

Inlier Confidence Calibration for Point Cloud Registration

Yongzhe Yuan^{1,2} Yue Wu^{1,2*} Xiaolong Fan^{1,3} Maoguo Gong^{1,3} Qiguang Miao^{1,2} Wenping Ma⁴

¹MoE Key Lab of Collaborative Intelligence Systems, Xidian University

²School of Computer Science and Technology, Xidian University

³School of Electronic Engineering, Xidian University

⁴School of Artificial Intelligence, Xidian University

yyz@stu.xidian.edu.cn, {ywu, qgmiao}@xidian.edu.cn,

xiaolongfan@outlook.com, gong@ieee.org, wpma@mail.xidian.edu.cn

Abstract

Inliers estimation constitutes a pivotal step in partially overlapping point cloud registration. Existing methods broadly obey coordinate-based scheme, where inlier confidence is scored through simply capturing coordinate differences in the context. However, this scheme results in massive inlier misinterpretation readily, consequently affecting the registration performance. In this paper, we explore to extend a new definition called inlier confidence calibration (ICC) to alleviate the above issues. Firstly, we provide finely initial correspondences for ICC in order to generate high quality reference point cloud copy corresponding to the source point cloud. In particular, we develop a soft assignment matrix optimization theorem that offers faster speed and greater precision compared to Sinkhorn. Benefiting from the high quality reference copy, we argue the neighborhood patch formed by inlier and its neighborhood should have consistency between source point cloud and its reference copy. Based on this insight, we construct transformation-invariant geometric constraints and capture geometric structure consistency to calibrate inlier confidence for estimated correspondences between source point cloud and its reference copy. Finally, transformation is further calculated by the weighted SVD algorithm with the calibrated inlier confidence. Our model is trained in an unsupervised manner, and extensive experiments on synthetic and real-world datasets illustrate the effectiveness of the proposed method.

1. Introduction

With the rapid development of 3D data acquisition technology, point cloud data collected by LiDAR [45], Structured Light Sensors [30], and Stereo Cameras [9] has become

*Corresponding author.

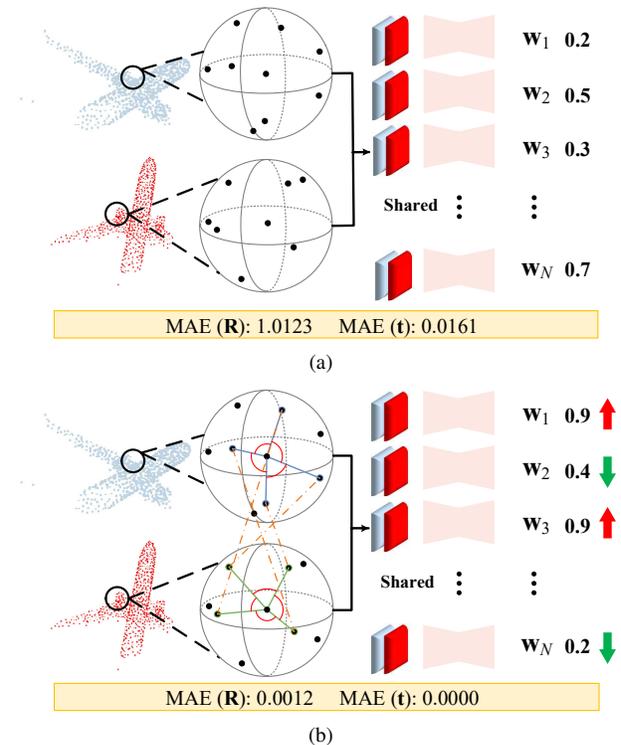


Figure 1. A toy example of proposed geometric constraints for inlier confidence calibration. The indicative center points pair is inlier pair in reality. (a) shows the situation without constraints. (b) demonstrates the situation with reliable geometric constraints. The experimental results are derived from ablation study.

ubiquitous in various 3D computer vision and robotics applications [6, 20], such as autopilot [15], surgical navigation [23], and simultaneous localization and mapping [11]. In such applications, rigid body point cloud registration plays an essential role, which aims to find a rigid transformation to align one point cloud to another.

The recent advances have been dominated by learning-

based methods. Most of these methods focus on solving the point cloud registration task in a supervised manner [16, 21, 27, 33, 38–41]. They require labeled data as the supervision signal to learn effective representations. In addition, obtaining labeled data is cumbersome and time-consuming, which may hinder applications in real world. To overcome this limitation, unsupervised point cloud registration is gradually garnering the interest of researchers [17, 24, 32]. However, in partially overlapping scenarios, learning-based methods must achieve robust inlier estimation, as inlier estimation serves as a fundamental source for identifying overlapping regions and predicting transformation. Existing methods broadly obey coordinate-based scheme, where inlier confidence are scored through simply capturing coordinate differences in the context. Meanwhile, the current guiding principle is that the reduced confidence of points pair indicates to the diminished probability that they represent inliers. But this assertion is not universally valid. The significant differences in coordinate present among point clouds pose a considerable challenge in differentiating inliers, thereby elevating the risk of misinterpretation. Thus, a crucial question remains to be addressed in order to make point cloud registration a success: *How to alleviate the bias in inlier confidence caused by massive coordinate differences?*

We provide answer to this question in Figure 1. Our key insight is that simple coordinates cannot provide strong evidence for scoring inlier confidence. In Figure 1a, the indicative center points pair is very difficult to distinguish because of coordinate differences, which diminish the confidence of this points pair and further classify it as an outlier pair. To alleviate this issue, we construct transformation-invariant geometric constraints for center points pair in Figure 1b. This strategy is based on two intuitive observations: (i) a rigorously defined property of rigid transformation, *i.e.*, it does not change the length of a vector and an angle between two vectors (isometric isomorphism). (ii) the neighborhood patch formed by inlier and its neighborhood should have consistency. Under strict constraints, the inlier confidence in Figure 1a can be effectively calibrate.

Motivated by the discussion above, in this paper, we explore to extend a new definition called inlier confidence calibration (ICC) for point cloud registration. Note that our method adheres to the pipeline of “pseudo correspondences-to-inlier estimation” [32, 38], which encompass critical task reference point cloud copy generation and scoring inlier confidence, respectively. A prerequisite for the success of ICC is the provision of high quality reference copy. Firstly, we provide finely initial correspondences to generate high quality reference point cloud copy corresponding to the source point cloud. In particular, we develop a soft assignment matrix optimization theorem that offers faster speed and greater precision compared to Sinkhorn [25]. This the-

orem is founded on the unique properties of the Gaussian distribution and cancels independent row and column normalization iteratively, along with the extra dustbin parameters in Sinkhorn. Benefiting from the high quality reference copy, we argue the neighborhood patch formed by inlier and its neighborhood should have consistency between source point cloud and its reference copy. Based on this insight, we construct transformation-invariant geometric constraints and capture geometric structure consistency to calibrate inlier confidence for each point pair between source point cloud and its reference copy. Finally, transformation are further calculated by the weighted SVD algorithm with the estimated correspondences and the calibrated inlier confidence. Our model is trained in an unsupervised manner, and extensive experiments on the synthetic datasets ModelNet40 [35], Augmented ICL-NUIM [7] and real-world dataset 7Scenes [43] demonstrate our model achieves competitive performance.

