

Mining and Transferring Feature-Geometry Coherence for Unsupervised Point Cloud Registration

Kezheng Xiong^{ab}, Haoen Xiang^{ab}, Qingshan Xu^c, Chenglu Wen^{ab*}, Siqi Shen^{ab},
Jonathan Li^d, Cheng Wang^{ab}

^aFujian Key Laboratory of Sensing and Computing for Smart Cities,
Xiamen University, China.

^bKey Laboratory of Multimedia Trusted Perception and Efficient Computing,
Ministry of Education of China, Xiamen University, China.

^cNanyang Technological University, Singapore.

^dUniversity of Waterloo, Waterloo, Canada

{xiongkezheng, haoenxiang}@stu.xmu.edu.cn

{clwen, siqishen, cwang}@xmu.edu.cn

qingshan.xu@ntu.edu.sg, junli@uwaterloo.ca

Abstract

Point cloud registration, a fundamental task in 3D vision, has achieved remarkable success with learning-based methods in outdoor environments. Unsupervised outdoor point cloud registration methods have recently emerged to circumvent the need for costly pose annotations. However, they fail to establish reliable optimization objectives for unsupervised training, either relying on overly strong geometric assumptions, or suffering from poor-quality pseudo-labels due to inadequate integration of low-level geometric and high-level contextual information. We have observed that in the feature space, latent new inlier correspondences tend to cluster around respective positive anchors that summarize features of existing inliers. Motivated by this observation, we propose a novel unsupervised registration method termed INTEGER to incorporate high-level contextual information for reliable pseudo-label mining. Specifically, we propose the Feature-Geometry Coherence Mining module to dynamically adapt the teacher for each mini-batch of data during training and discover reliable pseudo-labels by considering both high-level feature representations and low-level geometric cues. Furthermore, we propose Anchor-Based Contrastive Learning to facilitate contrastive learning with anchors for a robust feature space. Lastly, we introduce a Mixed-Density Student to learn density-invariant features, addressing challenges related to density variation and low overlap in the outdoor scenario. Extensive experiments on KITTI and nuScenes datasets demonstrate that our INTEGER achieves competitive performance in terms of accuracy and generalizability. [Code Release]

1 Introduction

Point cloud registration is a fundamental task in autonomous driving and robotics. It aims to align two partially overlapping point clouds with a rigid transformation. Learning-based methods have achieved remarkable success in outdoor point cloud registration[1–5]. PCAM[1] pioneered the integration of low-level geometric and high-level contextual information, inspiring subsequent works[2–5]. However, these supervised methods suffer from poor generalizability and reliance on costly pose

*Corresponding author

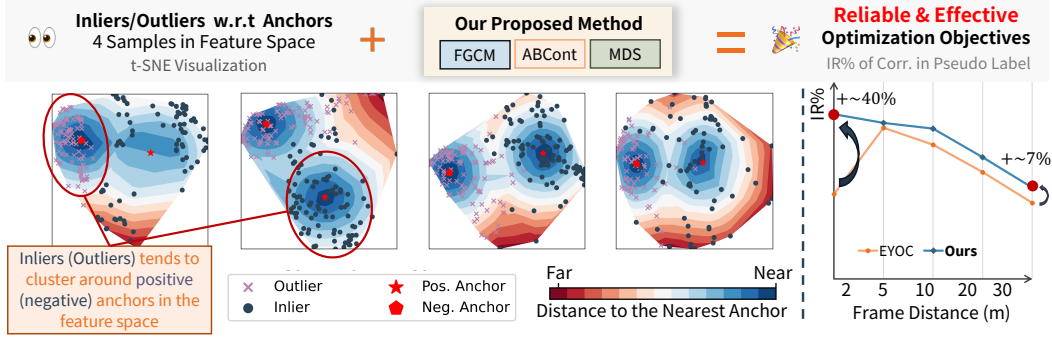


Figure 1: (1) **Motivation**: new inliers (outliers) tend to cluster around latent *positive (negative)* anchors that represent existing inliers (outliers) in the feature space, respectively. (2) **Performance**: pseudo-labels from INTEGER are more robust and accurate than the previous state-of-the-art EYOC[12].

annotations[6–8], underscoring the need for unsupervised methods to address these challenges in real-world applications.

Despite recent progress[9–12] in unsupervised registration methods, the task remains challenging and underexplored, especially in outdoor scenarios where LiDAR point clouds are large-scale and complexly distributed. Some methods[9–11] optimize photometric and depth consistency, limiting their applicability to indoor scenarios where RGB-D data and differentiable rendering are feasible. Others[13, 14] learn global alignment and neighborhood consensus, but struggle with low overlap and density variation in outdoor settings. Recent advances resort to pseudo-label-based frameworks, achieving promising results in outdoor scenarios[12, 15]. However, they rely solely on geometric cues to mine and filter pseudo-labels, neglecting the complementarity of high-level contextual information in feature space. Their partiality results in incomplete scene perception, leading to noisy and suboptimal optimization objectives.

Various 2D [16, 17] and 3D vision tasks [18–20, 2, 5, 4] have benefited from integrating *both* low-level *and* high-level information. In point cloud registration, as illustrated in Fig. 1 (Left), we observe that potential inliers (outliers) tend to cluster around *positive (negative)* anchors that summarize the features of existing inliers (outliers) in the feature space, respectively. This suggests that high-level contextual information is adept at discovering inliers from a global perspective of the scene. Meanwhile, low-level geometric cues have proven effective in rejecting outliers[13, 21–23]. Inspired by this, we propose a novel method, termed **INTEGER**, which adopts a *teacher-student framework* to **m**INE and **T**ransfer **f**Eature-**G**Eometry cohe**R**ence for unsupervised point cloud registration.

