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]

[a]Fujian Key Laboratory of Sensing and Computing for Smart Cities,
Xiamen University, China.
[b]Key Laboratory of Multimedia Trusted Perception and Efficient Computing,
Ministry of Education of China, Xiamen University, China.
[c]Nanyang Technological University, Singapore.
[d]University 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

38th Conference on Neural Information Processing Systems (NeurIPS 2024).
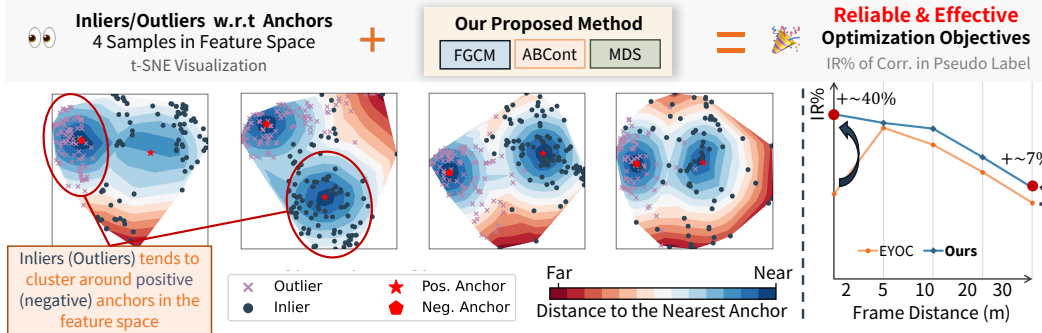
Figure 1: (1) **Motivation**: new inliers (outliers) tend to cluster around latent *positive (negative) anchors* that represent existing inliers (outliers) in the feature space, respectively. (2) **Performance**: pseudo-labels from INTEGER are more robust and accurate than the previous state-of-the-art 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**IN**e and **T**ransfer f**E**ature-**GE**ometry 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 **F**eature-**G**eometry **C**oherence **M**ining (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 **A**nchor-**B**ased **Cont**rastive Learning (ABCont) for contrastive learning with anchors. Meanwhile, we design a succinct and efficient **M**ixed-**D**ensity **S**tudent (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:

2

- 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 $\mathbf{R} \in \mathrm{SO}(3)$ 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 $d \in [5\mathrm{m}, 50\mathrm{m}]$, 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

3
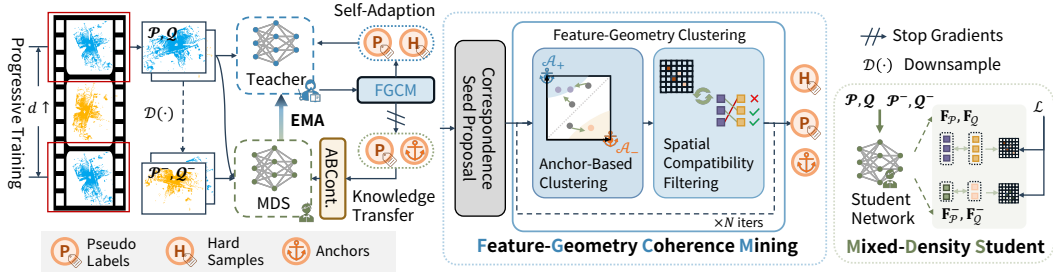
Figure 2: **The Overall Pipeline.** FGCM(Sec. 3.2) first adapt the teacher model to a *data-specific teacher* for the current mini-batch, and then mine reliable pseudo-labels. Next, MDS(Sec. 3.4) learns density-invariant features from pseudo-labels. ABCont(Sec. 3.3) is applied for adapting the teacher and transferring knowledge to the student in the feature space.

defined as $\mathbf{F}_{\mathcal{C}}^{(i,j)} = \mathbf{F}_i^{\mathcal{P}} - \mathbf{F}_j^{\mathcal{Q}}$. 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 $\mathcal{C}_-$