To summarize, our contributions are as follows:

- An effective ICC method for point cloud registration by constructing transformation-invariant geometric constraints.
- A soft assignment matrix optimization theorem that offers faster speed and greater precision compared to Sinkhorn, which can provide finely initial correspondences for ICC in order to generate the reference point cloud copy corresponding to the source point cloud.
- Under the unsupervised setting, the constructed geometric constraints provides self-supervised signal and optimization objective for model training.

2. Related Works

Traditional point cloud registration methods. Most traditional methods require a good initial transformation and converge to the local minima near the initialization point. One of the most profound methods is the Iterative Closest Point (ICP) algorithm [4], which begins with an initial transformation and iteratively alternates between solving two trivial subproblems: finding the closest points as correspondence under current transformation, and computing optimal transformation by SVD [19] based on identified correspondences. Though ICP can complete a high-precision registration, it is susceptible to the initial perturbation. In recent years, variants of ICP have been proposed [5, 14, 28, 31], and they can improve the defects of ICP and enhance the registration accuracy [3]. However, these methods retain a few essential drawbacks. Firstly, they depend strongly on the initialization. Secondly, it is difficult to integrate them into the deep learning pipeline as they lack differentiability. Thirdly, explicit estimation of corresponding points leads to quadratic complexity scaling with the number of points [29], which can introduce significant computational challenges.

Learning-based registration methods. At present, most learning-based methods are based on supervision [12, 27, 33, 40]. PointnetLK [1] is a classical correspondence-free method, which calculates global feature descriptors through PointNet and iteratively uses the Inverse Compositional formulation and Lucas & Kanade algorithm [22, 37] to minimize distance between the descriptors to achieve registration. RPM-Net [38] utilizes the differentiable Sinkhorn layer and annealing to get soft assignments of point correspondences from hybrid features learned from both spatial coordinates and local geometry.

Recently, unsupervised point cloud registration has gained increasing attention due to its applicability in scenarios where labeled training data is scarce or unavailable. Some methods have been proposed to address this challenge and achieved promising results [2, 8, 10, 17, 18]. Feature-metric point cloud registration framework (FMR) [16] enforces the optimisation of registration by minimising a feature-metric projection error with a autoencoder-based network. CEMNet [17] treats the point cloud registration as a reinforcement learning task. However, the current methods lack a concept of calibrating inlier confidence, which poses a potential challenge for partially overlapping point cloud registration. In comparison, our method provide a reliable ICC by constructing strict constraints.

3. Method

Given two point clouds: source point cloud $\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}$ and reference point cloud $\mathcal{Q} = \{\mathbf{q}_j \in \mathbb{R}^3 \mid j = 1, \dots, M\}$, where each point is represented as a vector of (x, y, z) coordinates. Our goal is to estimate a rigid transformation $\{\mathbf{R}, \mathbf{t}\}$ which accurately aligns \mathcal{P} and \mathcal{Q} , with a 3D rotation $\mathbf{R} \in SO(3)$ and a 3D translation $\mathbf{t} \in \mathbb{R}^3$.

In this paper, our method adheres to the pipeline of “pseudo correspondences-to-inlier estimation” [32, 38], which encompass two critical task: reference point cloud copy generation and scoring inlier confidence. A prerequisite for the success of ICC is the provision of high quality reference copy. Firstly, we provide finely initial correspondences to generate high quality reference point cloud copy corresponding to the source point cloud (Section 3.1). In particular, we develop a soft assignment matrix optimization theorem that offers faster speed and greater precision compared to Sinkhorn. Benefiting from the high quality reference copy, we argue the neighborhood patch formed by inlier and its neighborhood should have consistency between source point cloud and its reference copy. Based on this insight, we construct transformation-invariant geometric constraints and capture geometric structure consistency to calibrate inlier confidence for each point pair between source point cloud and its reference copy (Section 3.2). The pipeline is illustrated in Figure 2.

3.1. Finely Initial Correspondences

To generate high quality reference point cloud copy corresponding to the source point cloud, constructing the matching map between \mathcal{P} and \mathcal{Q} to refine the initial correspondences is an essential prerequisite. Different from previous work [32, 38], the model input is expanded into four point clouds through a 1-NN strategy, a methodology whose advantages have been substantiated in [42].

Let 1-NN point cloud $\hat{\mathcal{P}}$ and $\hat{\mathcal{Q}}$ as the closest point for each point in \mathcal{P} and \mathcal{Q} , respectively. Next, we construct a local patch \mathcal{N}_o with K -nearest neighborhood for each point and extract local features with Dynamic Graph CNN [34] for input point cloud and 1-NN point cloud. Associated learned features are denoted as $\mathcal{F}_{\mathcal{P}}$, $\mathcal{F}_{\hat{\mathcal{P}}}$, $\mathcal{F}_{\mathcal{Q}}$ and $\mathcal{F}_{\hat{\mathcal{Q}}}$, respectively. Then, we aim to learn a soft assignment matrix \mathbf{M} ($\hat{\mathbf{M}}$) to explicitly indicate the matching between any two points of \mathcal{P} ($\hat{\mathcal{P}}$) and \mathcal{Q} ($\hat{\mathcal{Q}}$). Ideally, there is only a single prominent element dominating each row and column. To further enhance \mathbf{M} ($\hat{\mathbf{M}}$), the widely adopted method is Sinkhorn [25], which performs the softmax operation on column-wise and row-wise iteratively and alternatively. However, the efficiency of Sinkhorn is limited by iterative strategy and extra dustbin parameters. To tackle this issue, we develop a soft assignment matrix optimization theorem to improve \mathbf{M} ($\hat{\mathbf{M}}$) based on the following unique properties of the Gaussian distribution, which is a promotion for [44].

Definition 1. (Three-sigma rule). *The variation or dispersion of data can be measured by the three-sigma rule, which is based on the properties of the normal distribution curve, indicating that within a dataset 99.7% of data values fall within the range of three standard deviation from the mean.*

Recalling definitions of three-sigma rule, we have the following Theorem for $\mathbf{M} \in \mathbb{R}^{N \times M}$ (maybe a non-square matrix), and $\hat{\mathbf{M}}$ follows the same principles.