Specifically, our method starts by initializing a teacher with synthetic pairs generated from each point cloud scan, and then transfers to real point cloud pairs with a teacher-student framework. Building upon our observations, we introduce the **Feature-Geometry Coherence Mining (FGCM)** module for the *teacher*, which first adapts the teacher to each mini-batch of real data to establish a denoised feature space. Reliable pseudo-labels, including correspondences and anchors, are then generated based on our key observation by iteratively mining potential inliers based on their similarity to anchors and rejecting outliers via spatial compatibility [21]. These robust pseudo-labels mined by FGCM not only accurately include inlier correspondences as shown in Fig. 1 (Right), but also aggregate effective representations of inliers and outliers from the teacher. We refer to this characteristic as *feature-geometry coherence*. To further enhance robustness and transfer feature-geometry coherence to the student, we propose **Anchor-Based Contrastive Learning (ABCont)** for contrastive learning with anchors. Meanwhile, we design a succinct and efficient **Mixed-Density Student (MDS)** for the *student* to learn density-invariant features using teacher’s anchors, overcoming density variation and low overlap in distant scenarios.

We extensively evaluate our method on two large-scale outdoor datasets, KITTI and nuScenes. By exploiting feature-geometry coherence for reliable optimization objectives, INTEGER outperforms existing unsupervised methods by a considerable margin. It even performs competitively compared to state-of-the-art supervised methods, especially in distant scenarios. To the best of our knowledge, our approach is the first to integrate both low-level and high-level information for producing pseudo-labels of unsupervised point cloud registration. Overall, our contributions are threefold:

- We propose INTEGER, a novel method to exploit low-level and high-level information for unsupervised point cloud registration, achieving superior performance in complex outdoor scenarios.
- We introduce FGCM and MDS for the teacher and student, respectively, to mine reliable pseudo-labels and learn density-invariant features.
- We design ABCont to mitigate pseudo-label noise and facilitate contrastive learning with anchors for a robust feature space.

2 Related Works

Supervised Registration. There are two categories of supervised registration approaches: Correspondence-based methods [24–26, 4, 27, 28, 3] first extract point correspondences and then estimate the transformation with robust pose estimators. In contrast, direct registration methods [29–31] extract global feature vectors and regress the transformation directly with a neural network. Recently, a series of works [12, 32] have tackled distant point cloud registration, which is crucial for real-world applications.

Unsupervised Registration. Previous researches in unsupervised registration mainly focus on indoor scenes. BYOC[11] suggests that random 2D CNNs generate robust image correspondences for supervising 3D registration networks. Meanwhile, render-based methods[9, 10] leverage differentiable renders as the supervision signal of 3D registration. However, these methods are restricted to RGB-D input. To address this, Mei et al. [33] enforce consistencies between Gaussian Mixture Models for unsupervised training, using only point cloud as input. Shen et al. [13] introduce an inlier evaluation method based on neighborhood consensus. However, its performance drops when the overlap is low. SGP[15] proposes a teacher-student framework for self-supervised learning from hand-crafted feature descriptors. EYOC[12] introduces progressive training and spatial filtering to adapt the model to distant point cloud pairs gradually, demonstrating promising results in outdoor scenarios.

Robust Pose Estimators. Pose estimators evaluate inliers and estimate poses from input correspondence sets. Traditional methods such as RANSAC[34] suffer from inefficiencies. Learning-based methods[35–37] learn to predict inliers and poses using neural networks. However, they require training and are thus constrained to supervised settings. To address this, non-parametric methods have emerged. Chen et al. [21] introduced SC^2 -measurements for robust inlier selection. Graph-based methods such as MAC[23] and FastMAC[22] approximate maximal cliques for fast and accurate inlier evaluation.

3 Methodology

Problem Formulation. Given two point clouds $\mathcal{P} = \{\mathbf{p}_i\} \in \mathbb{R}^{m \times 3}$ and $\mathcal{Q} = \{\mathbf{q}_j\} \in \mathbb{R}^{n \times 3}$, the goal of point cloud registration is to uncover the rigid transformation $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$ that perfectly aligns \mathcal{P} to \mathcal{Q} , where $\mathbf{R} \in \text{SO}(3)$ is the rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is the translation vector. When the two point clouds are acquired at a large distance d such as when $d \in [5\text{m}, 50\text{m}]$, the registration task faces the challenges of low overlap and density variation [32, 12, 38]. Therefore, it is crucial to learn density-invariant features.

Overall Pipeline. INTEGER adopts a two-stage training scheme and a teacher-student framework. Training of INTEGER consists of two stages: First, we initialize the teacher with synthetic data. Then, we train a student model on *real data* with the reliable pseudo-labels mined by the teacher. The overall pipeline and proposed modules are illustrated in Fig. 2. During teacher-student training, FGCM first dynamically adapts the teacher model θ to a *data-specific teacher* ϕ designated for the current mini-batch, and then mines reliable pseudo-labels with the adapted teacher. Next, the MDS learns density-invariant features by learning to match regular and sparse views of point cloud pairs supervised by pseudo-labels mined by ϕ . A pseudo-label $\mathcal{I} = \{\mathcal{C}, \hat{\mathcal{C}}, \mathcal{A}_+, \mathcal{A}_-\}$ contains correspondences $\mathcal{C}, \hat{\mathcal{C}}$ to supervise dense matches and sparse matches, respectively. The feature-space positive and negative anchors, denoted respectively by \mathcal{A}_+ and \mathcal{A}_- , serve as overall representatives of inliers and outliers in the feature space. For a correspondence $^{(i,j)}\mathcal{C} = (\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{C}$, the correspondence features are

