam *et al.*, "Locus: LiDAR-based Place Recognition using Spatiotemporal Higher-Order Pooling," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[15] X. Xu, H. Yin *et al.*, "Disco: Differentiable scan context with orientation," *IEEE Robotics and Automation Letters*, 2021.

[16] J. Knights, P. Moghadam *et al.*, "InCloud: Incremental Learning for Point Cloud Place Recognition," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[17] M. A. Uy and G. H. Lee, "PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[18] E. Brachmann, M. Humenberger *et al.*, "On the Limits of Pseudo Ground Truth in Visual Camera Re-localisation," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6198–6208, 2021.

[19] H. Kim, Y. Kang *et al.*, "Single Domain Generalization for LiDAR Semantic Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 17 587–17 598.

[20] A. Xiao, J. Huang *et al.*, "3D Semantic Segmentation in the Wild: Learning Generalized Models for Adverse-Condition Point Clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 9382–9392.

[21] Q. Wang, O. Fink *et al.*, "Continual Test-Time Domain Adaptation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7191–7201, 2022.

[22] C. Saltori, F. Galasso *et al.*, "Cosmix: Compositional Semantic Mix for Domain Adaptation in 3D Lidar Segmentation," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 586–602.

[23] K. Ryu, S. Hwang *et al.*, "Instant Domain Augmentation for LiDAR Semantic Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 9350–9360.

[24] Y. Wei, Z. Wei *et al.*, "LiDAR Distillation: Bridging the Beam-Induced Domain Gap for 3D Object Detection," in *European Conference on Computer Vision (ECCV)*, 2022, pp. 179–195.

[25] A. Lehner, S. Gasperini *et al.*, "3D-VField: Adversarial Augmentation of Point Clouds for Domain Generalization in 3D Object Detection," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17 274–17 283, 2021.

[26] Y. Wang, J. Yin *et al.*, "SSDA3D: Semi-supervised Domain Adaptation for 3D Object Detection from Point Cloud," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[27] J. Yang, S. Shi *et al.*, "ST3D: Self-training for Unsupervised Domain Adaptation on 3D Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[28] R. Ding, J. Yang *et al.*, "DODA: Data-Oriented Sim-to-Real Domain Adaptation for 3D Semantic Segmentation," in *European Conference on Computer Vision (ECCV)*, S. Avidan, G. Brostow *et al.*, Eds., 2022, pp. 284–303.

[29] S. Huch, L. Scalerandi *et al.*, "Quantifying the LiDAR Sim-to-Real Domain Shift: A Detailed Investigation Using Object Detectors and Analyzing Point Clouds at Target-Level," *IEEE Transactions on Intelligent Vehicles*, 2023.

[30] N. Vodisch, D. Cattaneo *et al.*, "Continual SLAM: Beyond Lifelong Simultaneous Localization and Mapping through Continual Learning," in *International Symposium of Robotics Research*, 2022.

[31] K. Vidanapathirana, P. Moghadam *et al.*, "Spectral Geometric Verification: Re-Ranking Point Cloud Retrieval for Metric Localization," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2494–2501, 2023.

[32] K. Mason, J. Knights *et al.*, "Uncertainty-Aware Lidar Place Recognition in Novel Environments," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.

[33] P.-Y. Lajoie and G. A. Beltrame, "Self-Supervised Domain Calibration and Uncertainty Estimation for Place Recognition," *IEEE Robotics and Automation Letters*, vol. 8, pp. 792–799, 2022.

[34] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[35] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005.

[36] J. Knights, K. Vidanapathirana *et al.*, "Wild-Places: A Large-Scale Dataset for Lidar Place Recognition in Unstructured Natural Environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 11 322–11 328.

[37] G. Kim, Y. S. Park *et al.*, "MulRan: Multimodal Range Dataset for Urban Place Recognition," in *2020 IEEE International Conference on Robotics and Automation*, 2020.

[38] A. Geiger, P. Lenz *et al.*, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[39] P. Yin, S. Zhao *et al.*, "ALITA: A Large-scale Incremental Dataset for Long-term Autonomy," *arXiv preprint arXiv:2205.10737*, 2022.

[40] N. Carlevaris-Bianco, A. K. Ushani *et al.*, " University of Michigan North Campus long-term vision and lidar dataset ," *International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2015.

[41] F. Radenović, G. Tolias *et al.*, "Fine-Tuning CNN Image Retrieval with No Human Annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1655–1668, 2019.