Theorem 1. *Suppose the $\mathbf{M}_{i,j}$ obeys a series of Gaussian distributions, \exists two constants ϵ, ξ , such that there is only a single prominent element dominating each row by three-sigma rule.*

Proof. Given μ_i and σ_i are the mean and standard deviation of the i -th row of \mathbf{M} . $\mathbf{M}_{i,j}$ is normalized in row-wise and the normalized elements follow a standard normal distribution, i.e.:

$$h_{i,j} = \frac{\mathbf{M}_{i,j} - \mu_i}{\sigma_i} \sim \mathcal{N}(0, 1), j = 1, \dots, M. \quad (1)$$

Based on above setting, we set tuning constants ϵ and ξ to $h_{i,j}$ and obtain

$$\tilde{h}_{i,j} = \epsilon \cdot h_{i,j} + \xi \sim \mathcal{N}(\xi, \epsilon^2 \mathbb{I}). \quad (2)$$

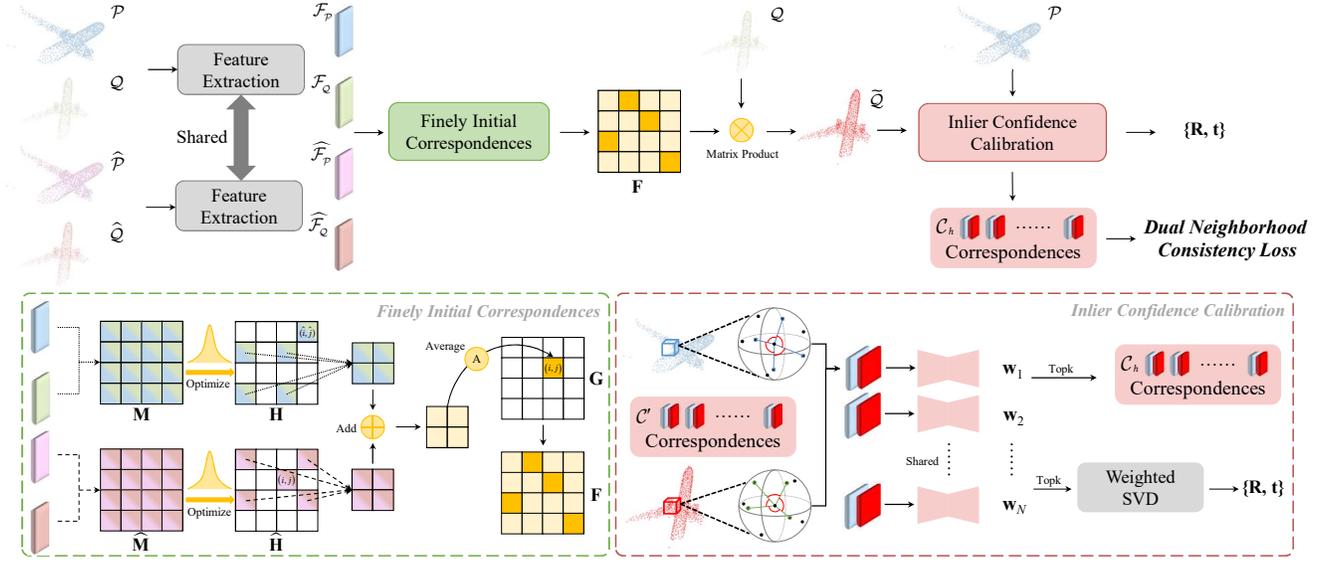


Figure 2. The pipeline of the proposed method. We first extract local features of input point cloud and 1-NN point cloud with DGCNN. Then, we construct matching map to generate high quality reference point cloud copy \tilde{Q} for fining initial correspondences. Based on the observation that the neighborhood graph formed by inlier and its neighborhood should have geometric structure consistency between \mathcal{P} and \tilde{Q} , we construct transformation-invariant geometric structure constraints (edge and included angle) and captures their consistency to score and calibrate the inlier confidence for each estimated correspondences between \mathcal{P} and \tilde{Q} . Finally, the transformation is estimated with SVD method via the calibrated inlier confidence.

For i -th row of $\mathbf{H} \in \mathbb{R}^{N \times M}$ ($\tilde{h}_{i,j} \subset \mathbf{H}$), the number of elements whose values are not less than a threshold Δ is calculated:

$$b_i = |\{\tilde{h}_{i,j} \mid \tilde{h}_{i,j} \geq \Delta, j = 1, 2, \dots, M\}|. \quad (3)$$

We aim to prompt b_i closer to 1, which means that in each row there is only a single element dominates the row. In addition, the N rows of \mathbf{H} can be regarded as N independent and identically distributed events, thus the set $\{b_1, b_2, \dots, b_N\}$ also follows a Gaussian distribution with the expectation μ_b and variance σ_b depending on the tuning constants ϵ, ξ . Hence, by the Definition 1, we can select suitable values of ϵ, ξ to control the bound $[\mu_b - 3\sigma_b, \mu_b + 3\sigma_b]$, which is centered around 1, such that the expectation of only a single element dominates the row can be implemented. \square

Theorem 1 tells us that column-wise and row-wise iteratively and alternatively may be not essential. Operating solely on the rows can achieve the objective with higher efficiency and accuracy (Please refer to Section 4.4).

Based on the scores in soft assignment matrix \mathbf{H} and $\hat{\mathbf{H}}$, we construct dual neighborhood matching map by fusing and averaging matching scores of each \mathcal{N}_o :

$$\mathbf{G}_{i,j} = \frac{1}{K} \sum_{\mathbf{p}_{i'} \in \mathcal{N}_{\mathbf{p}_i}} \sum_{\mathbf{q}_{j'} \in \mathcal{N}_{\mathbf{q}_j}} \mathbf{H}_{i',j'} + \frac{1}{K} \sum_{\hat{\mathbf{p}}_{i'} \in \mathcal{N}_{\hat{\mathbf{p}}_i}} \sum_{\hat{\mathbf{q}}_{j'} \in \mathcal{N}_{\hat{\mathbf{q}}_j}} \hat{\mathbf{H}}_{i',j'}. \quad (4)$$

Finally, we formulate the final matching map as:

$$\mathbf{F}_{i,j} = \text{softmax}(-\mathbf{D}'_{i,1}, -\mathbf{D}'_{i,2}, \dots, -\mathbf{D}'_{i,M})_j, \quad (5)$$

$$\mathbf{D}'_{i,j} = \exp(\alpha - \mathbf{G}_{i,j}) * (\mathbf{D}_{i,j} + \hat{\mathbf{D}}_{i,j}),$$

where $\mathbf{D}_{i,j} = \|\mathcal{F}_{\mathbf{p}_i} - \mathcal{F}_{\mathbf{q}_j}\|_2$ and $\hat{\mathbf{D}}_{i,j} = \|\mathcal{F}_{\hat{\mathbf{p}}_i} - \mathcal{F}_{\hat{\mathbf{q}}_j}\|_2$ denote the Euclidean distance between learned local features, $\mathbf{D}'_{i,j}$ is negatively related to the matching score $\mathbf{G}_{i,j}$. We utilize the exponential strategy to control the changing ratio, along with a hyper-parameter α to control the influence of the dual neighborhood matching [32].