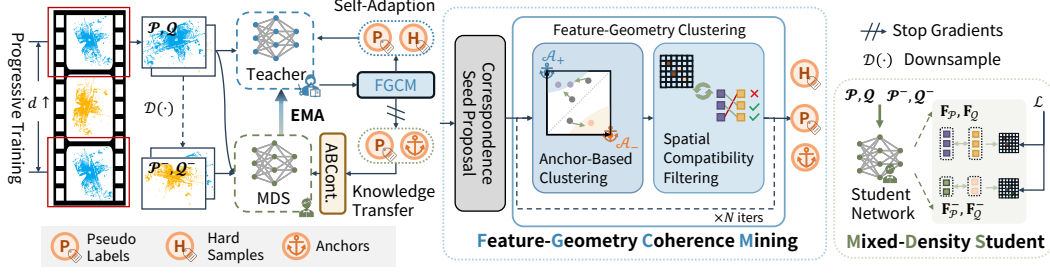


Figure 2: **The Overall Pipeline.** FGCM(Sec. 3.2) first adapt the teacher model to a *data-specific teacher* for the current mini-batch, and then mine reliable pseudo-labels. Next, MDS(Sec. 3.4) learns density-invariant features from pseudo-labels. ABCont(Sec. 3.3) is applied for adapting the teacher and transferring knowledge to the student in the feature space.

defined as $\mathbf{F}_C^{(i,j)} = \mathbf{F}_i^P - \mathbf{F}_j^Q$. Then, the positive and negative anchors $\mathcal{A}_+, \mathcal{A}_-$ are computed as the average of the respective features of inliers \mathcal{C}_+ and outliers \mathcal{C}_-

$$\mathcal{A}_+ = \frac{1}{|\mathcal{C}_+|} \sum_{(\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{C}_+} \mathbf{F}_C^{(i,j)}, \quad \mathcal{A}_- = \frac{1}{|\mathcal{C}_-|} \sum_{(\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{C}_-} \mathbf{F}_C^{(i,j)}, \quad (1)$$

ABCont is applied to effectively learn a robust student guided by anchors from the teacher. Progressive training [12] is adopted to gradually train the student to adapt to pairs of distant point clouds.

3.1 Synthetic Teacher Initialization

To initialize a teacher model, Liu et al. [12] assume that two consecutive frames approximately have no relative transformation and pretrain the teacher with the *identity* transformation. However, the errors introduced in such approximation lead to suboptimal initial teachers. To address this, inspired by existing efforts[39, 40], we instead pretrain the teacher with synthetic pairs generated from each real scan. Specifically, we follow PointContrast[40] to generate two partially overlap fragments for each scan. We additionally apply periodic sampling[39] to remove points periodically with respect to a random center, simulating the irregular sampling of LiDAR. Please refer to the Appendix for a visualization of synthetic pairs.

3.2 Feature-Geometry Coherence Mining

With the teacher initialized on synthetic pairs, our goal is to provide reliable supervision for the student. Despite efforts to ensure an effective initialization, a distribution discrepancy persists between synthetic and real data. Hence, we introduce a *train-only* FGCM. As is depicted in Fig. 2, FGCM starts with Correspondence Seed Proposals for \mathcal{C}^0 using a simple similarity threshold. Subsequently, Feature-Geometry Clustering extends from \mathcal{C}^0 by mining additional reliable correspondences and anchors, which serve as effective optimization objectives.

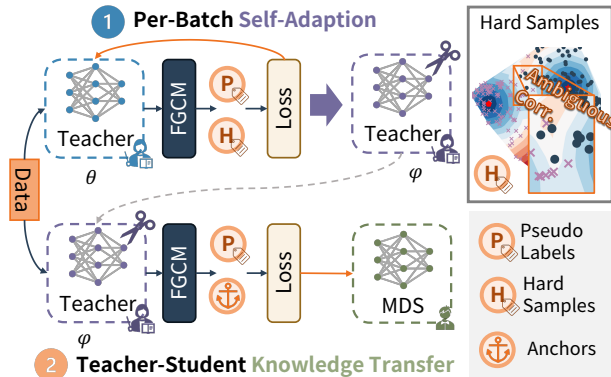


Figure 3: The two-pass usage of the proposed FGCM.

As is illustrated in Fig. 3, for each mini-batch, we use FGCM in a two-pass manner. In the first forward pass, we perform Per-Batch Self-Adaption on the teacher model θ to establish a denoised feature space, yielding a data-specific teacher ϕ . In the second forward pass, the adapted teacher ϕ and FGCM are used to mine reliable pseudo-labels \mathcal{I} , which are then used to train the student, achieving Teacher-Student Knowledge Transfer.

Algorithm 1: Feature-Geometry Clustering

Input: Initial correspondence seed proposals \mathcal{C}^0
Compute initial $\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}}$ and anchors $\mathcal{A}_+, \mathcal{A}_-$ with Eq. 1
for $i = 1$ **to** max_iters **do**
 Generate unclassified correspondences $\mathcal{C}_U \leftarrow \text{FeatureMatching}(\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}})$
 Select $\mathcal{C}_U^{\text{top-}k}$ with top- k S_c^+ satisfying $S_c^+ > S_c^-$ based on Eq. 2
 Update $\mathcal{C}^i \leftarrow \mathcal{C}^{i-1} \cup \mathcal{C}_U^{\text{top-}k}$ // Anchor-Based Clustering
 Filter \mathcal{C}^i with spatial compatibility to produce $\mathcal{C}_+^i, \mathcal{C}_-^i$ // Spatial Compatibility Filtering
 Update $\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}}$ and $\mathcal{A}_+, \mathcal{A}_-$ with Eq. 1
 if $|\mathcal{C}_+^i| = |\mathcal{C}_+^{i-1}|$ **or** $|\mathcal{C}_-^i| = |\mathcal{C}_-^{i-1}|$ **then**
 $\mathcal{C} \leftarrow \mathcal{C}_+^i$
 break
return $\mathcal{C}, \mathcal{A}_+, \mathcal{A}_-$

Feature-Geometry Clustering. *Feature-Geometry Clustering* is central to the FGCM, designed to extend initial correspondence proposals by integrating both high-level *feature* representations and low-level *geometric* cues. It iteratively includes speculative inliers based on feature-space clustering, followed by outlier rejection with spatial compatibility filtering. We empirically adopt SC²-PCR[21] for spatial compatibility filtering. Our experiments show that our method is agnostic to the choice of spatial compatibility measures.