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

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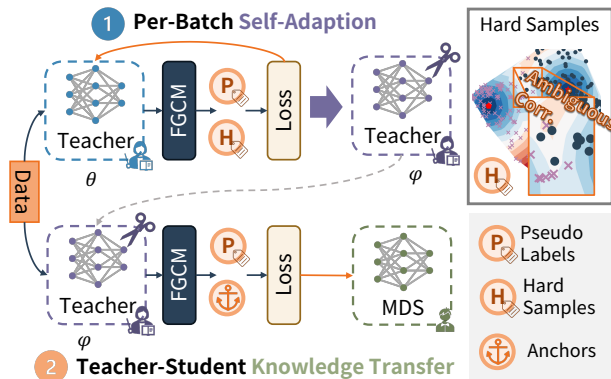
Figure 3: The two-pass usage of the proposed FGCM.

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.

4

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 $\mathcal{C}_U \leftarrow \text{FeatureMatching}\left(\mathbf{U}^{\mathcal{P}}, \mathbf{U}^{\mathcal{Q}}\right)$

    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 $\left|\mathcal{C}_+^i\right| = \left|\mathcal{C}_+^{i-1}\right|$ *or* $\left|\mathcal{C}_-^i\right| = \left|\mathcal{C}_-^{i-1}\right|$ **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 SC^2-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 $\mathbf{F}_{\mathcal{C}}^{(i,j)}$, The similarity S_c^+ and S_c^- w.r.t. respective anchors $\mathcal{A}_+$ and $\mathcal{A}_-$ is computed as:

$$S_c^+ = \min\{\text{D}_\text{E}(\mathcal{A}_+, \mathbf{F}_{\mathcal{C}}^{(i,j)}), \text{D}_\text{C}(\mathcal{A}_+, \mathbf{F}_{\mathcal{C}}^{(i,j)})\}, \ S_c^- = \min\{\text{D}_\text{E}(\mathcal{A}_-, \mathbf{F}_{\mathcal{C}}^{(i,j)}), \text{D}_\text{C}(\mathcal{A}_-, \mathbf{F}_{\mathcal{C}}^{(i,j)})\}, \ (2)$$

where $\text{D}_\text{E}(\mathbf{F}_1, \mathbf{F}_2) = 1 - \min(L_2(\mathbf{F}_1, \mathbf{F}_2), 1)$ and $\text{D}_\text{C}(\mathbf{F}_1, \mathbf{F}_2) = (\cos(\mathbf{F}_1, \mathbf{F}_2) + 1)/2$ 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:

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

Then, the algorithm takes an iterative approach, starting from the given initial correspondence set $\mathcal{C}^0$: during the ith 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 $\mathcal{S}_c$, 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.

5

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-adaption 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

- Pos. Samples
- Neg. Samples

Inevitable Noise In Pseudo Labels

⚓ Pos. Anchor
⚓ Neg. Anchor

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

the number of additionally-introduced pairwise relationships is reduced[46].

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 $\mathcal{L}_{\text{reg}}$ 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:

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

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 $\mathcal{L}_{\text{aux}}$ as a contrastive learning problem to distinguish inliers from outliers. InfoNCE[42] loss is then applied to these correspondence features:

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

where β_p^i and β_n^j are the distance between the ith positive correspondence and the jth 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

6

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:

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

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.

7

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@$d \in$				
				$[5, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$
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	98.6	97.1	80.6	63.1
	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 SC^2-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

Table 2: Different Pose Estimators in FGCM

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

estimation time. We choose SC^2-PCR[21] for FGCM by default due to its superior balance in performance and efficiency.

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.

---

[2]We directly use their official weights for evaluation.