Based on the high quality matching map $\mathbf{F} \in \mathbb{R}^{N \times M}$, we generate the reference point cloud copy $\tilde{Q} \in \mathbb{R}^{N \times 3}$ including finely initial correspondences for each point \mathbf{p}_i in source point cloud \mathcal{P} :

$$\mathcal{C}' = \{(\mathbf{p}_i, \tilde{\mathbf{q}}_i) \mid i = 1, \dots, N, \tilde{\mathbf{q}}_i \in \tilde{Q}\}, \quad (6)$$

where

$$\tilde{\mathbf{q}}_i = \sum_{j=1}^M \mathbf{F}_{i,j} \cdot \mathbf{q}_j. \quad (7)$$

3.2. Inlier Confidence Calibration

After generating reference copy \tilde{Q} , we score inlier confidence for each point pair in \mathcal{C}' . However, existing methods broadly obey coordinate-based scheme, where inlier confidence are scored through simply capturing coordinate differences in the context, and the current guiding principle

is that the reduced confidence of points pair indicates to the diminished probability that they represent inliers. But this assertion is not universally valid. The significant differences in coordinate pose a challenge in differentiating inliers, thereby elevating the risk of misinterpretation.

In this section, we explore to extend a new definition called ICC to alleviate the above issues. Benefiting from the high quality matching map \mathbf{F} in Section 3.1, we argue that the neighborhood patch formed by inlier and its neighborhood should have consistency between source point cloud and its reference copy, *i.e.*, if \mathbf{p}_i is an inlier, its neighborhood points also tend to be the inliers and $\tilde{\mathbf{q}}_i$ have geometric structure neighborhood consistency in the reference copy. Conversely, if \mathbf{p}_i is an outlier, its neighborhood may also be the outliers and spatial structure is prone to be significantly different from the reference copy neighborhood. Based on this observation, we construct transformation-invariant geometric constraints and formulate geometric structure consistency between $\mathcal{N}_{\mathbf{p}_i}$ and $\mathcal{N}_{\tilde{\mathbf{q}}_i}$ to calibrate inlier confidence for estimated correspondences between source point cloud and its reference copy.

We first construct a learnable neighborhood graph by transformation-invariant geometric structure constraints of $\mathcal{N}_{\mathbf{p}_i}$ and $\mathcal{N}_{\tilde{\mathbf{q}}_i}$, which consists of edge representation and angle representation:

$$\begin{aligned} \mathbf{e}_{i,k}^{\mathbf{p}} &= \mathbf{p}_i - \mathbf{p}_k, \mathbf{a}_{r,s}^{\mathbf{p}} = \angle(\mathbf{e}_{i,r}^{\mathbf{p}}, \mathbf{e}_{i,s}^{\mathbf{p}}), \\ \tilde{\mathbf{e}}_{i,k}^{\mathbf{q}} &= \tilde{\mathbf{q}}_i - \tilde{\mathbf{q}}_k, \tilde{\mathbf{a}}_{r,s}^{\mathbf{q}} = \angle(\tilde{\mathbf{e}}_{i,r}^{\mathbf{q}}, \tilde{\mathbf{e}}_{i,s}^{\mathbf{q}}), \\ i &\neq k, k = 1, \dots, K \quad r, s \in \{1, \dots, K\} \end{aligned} \quad (8)$$

where \mathbf{p}_k and $\tilde{\mathbf{q}}_k$ are the points in $\mathcal{N}_{\mathbf{p}_i}$ and $\mathcal{N}_{\tilde{\mathbf{q}}_i}$. Moreover, these representations express sensitive and discriminative geometric structure in the point cloud, and provide adequate geometric cues for subsequent pipeline.

In order to better capture the neighborhood relevance and promote contextual message propagation, we utilize Multi-layer Perceptron (MLP) f_θ with parameters θ to fuse representations and characterize the consistency between the neighborhoods by the subtraction of the fused geometric representations:

$$\mathbf{d}_{i,k} = f_\theta(\text{concat}(\mathbf{e}_{i,k}^{\mathbf{p}}, \mathbf{a}_{r,s}^{\mathbf{p}})) - f_\theta(\text{concat}(\tilde{\mathbf{e}}_{i,k}^{\mathbf{q}}, \tilde{\mathbf{a}}_{r,s}^{\mathbf{q}})). \quad (9)$$

Next, the model further adaptively learns the attention coefficients $\delta_{i,k}$ of each geometric structure consistency:

$$\delta_{i,k} = \text{softmax}(f_\mu(\mathbf{d}_{i,1}), f_\mu(\mathbf{d}_{i,2}), \dots, f_\mu(\mathbf{d}_{i,K}))_k, \quad (10)$$

where f_μ is another MLP with parameters μ . Then, inlier confidence \mathbf{w}_i of correspondence $(\mathbf{p}_i, \tilde{\mathbf{q}}_i)$ is calculated by aggregating the geometric structure consistency weighted:

$$\mathbf{w}_i = 1 - \text{Tanh}\left(\left|l\left(\sum_{k=1}^K \delta_{i,k} * \mathbf{d}_{i,k}\right)\right|\right), \quad (11)$$

where l is a linear function. Through strict geometric constraints, the scored inlier confidence can be effectively calibrated, ensuring the sustained high credibility of \mathbf{w}_i . However, some potential repetitive structures, textureless structures and other uncertain factors may still exist. In order to enhance the robustness of the proposed method and reduce the bias caused by the above issues, we select inlier correspondences from \mathcal{C}' with the largest N_c inlier confidence:

$$\mathcal{C}_h = \{(\mathbf{p}_h, \tilde{\mathbf{q}}_h) \mid h \in \text{Topk}(\mathbf{w}_i), i = 1, \dots, N_c\}. \quad (12)$$

Finally, transformation $\{\mathbf{R}_{est}, \mathbf{t}_{est}\}$ can be solved in closed form using weighted SVD, which has been shown to be differentiable in [26]:

$$\mathbf{R}_{est}, \mathbf{t}_{est} = \min_{\mathbf{R}, \mathbf{t}} \sum_{(\mathbf{p}_i, \tilde{\mathbf{q}}_i) \in \mathcal{C}_h} \mathbf{w}_i \|\mathbf{R} \cdot \mathbf{p}_i + \mathbf{t} - \tilde{\mathbf{q}}_i\|_2^2. \quad (13)$$

Besides, ICC provides useful byproduct for unsupervised learning, which can be reliable self-supervised signal. More details can be seen in the following section.