To discover latent correspondences in the feature space, it is necessary to measure the similarity between putative correspondences and anchors. Inspired by Xia et al. [41], we compute the feature similarity using both Euclidean distance and cosine similarity. Specifically, for a correspondence $(\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{C}$ with its features $\mathbf{F}_c^{(i,j)}$, The similarity S_c^+ and S_c^- w.r.t. respective anchors \mathcal{A}_+ and \mathcal{A}_- is computed as:

$$S_c^+ = \min\{\text{D}_E(\mathcal{A}_+, \mathbf{F}_c^{(i,j)}), \text{D}_C(\mathcal{A}_+, \mathbf{F}_c^{(i,j)})\}, S_c^- = \min\{\text{D}_E(\mathcal{A}_-, \mathbf{F}_c^{(i,j)}), \text{D}_C(\mathcal{A}_-, \mathbf{F}_c^{(i,j)})\}, \quad (2)$$

where $\text{D}_E(\mathbf{F}_1, \mathbf{F}_2) = 1 - \min(L_2(\mathbf{F}_1, \mathbf{F}_2), 1)$ and $\text{D}_C(\mathbf{F}_1, \mathbf{F}_2) = (\cos(\mathbf{F}_1, \mathbf{F}_2) + 1) / 2$ are *normalized* Euclidean distance and cosine similarity, respectively.

The algorithm is detailed in Alg. 1. Given correspondence set \mathcal{C}^i at i -th iteration, we define unclassified points $\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}}$ in \mathcal{P} and \mathcal{Q} as:

$$\mathbf{U}^{\mathcal{P}} = \{\mathbf{p} | \mathbf{p} \in \mathcal{P} \wedge (\mathbf{p}, *) \notin \mathcal{C}^i\}, \mathbf{U}^{\mathcal{Q}} = \{\mathbf{q} | \mathbf{q} \in \mathcal{Q} \wedge (*, \mathbf{q}) \notin \mathcal{C}^i\} \quad (3)$$

Then, the algorithm takes an iterative approach, starting from the given initial correspondence set \mathcal{C}^0 : during the i^{th} iteration, it (1) generates putative correspondences from $\mathbf{U}^{\mathcal{P}}$ and $\mathbf{U}^{\mathcal{Q}}$ via feature matching; (2) expands \mathcal{C}^{i-1} with top- k similar correspondences to positive anchors measured by S_c , yielding \mathcal{C}^i ; (3) filters the expanded correspondence set with spatial compatibility and updates the anchors based on Eq. 1; (4) updates $\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}}$ and $\mathcal{A}_+, \mathcal{A}_-$ according to Eq. 3. The iteration stops when the number of inliers and outliers converges, or the maximum iteration is reached.

With ample accurate correspondences included in \mathcal{C} , we can then estimate a more accurate transformation \mathbf{T} and compute $\hat{\mathcal{C}}$ using nearest neighbor search (NN-search) with \mathbf{T} . We do not directly apply Alg. 1 for sparse pairs because, in downsampled views, the features become less descriptive [38], hindering feature-based approaches.

Per-Batch Self-Adaption. Throughout the iterations in Alg. 1, positive and negative anchors gradually aggregate representative and discriminative features of inliers and outliers, respectively. In the first forward pass, noise exists due to distributional discrepancies, leading to the rejection of some correspondences by spatial compatibility. These rejected correspondences are *hard samples*: ambiguous correspondences that are closer to positive anchors in the feature space but are more likely to be outliers. We leverage these hard samples for teacher self-adaptation by applying the InfoNCE[42] loss, guiding the teacher to distinguish them from the positive anchors. This step results in the adapted teacher ϕ , which produces more discriminative features for the current mini-batch.

Focusing only on hard samples for self-adaptation is more efficient than simply using all correspondences for self-adaptation. Hard samples capture the key ambiguities in the feature space while introducing only a limited number of pairwise relationships. This defines a clear and reliable optimization objective for self-adaptation. In contrast, self-adaptation with all correspondences not only slows down the training, but also introduces too many already-distinguishable pairwise relationships, diluting the focus on feature-space ambiguity and thus hindering effective adaptation.

Teacher-Student Knowledge Transfer. After Per-Batch Self-Adaption, the feature space of the adapted teacher ϕ is expected to contain less noise. Consequently, the positive and negative anchors become sufficiently representative and discriminative now, enabling effective guidance for the student to learn a robust feature space. For Teacher-Student Knowledge Transfer, we utilize both the correspondences and anchors from the adapted teacher ϕ to train the student using the proposed ABCont. Unlike existing methods[15, 12] that rely solely on correspondences, our approach directly bridges the teacher and student in the feature space via anchors, providing a clear and effective optimization objective for the student.

3.3 Anchor-Based Contrastive Learning

Contrastive learning has been widely adopted to train registration models[43, 24, 4, 3, 20]. Recently, a surge of research on various tasks involves *anchor-based* or *proxy-based* approaches to facilitate contrastive learning due to their robustness against inconsistency and noise in the feature space[44, 45, 41], superiority in generalizability[46] and ability to learn discriminative features[47]. Therefore, we design ABCont to leverage positive and negative anchors to facilitate effective contrastive learning with the pseudo-labels, where noise and outliers are inevitable. As shown in Fig. 4, with anchor-based representations, ABCont sets up a convergence target that is more robust against label noise. Moreover, it is more efficient because the number of additionally-introduced pairwise relationships is reduced[46].