8
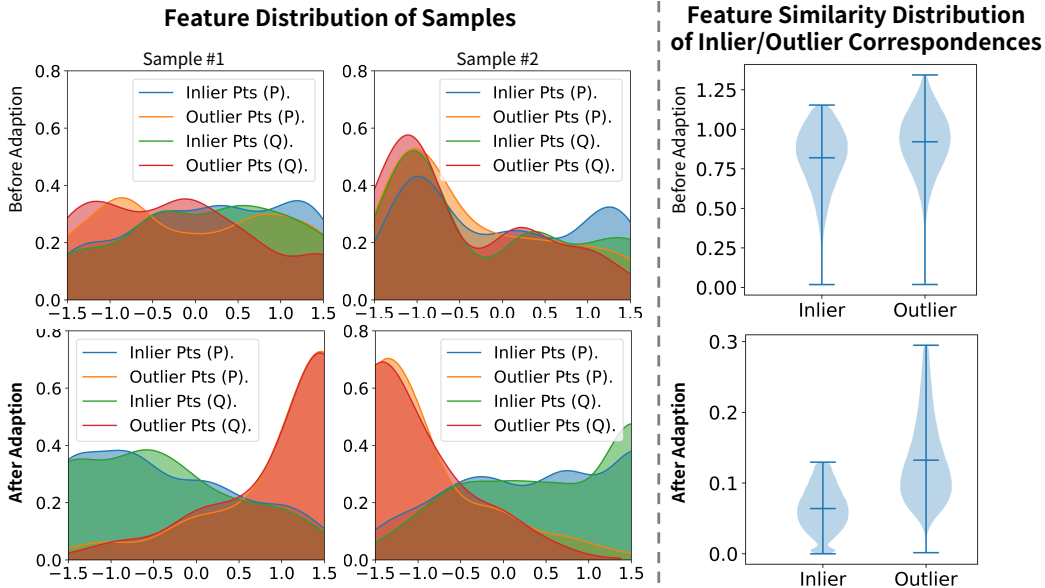
Figure 5: **Before v.s. After Self-Adaption in FGCM**: Point-wise Feature & Correspondence-wise Similarity Distribution indicate that the self-adaption results in more discriminative features.

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 **s**ynthetic **t**eacher **i**nitialization. PBSA and FGC denote Per-Batch Self-Adaption and Feature-Geometry Clustering, respectively

Methods	tIR@1st 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

9

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 Kretzschmar, 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. Unsupervisedr&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

[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.

11

[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.

12

[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.

[52] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[53] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

13

A Appendix

A.1 Evaluation Protocol

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".

Relative Rotation Error Relative rotation error is the geodesic distance in degrees between ground-truth and predicted rotation matrices:

$$\text{RRE}(\mathbf{R}_p, \mathbf{R}_{gt}) = \arccos\left(\frac{\text{trace}(\mathbf{R}_p^T \mathbf{R}_{gt}) - 1}{2}\right), \tag{7}$$

where R_{gt} and R_p denote the ground-truth and predicted rotation matrices, respectively.

Relative Translation Error Relative translation error is the Euclidean distance between ground-truth and predicted translation vectors:

$$\text{RTE}(\mathbf{t}_p, \mathbf{t}_{gt}) = \|\mathbf{t}_p - \mathbf{t}_{gt}\|_2, \tag{8}$$

where $\mathbf{t}_p$ and $\mathbf{t}_{gt}$ are the ground-truth and predicted translation vectors, respectively.

Registration Recall The registration recall is defined as the fraction of point cloud pairs whose RRE and RTE are simultaneously below the given thresholds (i.e.RRE $< 5°$ and RTE < 2m for KITTI datasets):

$$\text{RR} = \frac{1}{M}\sum_{i=1}^{M}[\![\text{RRE}_i < \tau_r \ \wedge \ \text{RTE}_i < \tau_t]\!], \tag{9}$$

where τ_r and τ_t are thresholds for RRE and RTE, respectively, and $[\![\cdot]\!]$ is the Iverson bracket. We compute the mean RRE and RTE only for point cloud pairs that are registered successfully, following common practices [50, 24, 43, 2, 51].