3.3. Optimization

In this section, we conduct four loss functions for model optimization. Moreover, the proposed model is trained in an unsupervised manner instead of using the ground-truth transformations.

Global Consistency Loss. We investigate the global consistency loss between the final transformed source point cloud \mathcal{P}' and the reference point cloud \mathcal{Q} . Specifically, the Huber function is utilized to assemble the global consistency loss, which is defined as follow:

$$\begin{aligned} \mathcal{L}_{gc} &= \sum_{\mathbf{p}' \in \mathcal{P}'} H_\beta \left(\min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{p}' - \mathbf{q}\|_2^2 \right) \\ &+ \sum_{\mathbf{q} \in \mathcal{Q}} H_\beta \left(\min_{\mathbf{p}' \in \mathcal{P}'} \|\mathbf{q} - \mathbf{p}'\|_2^2 \right). \end{aligned} \quad (14)$$

However, relying solely on global consistency loss is detrimental to the accuracy and reliability of our model. Since the model may still potentially converge to sub-optimization due to the existing outliers, leading massive potential information of point cloud is wasted. Hence, it is critical to mine the potential self-supervised signals in the point cloud and construct loss functions based on other existing elements.

Dual Neighborhood Consistency Loss. Based on reliable inlier correspondences in Equation 12, we denote the inliers set of the source and its reference point cloud copy as $\mathbf{X} \in \mathbb{R}^{N_c \times 3}$ and $\mathbf{Y} \in \mathbb{R}^{N_c \times 3}$, respectively. We utilize the neighborhood between the inliers to construct consistency objective, which aims to minimize the registration error between each neighborhood $\mathcal{N}_{\mathbf{x}_i}$ and $\mathcal{N}_{\mathbf{y}_i}$:

$$\mathcal{L}_{in} = \sum_{\mathbf{x}_i \in \mathbf{X}, \mathbf{y}_i \in \mathbf{Y}} \sum_{\mathbf{p}_j \in \mathcal{N}_{\mathbf{x}_i}, \tilde{\mathbf{q}}_j \in \mathcal{N}_{\mathbf{y}_i}} \|\mathbf{R}_{est} \mathbf{p}_j + \mathbf{t}_{est} - \tilde{\mathbf{q}}_j\|_2, \quad (15)$$

where \mathcal{N}_{x_i} is transformed by \mathcal{N}_{y_i} .

Geometric Structure Consistency Loss. The geometric signal buried in point cloud is readily ignored, which hinders ICC. To address this issue, we design a geometric neighborhood loss with the reliable geometric self-supervised signal proposed in Section 3.2:

$$\mathcal{L}_{gs} = \sum_{x_i \in \mathbf{X}, y_i \in \mathbf{Y}} \sum_{p_j \in \mathcal{N}_{x_i}, \tilde{q}_j \in \mathcal{N}_{y_i}} \left\| \mathbf{e}_{i,j}^p - \tilde{\mathbf{e}}_{i,j}^q \right\|_2 + \sum_{x_i \in \mathbf{X}, y_i \in \mathbf{Y}} \sum_{p_j \in \mathcal{N}_{x_i}, \tilde{q}_j \in \mathcal{N}_{y_i}} \left\| \mathbf{a}_{r,s}^p - \tilde{\mathbf{a}}_{r,s}^q \right\|_2, \quad (16)$$

where $\mathbf{a}_{r,s}^p$ and $\tilde{\mathbf{a}}_{r,s}^q$ is calculated by $\mathbf{e}_{i,j}^p$ and $\tilde{\mathbf{e}}_{i,j}^q$ shown in Equation 8.

Spatial Consistency Loss. We further explore to eliminate the spatial difference between the estimated correspondence and the real correspondence for each selected inlier \mathbf{x}_i and utilize spatial consistency loss with cross-entropy to sharpen matching map \mathbf{G} :

$$\mathcal{L}_{sc} = -\frac{1}{|\mathbf{X}|} \sum_{x_i \in \mathbf{X}} \sum_{j=1}^M \mathbb{I}[j = \arg \max_{j'} \mathbf{G}_{i,j'}] \log \mathbf{G}_{i,j}, \quad (17)$$

where $\mathbb{I}[\cdot]$ is the Iverson bracket. Spatial consistency loss encourages to improve the matching probability and thus the estimated correspondence point in reference copy tends to have an stable position.

Overall Loss. Since our work utilize an iterative scheme, we compute the loss at each iteration N_l and have the weighted sum loss:

$$\mathcal{L} = \sum_{l=1}^{N_l} (\mathcal{L}_{gc}^l + \gamma \mathcal{L}_{in}^l + \rho \mathcal{L}_{gs}^l + \lambda \mathcal{L}_{sc}^l), \quad (18)$$

where γ , ρ and λ are trade-off parameters to control corresponding loss function.

4. Experiments

4.1. Experimental Setup

We compare our method to traditional methods and recent state of the arts learning-based methods. The traditional methods include ICP [4], FGR [16] and FPFH + RANSAC [13]. The recent learning based methods include IDAM [21], FMR [16], RPMNet [38], CEMNet [17], REGTR [39], GeoTransformer [27] and RIE [32]. For consistency with previous work, we measure Mean Isotropic Error (MIE) and Mean Absolute Error (MAE).

4.2. ModelNet40 Dataset and Evaluation

Noted that, to simulate partial-to-partial registration, the reference point cloud \mathcal{Q} and the source point cloud \mathcal{P} are cropped respectively, and retain 70% of the points.

Method	MAE(R)	MAE(t)	MIE(R)	MIE(t)
ICP [4] (○)	3.4339	0.0114	6.7706	0.0227
FGR [16] (○)	0.5972	0.0021	1.1563	0.0041
FPFH+RANSAC [13] (○)	0.7031	0.0025	1.2772	0.0050
IDAM [21] (▲)	0.4243	0.0020	0.8170	0.0040
RPMNet [38] (▲)	0.0051	0.0000	<u>0.0201</u>	0.0000
FMR [16] (▲)	3.6497	0.0101	7.2810	0.0200
REGTR [39] (▲)	0.1894	0.0015	0.3493	0.0030
GeoTransformer [27] (▲)	0.5970	0.0057	0.9010	0.0102
CEMNet [17] (△)	0.1385	<u>0.0001</u>	0.2489	<u>0.0002</u>
RIE [32] (△)	<u>0.0033</u>	0.0000	0.0210	0.0000
Ours (△)	0.0012	0.0000	0.0192	0.0000

Table 1. Evaluation results on ModelNet40. Bold indicates the best performance and underline indicates the second-best performance. (○), (▲) and (△) denote the traditional, supervised and unsupervised methods, respectively.