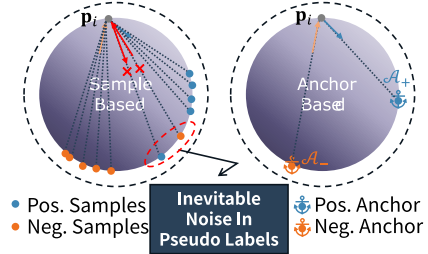


Figure 4: **Toy Example for ABCont.** Anchor-based methods introduce fewer pairwise relationships and are robust against inevitable label noise.

Specifically, we propose the ABCont loss $\mathcal{L}_{\text{ABCont}} = \mathcal{L}_{\text{reg}} + \lambda_{\text{corr}} \mathcal{L}_{\text{corr}}$, where $\mathcal{L}_{\text{corr}}$ is the anchor-based correspondence loss, weighted by a hyperparameter λ_{corr} to complement the registration loss \mathcal{L}_{reg} originally used by the feature extractors. The student’s feature matching results can be classified into inliers \mathcal{C}_+ and outliers \mathcal{C}_- based on the pseudo-labels from the teacher. Then, anchors $\{\mathcal{A}_+, \mathcal{A}_-\}$ from the teacher are designated as a *universal inlier* and a *universal outlier*, resulting in augmented inliers and outliers:

$$\mathcal{C}_+^* = \mathcal{C}_+ \cup \text{sg}(\mathcal{A}_+), \quad \mathcal{C}_-^* = \mathcal{C}_- \cup \text{sg}(\mathcal{A}_-), \quad (4)$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operator, preventing gradients from flowing back to the teacher. Following existing efforts[41, 40], we sample n_p correspondences randomly and formulate \mathcal{L}_{aux} as a contrastive learning problem to distinguish inliers from outliers. InfoNCE[42] loss is then applied to these correspondence features:

$$\mathcal{L}_{\text{corr}} = -\frac{1}{n_p} \sum_{i=1}^{n_p} \log \frac{\exp(\beta_p^i)}{\exp(\beta_p^i) + \sum_{j=1}^{n_n} \exp(\beta_n^j)}, \quad (5)$$

where β_p^i and β_n^j are the distance between the i^{th} positive correspondence and the j^{th} negative correspondence, respectively. ABCont is pivotal in transferring feature-geometry coherence from the teacher to the student: the accurate pseudo-labels for correspondences, combined with anchors, enable the student to learn discriminative features efficiently. Anchors from the teacher impose direct constraints on the student’s feature space, encouraging the student to replicate the teacher’s feature-space matchability. This leads to a more effective transfer of feature-geometry coherence.

3.4 Mixed-Density Student

The density of LiDAR point clouds varies greatly with the distance to the sensor, posing challenges for matching distance point clouds effectively[38]. To address this, it is crucial for a student model

to learn density-invariant features, ensuring robust correspondences across varying point densities. Previous methods [32, 38] have sought density invariance through auxiliary reconstruction tasks or by identifying positive groups, but these techniques are either computationally expensive or depend on precise supervision. Xia et al. [48] introduced a simple yet effective technique for density invariance in object detection by using features from downsampled views. Inspired by this, we propose Mixed-Density Student to learn density-invariant features from reliable correspondences.

Specifically, given point clouds \mathcal{P} and \mathcal{Q} , we compute their sparsely-downsampled views \mathcal{P}^- and \mathcal{Q}^- with increased voxel sizes. We then extract student’s features $(\mathbf{F}_{\mathcal{P}}, \mathbf{F}_{\mathcal{Q}})$ and $(\mathbf{F}_{\mathcal{P}^-}, \mathbf{F}_{\mathcal{Q}^-})$. Using features from point clouds of different density, we compute *dense matches* and *sparse matches* through matching respective features. We then apply ABCont to both sets of matches, encouraging the extraction of similar features at corresponding spatial locations across point clouds of varying densities, thereby promoting density-invariant feature learning.

Loss Aggregation. The student’s overall training loss is aggregated as a weighted combination of $\mathcal{L}_{\text{ABCont}}$ on both dense and sparse matches:

$$\mathcal{L} = \mathcal{L}_{\text{ABCont}}^{(\mathcal{P}, \mathcal{Q})} + \lambda_1 \mathcal{L}_{\text{ABCont}}^{(\mathcal{P}^-, \mathcal{Q}^-)}, \quad (6)$$

where λ_1 is a weight for the sparse match.

4 Experiments

We mainly evaluate INTEGER on two challenging public datasets: KITTI[6] and nuScenes[7]. Both datasets adhere to official splits. The evaluation protocol follows the standard setting of EYOC[12]. Please refer to the appendix for more details of implementation and experimental settings.

Metrics Following previous works[4, 2, 43, 38], we evaluate the registration performance using *Relative Rotation Error* (RRE), *Relative Translation Error* (RTE) and *Registration Recall* (RR). Related to the practical purpose of outdoor registration, we additionally report $\text{RR}@[d_1, d_2]$ and *mean Registration Recall*(mRR). $\text{RR}@[d_1, d_2]$ is registration recall w.r.t pairs with distance $d \in [d_1, d_2]$, following [12]. mRR is defined as the average of $\text{RR}@[d_1, d_2]$ for all $[d_1, d_2]$. To measure the quality of correspondences in pseudo-labels, we report *Inlier Ratio*(IR) of the *teacher* in the first epoch, denoted “tIR@1st Epoch”.

Baselines For supervised methods, We compare INTEGER with FCGF[24], Predator[43], SpinNet[49], D3Feat[50], CoFiNet[51], and Geometric Transformer(GeoTrans.)[4]. For unsupervised methods, we compare with RIENet[13] and EYOC[12]. Following Liu et al. [12], we report a variant of FCGF denoted as FCGF+C, which is FCGF trained with progressive training [12].