Following existing efforts[12], we denote the registration recall w.r.t pairs with distance $d \in [d_1, d_2]$ as $\text{RR@}[d_1, d_2]$. With the set of distance intervals $\mathcal{D} = \{[5, 10), [10, 20), [20, 30), [30, 40), [40, 50)\}$, we define the mean registration recall as:

$$\text{mRR} = \frac{1}{|\mathcal{D}|}\sum_{[d_1, d_2] \in \mathcal{D}} \text{RR@}[d_1, d_2]. \tag{10}$$

A.2 Implementation Details

We introduce the implementation details of INTEGER in this section. Firstly, we describe the network architectures of FCGF and SC^2-PCR, which we use by default to implement INTEGER. Then, we provide additional information about the implementation of FGCM. Finally, we provide the details of the implementation of visualization in Fig. 1 and Fig. 5.

FCGF We adopt the popular FCGF[24] for the registration network in INTEGER. As depicted in Fig. 6, the architecture incorporates Res-UNet structure and is implemented with sparse voxel convolution. It adopts three layers of skip connections with a roughly symmetric encoder-decoder structure. Before used for registration, features are normalized onto the unit sphere after the last convolutional layer.

14
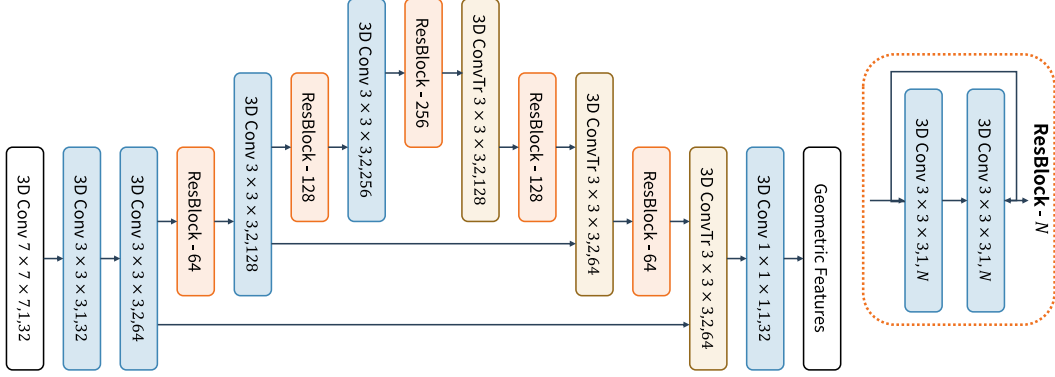
Figure 6: **The Architecture of FCGF.** We use the same architecture as their official implementation[24]. It adopts three layers of skip connections with a roughly symmetric encoder-decoder structure.

SC²-PCR   SC²-PCR is a robust pose estimator built upon the SC^2 measurement[21]. For two correspondence $\mathbf{c}_x = (\mathbf{p}_i, \mathbf{q}_j)$ and $\mathbf{c}_y = (\mathbf{p}_k, \mathbf{q}_l)$, Literatures prior to Chen et al. [21] utilized *first*-order spatial compatibility (SC) defined as:

$$\mathbf{M}_{xy} = |\|\mathbf{p}_i - \mathbf{p}_k\|_2 - \|\mathbf{q}_j - \mathbf{q}_l\|_2| \tag{11}$$

to measure the spatial compatibility between two correspondences, where $\mathbf{c}_x, \mathbf{c}_y \in \mathcal{C}$ and $\mathbf{M} \in \mathbb{R}^{|\mathcal{C}| \times |\mathcal{C}|}$ is the spatial compatibility matrix. To better distinguish inliers from outliers, Chen et al. [21] proposed a *second*-order spatial compatibility (SC^2) by using $\mathbf{M} \cdot \mathbf{M}^2$. Thus, the SC^2 scores for inliers will be skyrocketing and can easily be distinguished from outliers.

Built upon the SC^2 measurement, SC²-PCR first using a spectral technique to extract most promising seed correspondences and then iteratively refine the correspondences. The SC²-PCR is GPU-compatible and non-parametric. Therefore, it is efficient and has great generalizability for implementing our unsupervised method.