Unseen Objects. Our models are trained and tested on datasets comprising of samples belonging to the same categories, and both the training and test sets are obtained without any preprocessing or manipulation. We apply a random transformation on the reference point cloud \mathcal{Q} to generate corresponding source point cloud \mathcal{P} . Table 1 shows quantitative results of the various algorithms under current experimental settings. The proposed method substantially outperforms all baseline in all metrics. We can observe our method can even outperform the supervised IDAM, RPMNet, FMR, REGTR and GeoTransformer by a large margin. Benefiting from ICC, our method attains highly accurate registration and improves the registration accuracy by an order of magnitude. In order to show the effect of our proposed approach clearly, a qualitative comparison of the registration results can be found in Figure 3. Our method ensures minimal impact on changing of the shape, and achieves the best performance even on asymmetric shape.

Unseen Categories. To verify the generalization ability on categories, we train the models on the first 20 categories and test on the remaining unseen categories. The results are summarized in Table 2. We can observe that the majority of baseline consistently exhibit lower performance on the unseen categories, especially learning-based methods. In contrast, traditional algorithms are less susceptible to this issue due to the insensitivity of handcrafted methods to shape variance [36]. Our registration process remains highly precise, achieving the lowest error across all metrics, while also maintaining acceptable levels of fluctuation.

Gaussian Noise. In order to assess performance in the presence of noise, which is commonly encountered in real-world point clouds, we train our model on noise-free data and then evaluate all baselines using a test set featuring Gaussian noise. We randomly and independently generate noisy points to introduce noise into in source point cloud and reference point cloud by sampling from $\mathcal{N}(0, 0.5)$ and

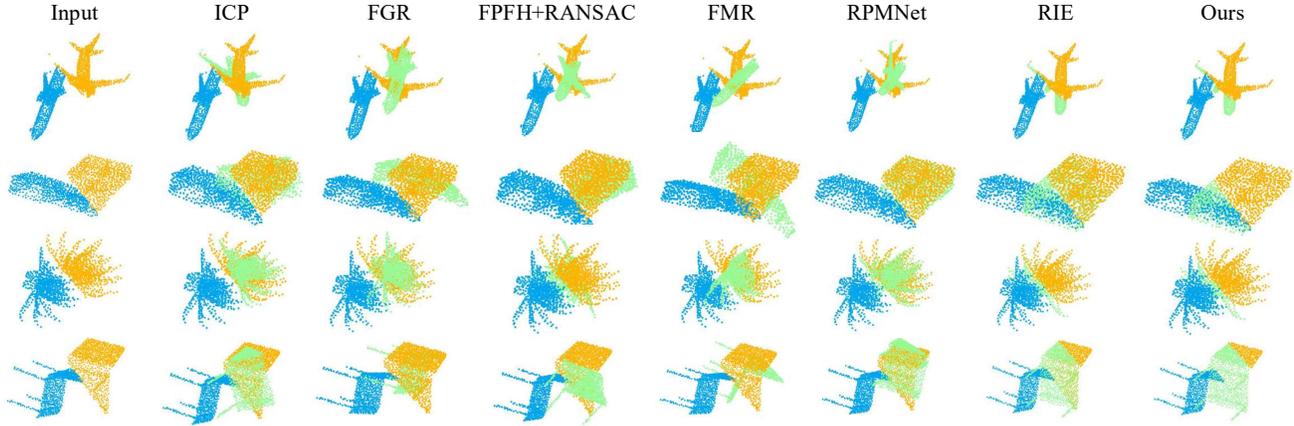


Figure 3. Qualitative comparison of the registration results on unseen objects data (blue: source point cloud, yellow: reference point cloud, green: transformed source point cloud).

clipped to $[-1.0, 1.0]$. This experiment is significantly challenging, as constructing matching map and ICC become much more difficult. As shown in Table 2, our method outperforms other baselines. The experimental results demonstrate that our method is robust to the noise. In addition, experimental results indirectly confirms the generality and high adaptability of ICC in point cloud registration with partially overlapping.

Method	MAE(R)	MAE(t)	MIE(R)	MIE(t)
Unseen Categories				
ICP [4] (○)	3.6099	0.0116	7.0556	0.0228
FGR [16] (○)	0.4579	0.0016	0.8442	0.0032
FPFH+RANSAC [13] (○)	0.4427	0.0021	0.9447	0.0043
IDAM [21] (▲)	0.4809	0.0028	0.9157	0.0055
RPMNet [38] (▲)	0.0064	0.0001	<u>0.0207</u>	<u>0.0001</u>
FMR [16] (▲)	3.8594	0.0114	7.6450	0.0225
CEMNet[17] (△)	0.0804	<u>0.0002</u>	0.1405	0.0003
RIE [32] (△)	<u>0.0059</u>	0.0000	0.0228	<u>0.0001</u>
Ours (△)	0.0022	0.0000	0.0189	0.0000
Gaussian Noise				
ICP [4] (○)	4.6441	0.0167	9.2194	0.0333
FGR [16] (○)	1.0676	0.0036	2.0038	0.0072
FPFH+RANSAC [13] (○)	1.4316	0.0061	2.5345	0.0120
IDAM [21] (▲)	2.3076	0.0124	4.5332	0.0246
RPMNet [38] (▲)	1.5890	0.0175	2.9830	0.0378
FMR [16] (▲)	18.0355	0.0536	35.7986	0.1063
CEMNet[17] (△)	10.7026	0.0393	21.1836	0.0781
RIE [32] (△)	<u>0.0431</u>	<u>0.0004</u>	<u>0.0776</u>	<u>0.0008</u>
Ours (△)	0.0088	0.0001	0.0245	0.0002

Table 2. Evaluation results on ModelNet40. Bold indicates the best performance and underline indicates the second-best performance. (○), (▲) and (△) denote the traditional, supervised and unsupervised methods, respectively.

4.3. Other Datasets and Evaluation

We further conduct comparison evaluation on other datasets: 7Scenes and Augmented ICL-NUIM. We sample the reference point clouds to 2,048 points and randomly sample three Euler angle rotations within $[0^\circ, 45^\circ]$ and translations within $[-0.5, 0.5]$ on each axis as the rigid transformation to obtain source point clouds, then down-sample the point clouds to 1,536 points to generate the partial data. As demonstrated in Table 4, our method exhibits extremely higher registration precision on all criteria on Augmented ICL-NUIM and 7Scenes, especially the rotation error. We can summarize our method has best performance and is comfortable with real-world dataset.