4.1 Performance Comparison with State-of-the-Art

Quantitative results are presented in Table 1. Our method outperforms existing unsupervised approaches and achieving state-of-the-art performance across all datasets and demonstrates superior generalizability. Notably, our unsupervised approach maintains competitive performance compared to supervised methods and even surpasses them in distant scenarios, highlighting its potential for real-world application.

Overall Performance Compared to existing methods, our approach excels in performance. RIENet, an end-to-end unsupervised registration method for outdoor scenes, exhibits suboptimal performance, particularly in low-overlap scenarios and environments with low LiDAR resolution, such as nuScenes. Both EYOC and INTEGER adopt a teacher-student framework for unsupervised training. However, our method demonstrates superior accuracy across all evaluation metrics, overcoming challenges associated with pseudo-label discovery and the absence of feature-space knowledge transfer in EYOC.

Generalizability We assess generalizability on nuScenes using weights trained on KITTI. Variations in LiDAR resolutions between nuScenes and KITTI may lead to different point densities, potentially degrading extracted features. Compared to existing unsupervised methods, our approach exhibits superior generalizability to unseen datasets. This superiority can be attributed to INTEGER’s design, which learns density-invariant features.

Table 1: **Comparisons with State-of-the-Art Methods.** “✓” in the column “U” denotes the methods are *Unsupervised*. Otherwise, they are supervised. The best *unsupervised* results are highlighted in **bold**. “KITTI→nuScenes” denotes generalizability results from KITTI to nuScenes.

Dataset	Method	U	mRR	RR@d ∈				
				[5, 10)	[10, 20)	[20, 30)	[30, 40)	[40, 50)
KITTI	FCGF	–	77.4	98.4	95.3	86.8	69.7	36.9
	FCGF+C	–	84.6	100.0	97.5	90.1	79.1	56.3
	Predator	–	87.9	100.0	97.5	90.1	79.1	56.3
	SpinNet	–	39.1	99.1	82.5	13.7	0.0	0.0
	D3Feat	–	66.4	99.8	98.2	90.7	38.6	4.5
	CoFiNet	–	82.1	99.9	99.1	94.1	78.6	38.7
	GeoTrans.	–	42.2	100.0	93.9	16.6	0.7	0.0
	EYOC	✓	83.2	99.5	96.6	89.1	78.6	52.3
	RIENet	✓	50.7	96.3	72.1	38.2	24.4	22.6
	Ours	✓	84.0	99.5	97.1	89.6	79.6	54.2
nuScenes	FCGF	–	39.5	87.9	63.9	23.6	11.8	10.2
	FCGF+C	–	59.3	96.2	85.1	59.6	35.8	20.0
	Predator	–	51.0	99.7	72.2	52.8	16.2	14.3
	EYOC	✓	61.7	96.7	85.6	61.8	37.5	26.9
	RIENet	✓	47.1	96.5	57.9	36.6	25.8	18.9
	Ours	✓	63.1	97.1	86.9	62.9	39.6	29.4
KITTI ↓ nuScenes	EYOC	✓	55.3	96.2	75.6	58.7	26.6	19.7
	RIENet	✓	46.2	83.3	73.2	43.5	19.8	11.1
	Ours	✓	62.6	97.5	84.6	62.6	37.8	30.2

4.2 Analysis

Different Choices of Pose Estimator in FGCM We contend that the robustness and efficacy of FGCM are not contingent upon a specific pose estimator. To substantiate this claim, we conduct experiments employing various robust pose estimators within FGCM. The results are detailed in Table 2. For different robust pose estimators in FGCM module, we experiment PointDSC[35]², MAC[23], FastMAC[22] and SC²-PCR[21]. The results demonstrate that the effectiveness of FGCM is agnostic to choices of pose estimators, despite marginal performance discrepancies are observed. Given the iterative nature of FGCM, the efficiency of pose estimators holds paramount importance, as the module’s runtime is proportional to pose estimation time. We choose SC²-PCR[21] for FGCM by default due to its superior balance in performance and efficiency.

Table 2: Different Pose Estimators in FGCM

Pose Estimators	tIR@1 st Epoch	Time (s)
PointDSC	81.3	1.13
MAC	80.1	28.2
FastMAC	79.3	0.67
SC ² -PCR	<u>81.2</u>	<u>0.75</u>

Effectiveness of Self-Adaption for Discriminative Features. To further understand the effectiveness of self-adaption in FGCM, we visualize the *point-level feature distribution* and *correspondence-level similarity distribution* in Fig. 5 (Please refer to the Appendix for implementation details.). The two representative samples are taken from KITTI dataset. In Fig. 5, the smaller overlap regions of point-level feature distribution between points from inliers and outliers indicate the features of inliers and outliers distribute more distant, and thus, the features are more discriminative. For correspondence-level similarity, inlier similarity should be distinct from outlier similarity to effectively differentiate between the two. With the self-adaption in FGCM, the data-specific teacher produces more discriminative features, resulting in a less noisy feature space conducive to the subsequent feature-based approach employed in FGCM.

²We directly use their official weights for evaluation.