Details in FGCM   Following existing efforts[24], we have to sample $n_c = 1024$ correspondences for training with Hardest-Contrastive Loss. Thus, when FGCM mines more than n_c correspondences in pseudo-labels, we randomly sample n_c correspondences for training. However, in some cases, FGCM may fails to mine enough correspondences, resulting in less than n_c correspondences. For effective training, we additionally perform a NN-search to complement the correspondences under such circumstance. This typically happens in very early stages in training when the teacher model has not adequately adapted to real data yet, and very late stages when the overlap between two point clouds is too low.

Visualization in Fig. 1 and Fig. 5   In Fig. 1, we visualize the features of points and anchors by projecting them into 2D space using t-SNE[52]. In Fig. 5, we visualize the point-wise feature distribution and correspondence-wise similarity distribution. For point-level feature distribution, we first project the point features into *scalar* space using t-SNE[52]. Then, we estimate the probability density for points associated with inlier and outlier correspondences using Kernel Density Estimation (KDE)[53]. To estimate correspondence-level similarity distribution, we compute the L_2-distance in feature space between two points in each correspondence and visualize the distribution of similarities for inliers and outliers. We use the sklearn[54] library to implement t-SNE and KDE.

A.3   Training Details

To train INTEGER, we use the SGD optimizer with an initial learning rate of 0.3 and a weight decay of $1e-4$. We train INTEGER for 400 epochs with a batch size of 8. The training process takes approximately 6 days on a single NVIDIA RTX 3090 GPU running at 1.70 GHz with 24 GiB of GPU memory.

A.4   Additional Experiments

15

(a) k (b) Max Iterations
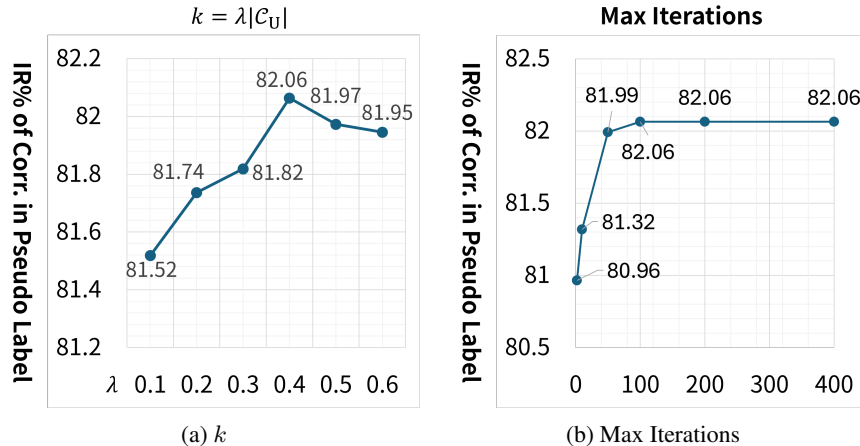
Figure 7: Sensitivity of hyperparameters in FGCM module.

Hyperparameter Sensitivity in FGCM   We evaluate the sensitivity of hyperparameters in FGCM, including the number k of putative correspondences to enlarge $\mathcal{C}^{i-1}$ and the max number of iterations. The results are illustrated in Fig. 7. The performance of INTEGER is relatively stable with respect to the hyperparameters, indicating the robustness of our method to hyperparameter settings. For best performance, we set $k = 0.4\mathcal{C}_{\mathrm{U}}$ and the max number of iterations to 100.

Table 4: Adaptability of INTEGER for Different Registration Networks

Network	mRR	RR@$d \in$				
		$[5, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$
FCGF	84.0	99.5	97.1	89.6	79.6	54.2
Predator	61.6	91.6	76.2	60.7	44.5	32.7

Adaptability to Different Registration Networks.   We assess the adaptability of INTEGER by training it with different registration networks, namely Predator[43] and FCGF[24]. The results presented in Table 4 indicate that INTEGER exhibits compatibility with diverse registration networks, underscoring its superior adaptability. However, we indeed observe that the performance gap between supervised and unsupervised settings widens when using Predator as the registration network. This may be attributed to the more sophisticated architecture of Predator compared to FCGF, which may require more accurate supervision signals for effective training. It will be our future work to improve the adaptability of INTEGER to more complicated registration networks.