4.4. Ablation Study and Analysis

Soft Assignment Matrix Optimization Theorem. We compare the developed theorem with the Sinkhorn, and evaluate precision and efficiency against Sinkhorn, which iterate 1 (Sinkhorn-1) and 5 (Sinkhorn-5) times. As shown in Figure 4, we can observe that under the same tolerant error, the developed theorem demonstrates superior performance, especially the optimization for rotation. Additionally, the proposed theorem improves the efficiency up to nearly $2\times$ compared to Sinkhorn-1 and nearly up to $5\times$ compared to Sinkhorn-5. It needs to be emphasized once again that the proposed theorem has a smaller computational complexity than Sinkhorn as shown in Table 5, since this theorem does not normalize the rows and columns repeatedly, and only operate on the rows.

Geometric Constraints. To investigate whether geometric constraints are meaningful to ICC, we conduct an experiment to verify our motivation. Experimental results are shown in the first row of Table 3. Compared to constraints-free, which score inlier confidence by simple coordinates, geometric constraints effectively generate more convincing

Geometric Constraints	\mathcal{L}_{gc}	\mathcal{L}_{in}	\mathcal{L}_{gs}	\mathcal{L}_{sc}	ModelNet40				7Scenes			
					MAE(R)	MAE(t)	MIE(R)	MIE(t)	MAE(R)	MAE(t)	MIE(R)	MIE(t)
	✓	✓	N/A	✓	1.0123	0.0161	1.8398	0.0350	2.2238	0.0148	4.3698	0.0286
✓	✓	✓	✓	✓	0.3998	0.0061	0.7665	0.0129	0.0042	0.0000	0.0217	0.0001
✓	✓	✓	✓	✓	22.5199	0.1340	42.4428	0.2710	3.3087	0.0340	5.8770	0.0657
✓	✓	✓	✓	✓	0.0070	0.0001	0.0235	0.0002	0.0037	0.0000	0.0186	0.0001
✓	✓	✓	✓	✓	0.1125	0.0014	0.01862	0.0030	0.0142	0.0001	0.0345	0.0003
✓	✓	✓	✓	✓	0.0012	0.0000	0.0192	0.0000	0.0033	0.0000	0.0181	0.0001

Table 3. Ablation study of different components. \mathcal{L}_{gc} , \mathcal{L}_{in} , \mathcal{L}_{gs} and \mathcal{L}_{sc} : Each Loss Function in Equation 18.

Method	MAE(R)	MAE(t)	MIE(R)	MIE(t)
Augmented ICL-NUIM				
ICP [4] (○)	2.4022	0.0699	4.4832	0.1410
FGR [16] (○)	2.2477	0.0808	4.1850	0.1573
FPFH+RANSAC [13] (○)	1.2349	0.0429	2.3167	0.0839
IDAM [21] (▲)	4.4153	0.1385	8.6178	0.2756
RPMNet [38] (▲)	0.3267	0.0125	0.6277	0.0246
FMR [16] (▲)	1.1085	0.0398	2.1323	0.0786
CEMNet[17] (△)	0.2374	<u>0.0005</u>	0.3987	<u>0.0010</u>
RIE [32] (△)	<u>0.0492</u>	0.0023	<u>0.0897</u>	0.0049
Ours (△)	0.0048	0.0002	0.0199	0.0003
7Scenes				
ICP [4] (○)	6.0091	0.0130	13.0484	0.0260
FGR [16] (○)	0.0919	0.0004	0.1705	0.0008
FPFH+RANSAC [13] (○)	1.2325	0.0062	2.1875	0.0124
IDAM [21] (▲)	5.6727	0.0303	11.5949	0.0629
RPMNet [38] (▲)	0.3885	0.0021	0.7649	0.0042
FMR [16] (▲)	2.5438	0.0072	4.9089	0.0150
CEMNet[17] (△)	0.0559	<u>0.0001</u>	0.0772	<u>0.0003</u>
RIE [32] (△)	<u>0.0121</u>	<u>0.0001</u>	<u>0.0299</u>	0.0001
Ours (△)	0.0033	0.0000	0.0181	0.0001

Table 4. Evaluation results on Augmented ICL-NUIM and 7Scenes. Bold indicates the best performance and underline indicates the second-best performance. (○), (▲) and (△) denote the traditional, supervised and unsupervised methods, respectively.

	Complexity
Sinkhorn with dustbin parameters	$\mathcal{O}((N+1) \times (M+1))$
Sinkhorn with iterate G times	$\mathcal{O}(G \times (N \times M))$
Proposed theorem	$\mathcal{O}(N)$

Table 5. Computational complexity of Sinkhorn and the proposed theorem comparison.

scoring mechanism. Noted that geometric constraints provide a self-supervised signal for our model, and \mathcal{L}_{gs} is designed based on this signal. Therefore, when geometric constraints is deleted, the loss function \mathcal{L}_{gs} should not participate in optimization.

Loss Function. We evaluate the performance of the model

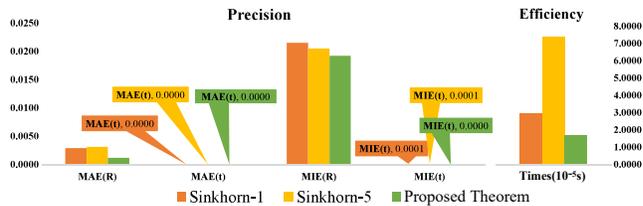


Figure 4. Comparisons between the proposed theorem and the Sinkhorn in terms of precision and efficiency.

with different loss functions. Comprehensively, lacking any part of \mathcal{L} will degrade the performance of the model. Comparing 2~5 rows in Table 3, the error observed in the third row is extremely large, primarily due to the absence of any optimization objective related to transformation in the loss function. This significantly reduces the registration performance of the model. In particular, this ablation study confirms the byproduct of ICC \mathcal{L}_{gs} can provide reliable and effective self-supervised signal for model optimization.

5. Conclusion

In this paper, we explore to extend a new definition called inlier confidence calibration (ICC). Existing methods broadly obey coordinate-based scheme, where inlier confidence is scored through simply capturing coordinate differences in the context. However, this scheme results in massive inlier misinterpretation readily, consequently affecting the registration accuracy. Thus, we construct transformation-invariant geometric constraints and capture geometric structure consistency to calibrate inlier confidence. Extensive experiments on synthetic and real-world datasets demonstrate the effectiveness of the proposed methods. In the future, we would like to extend our method to crossmodality (e.g., 2D-3D) registration with richer applications.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (62036006, 62276200), the CAAI-Huawei MINDSPORE Academic Open Fund, the Key Research and Development Project of Shaanxi Province (2022QCY-LL-35), the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University.