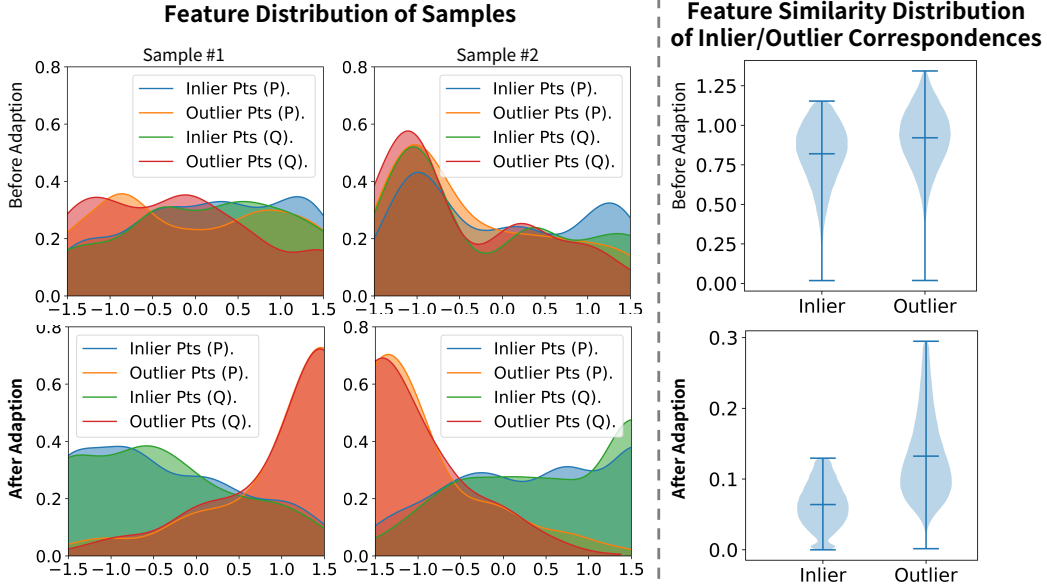


Figure 5: **Before v.s. After Self-Adaption in FGCM:** Point-wise Feature & Correspondence-wise Similarity Distribution indicate that the self-adaption results in more discriminative features.

4.3 Ablation Study

We conduct ablation studies to evaluate the efficacy of INTEGER on the KITTI dataset. We present mRR and registration errors in a distant scenario where $d \in [40, 50)$. Various alternative configurations of INTEGER are compared in Table 3. Our method exhibits superior performance compared to alternatives, demonstrating the effectiveness of our design. This superiority may be attributed to the features-geometry coherence: With FGCM, correspondences in pseudo-labels possess discriminative features, facilitating effective knowledge transfer in feature space using ABCont. Additionally, MDS significantly enhances performance in distant scenarios. The combination of Per-Batch Self-Adaption and Feature-Geometry Clustering in the FGCM module yields more substantial improvement than using either alone. The removal of Per-Batch Self-Adaption marginally degrades the quality of pseudo-labels, emphasizing the importance of denoising the feature space. When synthetic teacher initialization is removed (w/o S.T.I), we employed the same way as EYOC to pretrain the teacher. We find that synthetic teacher initialization greatly enhances the initial teacher’s performance. Please refer to the Appendix for more qualitative results on generated synthetic pairs.

Table 3: **Ablation Study of INTEGER.** S.T.I denotes synthetic teacher initialization. PBSA and FGC denote Per-Batch Self-Adaption and Feature-Geometry Clustering, respectively

Methods	tIR@1 st Epoch	mRR	$d \in [40, 50)$		
			RR	RRE	RTE
Full	81.2	84.0	54.2	1.1	0.54
w/o ABCont	80.3	83.5	53.7	1.3	0.58
w/o PBSA	43.3	80.9	50.2	1.7	0.79
w/o FGC	67.6	82.8	52.7	1.4	0.61
w/o MDS	81.2	82.7	52.3	1.3	0.71
w/o S.T.I	71.9	83.7	53.7	1.2	0.55

5 Conclusion

In this paper, we present INTEGER, a novel unsupervised method for point cloud registration that integrates low-level geometric and high-level contextual information for reliable pseudo-labels. Our method introduces Feature-Geometry Coherence Mining for dynamic teacher self-adaption and robust pseudo-label mining based on both feature and geometric spaces. Then, we propose Mixed-Density Student to learn density-invariant features. We also introduce Anchor-Based Contrastive Learning

for effective contrastive learning using anchors. Extensive experiments on two large-scale outdoor datasets validate our method’s efficacy. Despite being unsupervised, it achieves results comparable to state-of-the-art supervised methods and surpasses existing unsupervised methods, particularly in distant scenarios. Furthermore, our approach exhibits superior generalizability to unseen datasets.

Limitations. The main limitations of the proposed method are twofold:

- Our method is subject to the quality of the teacher. If the teacher is inaccurate, the feature space may become too noisy, potentially impeding Feature-Geometry Clustering in FGCM, especially in distant scenarios. One potential remedy is to devise a more robust strategy for initializing the teacher.
- Our method is slightly slower to obtain pseudo-labels compared to existing efforts[12] due to the proposed iterative method used in FGCM of the FGCM module. Future work may involve devising a more efficient strategy for mining pseudo-labels.

6 Acknowledgements

This work was supported in part by the National Key R&D Program of China under Grant 2021YFF0704600, the Fundamental Research Funds for the Central Universities (No. 20720220064).

References

- [1] Anh-Quan Cao, Gilles Puy, Alexandre Boulch, and Renaud Marlet. Pcam: Product of cross-attention matrices for rigid registration of point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13229–13238, 2021.
- [2] Fan Lu, Guang Chen, Yinlong Liu, Lijun Zhang, Sanqing Qu, Shu Liu, and Rongqi Gu. Hregnet: A hierarchical network for large-scale outdoor lidar point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16014–16023, 2021.
- [3] Kezheng Xiong, Maoji Zheng, Qingshan Xu, Chenglu Wen, Siqi Shen, and Cheng Wang. Speal: Skeletal prior embedded attention learning for cross-source point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6279–6287, 2024.
- [4] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11143–11152, 2022.
- [5] Jiuming Liu, Guangming Wang, Zhe Liu, Chaokang Jiang, Marc Pollefeys, and Hesheng Wang. Regformer: an efficient projection-aware transformer network for large-scale point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8451–8460, 2023.
- [6] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012.
- [7] 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 multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.
- [8] Pei Sun, Henrik Kretschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.
- [9] Mohamed El Banani, Luya Gao, and Justin Johnson. Unsuperviseddr&r: Unsupervised point cloud registration via differentiable rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7129–7139, 2021.