Table 5: **Additional Generalizability Results Compared with Supervised 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@$d \in$				
				$[5, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$
	FCGF	–	60.1	93.9	84.5	58.9	39.2	23.9
KITTI	FCGF+C	–	60.3	95.9	84.5	61.1	35.8	24.1
↓	Predator	–	32.9	91.1	50.2	11.2	5.9	6.1
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**

16

Additional Generalizability Results. We provide additional generalizability results in Table 5. For supervised methods, FCGF and its variants surprisingly gains improved performance by generalize from KITTI, which may be attributed to the lower resolution of LiDAR in nuScenes(64 beams, 100m range in KITTI v.s. 32 beams, 70m range in nuScenes), which may hinder training from scratch. This phenomenon is also observed by existing efforts[12]. Our method show superior generalizability even compared with supervised methods.

(a) A Sample from KITTI

(b) A Sample from nuScenes
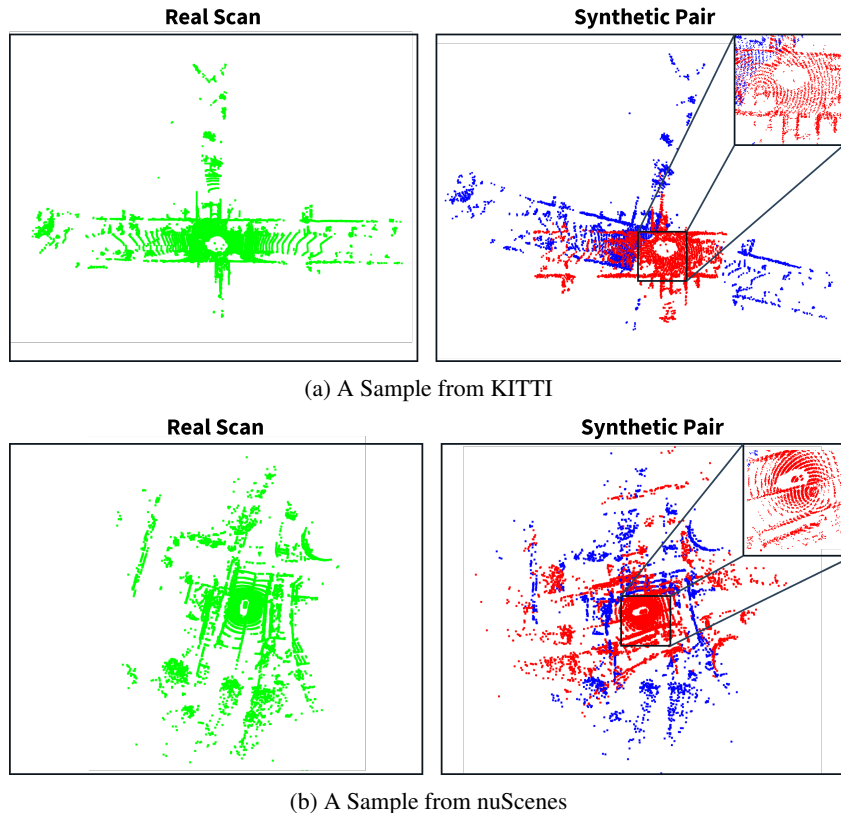
Figure 8: **Visualization of Generated Pairs in the Synthetic Pretraining Stage.** The synthetic pair on the right is generated from a single real scan on the right. We additionally visualize the result of periodic sampling in the top-right corner of synthetic pairs.

Qualitative Results on Synthetic Pairs. In the synthetic pretraining stage, we generate synthetic pairs from each real scan to train the teacher. We provide visualization on generated pairs in Fig. 8. The synthetic pair on the right is generated from a single real scan on the right. We additionally visualize the result of periodic sampling in the top-right corner of synthetic pairs. The periodic sampling strategy effectively simulates the irregular sampling of LiDAR point clouds. By randomly cropping the point clouds, the synthetic pairs are partially overlapped, which facilitates the training of the registration network.