References

- [1] 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. 3
- [2] Dominik Bauer, Timothy Patten, and Markus Vincze. Reagent: Point cloud registration using imitation and reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14586–14594, 2021. 3
- [3] Ben Bellekens, Vincent Spruyt, Rafael Berkvens, Rudi Penne, and Maarten Weyn. A benchmark survey of rigid 3d point cloud registration algorithms. In *Proceedings of International Conference on Ambient Computing, Applications, Services and Technologies*, pages 118–127, 2015. 2
- [4] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: Control Paradigms and Data Structures*, pages 586–606, 1992. 2, 6, 7, 8
- [5] Sofien Bouaziz, Andrea Tagliasacchi, and Mark Pauly. Sparse iterative closest point. In *Computer Graphics Forum*, pages 113–123, 2013. 2
- [6] Wentao Cheng, Weisi Lin, Xinfeng Zhang, Michael Goesele, and Ming-Ting Sun. A data-driven point cloud simplification framework for city-scale image-based localization. *IEEE Transactions on Image Processing*, 26(1):262–275, 2016. 1
- [7] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5556–5565, 2015. 2
- [8] 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. 3
- [9] Jakob Engel, Jörg Stückler, and Daniel Cremers. Large-scale direct slam with stereo cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1935–1942, 2015. 1
- [10] Wanquan Feng, Juyong Zhang, Hongrui Cai, Haofei Xu, Junhui Hou, and Hujun Bao. Recurrent multi-view alignment network for unsupervised surface registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10297–10307, 2021. 3
- [11] Nicola Fioraio and Kurt Konolige. Realtime visual and point cloud slam. In *Proceedings of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf.(RSS)*, 2011. 1
- [12] Kai Fischer, Martin Simon, Florian Olsner, Stefan Milz, Horst-Michael Gross, and Patrick Mader. Stickypillars: Robust and efficient feature matching on point clouds using graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 313–323, 2021. 3
- [13] 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. 6, 7, 8
- [14] Andrew W Fitzgibbon. Robust registration of 2d and 3d point sets. *Image and Vision Computing*, 21(13-14):1145–1153, 2003. 2
- [15] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. 1
- [16] 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. 2, 3, 6, 7, 8
- [17] Haobo Jiang, Yaqi Shen, Jin Xie, Jun Li, Jianjun Qian, and Jian Yang. Sampling network guided cross-entropy method for unsupervised point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6128–6137, 2021. 2, 3, 6, 7, 8
- [18] Pranav Kadam, Min Zhang, Shan Liu, and C-C Jay Kuo. Unsupervised point cloud registration via salient points analysis. In *2020 IEEE International Conference on Visual Communications and Image Processing*, pages 5–8, 2020. 3
- [19] Akiyoshi Kurobe, Yusuke Sekikawa, Kohta Ishikawa, and Hideo Saito. Corsnet: 3d point cloud registration by deep neural network. *IEEE Robotics and Automation Letters*, 5(3):3960–3966, 2020. 2
- [20] Huan Lei, Guang Jiang, and Long Quan. Fast descriptors and correspondence propagation for robust global point cloud registration. *IEEE Transactions on Image Processing*, 26(8):3614–3623, 2017. 1
- [21] Jiahao Li, Changhao Zhang, Ziyao Xu, Hangning Zhou, and Chi Zhang. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration. In *European Conference on Computer Vision*, pages 378–394. Springer, 2020. 2, 6, 7, 8
- [22] Simon Lucey, Rajitha Navarathna, Ahmed Bilal Ashraf, and Sridha Sridharan. Fourier lucas-kanade algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1383–1396, 2012. 3
- [23] Longfei Ma, Hanying Liang, Boxuan Han, Shizhong Yang, Xinran Zhang, and Hongen Liao. Augmented reality navigation with ultrasound-assisted point cloud registration for percutaneous ablation of liver tumors. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–10, 2022. 1
- [24] 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. 2
- [25] Gonzalo Mena, David Belanger, Scott Linderman, and Jasper Snoek. Learning latent permutations with gumbel-sinkhorn networks. *arXiv preprint arXiv:1802.08665*, 2018. 2, 3

- [26] Théodore Papadopoulos and Manolis IA Lourakis. Estimating the jacobian of the singular value decomposition: Theory and applications. In *European Conference on Computer Vision*, pages 554–570, 2000. 5
- [27] 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. 2, 3, 6
- [28] Szymon Rusinkiewicz. A symmetric objective function for icp. *ACM Transactions on Graphics*, 38(4):1–7, 2019. 2
- [29] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling*, pages 145–152, 2001. 2
- [30] Joaquim Salvi, Jordi Pages, and Joan Battle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004. 1
- [31] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: Science and Systems*, 2009. 2
- [32] 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*, pages 2198–2206, 2022. 2, 3, 4, 6, 7, 8
- [33] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3523–3532, 2019. 2, 3
- [34] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions On Graphics*, 38(5):1–12, 2019. 3
- [35] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015. 2
- [36] Hao Xu, Nianjin Ye, Shuaicheng Liu, Guanghui Liu, and Bing Zeng. Finet: Dual branches feature interaction for partial-to-partial point cloud registration. *arXiv preprint arXiv:2106.03479*, 2021. 6
- [37] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics*, 37(2):314–333, 2021. 3
- [38] Zi Jian Yew and Gim Hee Lee. Rpm-net: Robust point matching using learned features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11824–11833, 2020. 2, 3, 6, 7, 8
- [39] 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. 6
- [40] Wentao Yuan, Benjamin Eckart, Kihwan Kim, Varun Jampani, Dieter Fox, and Jan Kautz. Deepgmr: Learning latent gaussian mixture models for registration. In *European Conference on Computer Vision*, pages 733–750, 2020. 3
- [41] Yongzhe Yuan, Yue Wu, Xiaolong Fan, Maoguo Gong, Wenping Ma, and Qiguang Miao. Egst: Enhanced geometric structure transformer for point cloud registration. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–13, 2023. 2
- [42] Yongzhe Yuan, Yue Wu, Maoguo Gong, Qiguang Miao, and AK Qin. One-nearest neighborhood guides inlier estimation for unsupervised point cloud registration. *arXiv preprint arXiv:2307.14019*, 2023. 3
- [43] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1802–1811, 2017. 2
- [44] Yiming Zeng, Yue Qian, Zhiyu Zhu, Junhui Hou, Hui Yuan, and Ying He. Cornnet3d: Unsupervised end-to-end learning of dense correspondence for 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6052–6061, 2021. 3
- [45] Ji Zhang and Sanjiv Singh. Loam: Lidar odometry and mapping in real-time. In *Robotics: Science and Systems*, 2014. 1