- [10] Mingzhi Yuan, Kexue Fu, Zhihao Li, Yucong Meng, and Manning Wang. Pointmbf: A multi-scale bidirectional fusion network for unsupervised rgb-d point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17694–17705, 2023.
- [11] Mohamed El Banani and Justin Johnson. Bootstrap your own correspondences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6433–6442, 2021.
- [12] Quan Liu, Hongzi Zhu, Zhenxi Wang, Yunsong Zhou, Shan Chang, and Minyi Guo. Extend your own correspondences: Unsupervised distant point cloud registration by progressive distance extension. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11366–11374, 2024.
- [13] Yaqi Shen, Le Hui, Haobo Jiang, Jin Xie, and Jian Yang. Reliable inlier evaluation for unsupervised point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2198–2206, 2022.
- [14] Pengcheng Shi, Jie Zhang, Haozhe Cheng, Junyang Wang, Yiyang Zhou, Chenlin Zhao, and Jihua Zhu. Overlap bias matching is necessary for point cloud registration, 2023. URL <https://arxiv.org/abs/2308.09364>.
- [15] Heng Yang, Wei Dong, Luca Carlone, and Vladlen Koltun. Self-supervised geometric perception. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2021.
- [16] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.
- [17] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.
- [18] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019.
- [19] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10529–10538, 2020.
- [20] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Learning semantic segmentation of large-scale point clouds with random sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8338–8354, 2021.
- [21] Zhi Chen, Kun Sun, Fan Yang, and Wenbing Tao. Sc2-pcr: A second order spatial compatibility for efficient and robust point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13221–13231, 2022.
- [22] Yifei Zhang, Hao Zhao, Hongyang Li, and Siheng Chen. Fastmac: Stochastic spectral sampling of correspondence graph. *arXiv preprint arXiv:2403.08770*, 2024.
- [23] Xiyu Zhang, Jiaqi Yang, Shikun Zhang, and Yanning Zhang. 3d registration with maximal cliques. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17745–17754, 2023.
- [24] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8958–8966, 2019.
- [25] Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European conference on computer vision (ECCV)*, pages 602–618, 2018.

- [26] Zan Gojcic, Caifa Zhou, Jan D Wegner, and Andreas Wieser. The perfect match: 3d point cloud matching with smoothed densities. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5545–5554, 2019.
- [27] Zi Jian Yew and Gim Hee Lee. Regtr: End-to-end point cloud correspondences with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6677–6686, 2022.
- [28] Junle Yu, Luwei Ren, Yu Zhang, Wenhui Zhou, Lili Lin, and Guojun Dai. Peal: Prior-embedded explicit attention learning for low-overlap point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17702–17711, 2023.
- [29] Hao Xu, Shuaicheng Liu, Guangfu Wang, Guanghui Liu, and Bing Zeng. Omnet: Learning overlapping mask for partial-to-partial point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3132–3141, 2021.
- [30] Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7163–7172, 2019.
- [31] Xiaoshui Huang, Guofeng Mei, and Jian Zhang. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11366–11374, 2020.
- [32] Quan Liu, Yunsong Zhou, Hongzi Zhu, Shan Chang, and Minyi Guo. Apr: online distant point cloud registration through aggregated point cloud reconstruction. *arXiv preprint arXiv:2305.02893*, 2023.
- [33] Guofeng Mei, Hao Tang, Xiaoshui Huang, Weijie Wang, Juan Liu, Jian Zhang, Luc Van Gool, and Qiang Wu. Unsupervised deep probabilistic approach for partial point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13611–13620, 2023.
- [34] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [35] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15859–15869, 2021.
- [36] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2514–2523, 2020.
- [37] Junha Lee, Seungwook Kim, Minsu Cho, and Jaesik Park. Deep hough voting for robust global registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15994–16003, 2021.
- [38] Quan Liu, Hongzi Zhu, Yunsong Zhou, Hongyang Li, Shan Chang, and Minyi Guo. Density-invariant features for distant point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18215–18225, 2023.
- [39] Sofiane Horache, Jean-Emmanuel Deschaud, and François Goulette. 3d point cloud registration with multi-scale architecture and unsupervised transfer learning. In *2021 international conference on 3D vision (3DV)*, pages 1351–1361. IEEE, 2021.
- [40] Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 574–591. Springer, 2020.

- [41] Qiming Xia, Jinhao Deng, Chenglu Wen, Hai Wu, Shaoshuai Shi, Xin Li, and Cheng Wang. Coin: Contrastive instance feature mining for outdoor 3d object detection with very limited annotations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6254–6263, 2023.
- [42] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
- [43] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4267–4276, 2021.
- [44] Yan Bai, Jile Jiao, Yihang Lou, Shengsen Wu, Jun Liu, Xuetao Feng, and Ling-Yu Duan. Dual-tuning: Joint prototype transfer and structure regularization for compatible feature learning. *IEEE Transactions on Multimedia*, 2022.
- [45] Oriol Barbany, Xiaofan Lin, Muhammet Bastan, and Arnab Dhua. Procsim: Proxy-based confidence for robust similarity learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1308–1317, 2024.
- [46] Xufeng Yao, Yang Bai, Xinyun Zhang, Yuechen Zhang, Qi Sun, Ran Chen, Ruiyu Li, and Bei Yu. Pcl: Proxy-based contrastive learning for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7097–7107, 2022.
- [47] Zhuangzhuang Chen, Jin Zhang, Zhuonan Lai, Jie Chen, Zun Liu, and Jianqiang Li. Geometry-aware guided loss for deep crack recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4703–4712, 2022.
- [48] Qiming Xia, Jinhao Deng, Chenglu Wen, Hai Wu, Shaoshuai Shi, Xin Li, and Cheng Wang. Hinted: Hard instance enhanced detector with mixed-density feature fusion for sparsely-supervised 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6254–6263, 2024.
- [49] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11753–11762, 2021.
- [50] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6359–6367, 2020.
- [51] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884, 2021.