17

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: It accurately reflects our contributions and scope. The scope of the paper is illustrated in Sec. 1. We detail our method in Sec. 3. Experimental validation are detailed in Sec. 4.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: It has been discussed in the Appendix. Furthermore, we also mention the location to find this discussion of limitations in the Conclusion section (Sec. 5) for clarity.

   Guidelines:

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

3. **Theory Assumptions and Proofs**

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

18

Answer: [NA]

Justification: Our work incorporates only experimental results.

Guidelines:

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

4. **Experimental Result Reproducibility**

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

Answer: [Yes]

Justification: Related information has been detailed in the paper. Sec. 4 includes brief information, whereas Sec. A.1, Sec. A.2 and Sec. A.3 in the Appendix include detailed information regarding reproducibility. Moreover, code will be made publicly available if the paper get accepted.

Guidelines:

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

19

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

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

Answer: [Yes]

Justification: Codes will be released. Instructions to faithfully reproduce the main experimental results are included in Sec. A.1, Sec. A.2 and Sec. A.3 in the Appendix.

Guidelines:

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

6. **Experimental Setting/Details**

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

Answer: [Yes]

Justification: They have been detailed in the appendix. See Sec. A.1, Sec. A.2 and Sec. A.3 in the Appendix.

Guidelines:

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

7. **Experiment Statistical Significance**

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

Answer: [No]

Justification: It is too computationally expensive for the proposed method. As shown in Sec. A.3 in the Appendix, our method currently requires a long training procedure. We have demonstrated the robustness and reliability of the proposed method by conducting experiments on generalizability. Moreover, we have to note that a considerable amount of existing research in point cloud registration[12, 4, 32, 38, 21, 51] does not include results on statistical significance, whereas a few[2, 5, 24] includes simple results such as variances for statistical results.

Guidelines:

20

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

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

   Answer: [Yes]

   Justification: This information has been detailed in Sec. A.3 in the Appendix.

   Guidelines:

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

9. **Code Of Ethics**

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

   Answer: [Yes]

   Justification: We have read the code of ethics. Our research involves no human subjects or participants, and the usage of KITTI and nuScenes datasets complies to their guidelines for privacy and ethical use cases. Information on reproducibility is provided in Sec. A.1, Sec. A.2 and Sec. A.3 in the Appendix. Our research on point cloud registration currently involves no ethical considerations, and our usage of synthetic data adheres to ethical guidelines, because it is randomized, happens on-the-fly during training, and thus will not harm reproducibility.

   Guidelines:

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

10. **Broader Impacts**

21

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

Answer: [Yes]

Justification: Our work performs positive societal impact, such as its benefits for autonomous driving. We discuss it in Sec. 1. We have not found negative societal impact yet for unsupervised point cloud registration.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

Answer: [NA]

Justification: Our paper poses no such risks. Our usage of data adhere to their official instructions and guidelines, as is stated in Sec. 4.

Guidelines:

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

12. **Licenses for existing assets**

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

Answer: [Yes]

22

Justification: They are credited and cited properly in Sec. 4. Release information on datasets has been included in their paper. KITTI uses CC BY-NC-SA 3.0 DEED, whereas nuScenes uses CC BY-NC-SA 4.0. We properly respect the license and terms of use. We also leverage official implementation for FCGF and Predator. We have cited their paper and their paper contains release information.

Guidelines:

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

13. **New Assets**

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

Answer: [NA]

Justification: Our paper does not release new data assets. Codes will be released *if the paper get accepted*. Instructions to faithfully reproduce the main experimental results are included in Sec. A.1, Sec. A.2 and Sec. A.3 in the Appendix. More detailed information such as exact commands for running the experiments, environment setup details and necessary data preprocessing steps will come with code release.

Guidelines:

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

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

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

Answer: [NA]

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

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

23

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

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

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

Answer: [NA]

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

Guidelines:

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

24