

Unsupervised RGB-D Point Cloud Registration for Scenes with Low Overlap and Photometric Inconsistency

Yejun Shou^{1,2}, Haocheng Wang^{1,2}, Lingfeng Shen^{1,2}, Qian Zheng^{3*}, Gang Pan³, Yanlong Cao^{1,2*}

¹The State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, China

²The Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, China

³College of Computer Science and Technology, Zhejiang University, China

{syj12125092, 00whc, lfs, qianzheng, gpan, sdcaoyl}@zju.edu.cn

Abstract

Point cloud registration is a fundamental task in 3D vision, playing a crucial role in various fields. With the rapid advancement of RGB-D sensors, unsupervised point cloud registration methods based on RGB-D sequences have demonstrated excellent performance. However, existing methods struggle in scenes with low overlap and photometric inconsistency. Low overlap results in numerous correspondence outliers, while photometric inconsistency hinders the model's ability to extract discriminative features. To address these challenges, we first propose the **Overlapping Constraint for Inliers Detection (OCID)** module, which filters and optimizes the initial correspondence set using an overlapping constraint. This module robustly selects reliable correspondences within the overlapping region while maintaining a balance between accuracy and efficiency. Additionally, we introduce a novel scene representation, **3DGS**, which integrates both geometric and texture information, making it particularly well-suited for RGB-D registration tasks. Building on this, we propose the **Gaussian Rendering for Photometric Adaptation (GRPA)** module, which refines the geometric transformation and enhances the model's adaptability to scenes with inconsistent photometric information. Extensive experiments on ScanNet and ScanNet1500 demonstrate that our method achieves state-of-the-art performance. The code will be released at [OG-UPCR](#).

1. Introduction

Point cloud registration is a critical task in 3D vision, with applications in navigation [30, 44, 56], pose estimation [10, 17], and 3D reconstruction [42, 43]. Typically, two frames of point clouds rely on feature extraction to estimate correspondences, followed by robust geometric fitting techniques to obtain the transformation.

*Corresponding author

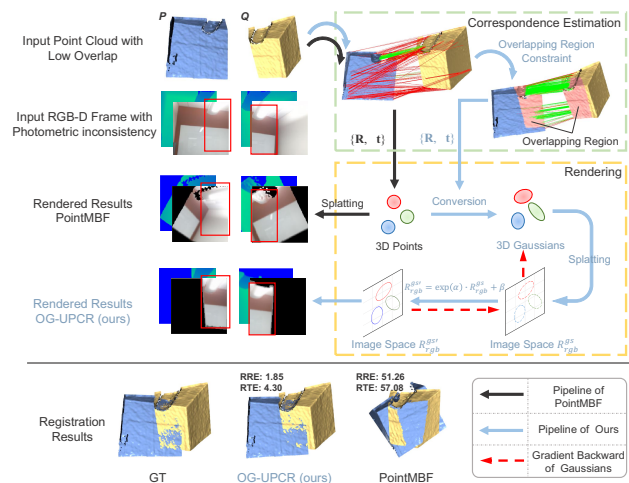


Figure 1. The objective of the proposed OG-UPCR is to effectively register scenes with low overlap and photometric inconsistency. The overlapping region constraint helps reduce correspondence outliers caused by low overlap, while photometric inconsistency is adaptively adjusted through the optimizable parameters contained in 3D Gaussians [24]. Compared to PointMBF [52], the proposed OG-UPCR produces more consistent results between input and rendered frames (highlighted in red boxes) and yields better registration outcomes.

With the rapid development of RGB-D sensors, unsupervised registration methods [12–14, 27, 40, 45, 52] have demonstrated impressive performance. These methods rasterize 3D points into images using a differentiable renderer [32] and employ geometric and photometric consistency losses between the input and rendered frames to extract features, thereby estimating the geometric transformation between nearby frames without the need for ground truth. However, all these methods ignore that the geometric and photometric information between frames to be registered is not always consistent, particularly in scenes with low overlap and photometric inconsistency. Low overlap

complicates feature identification and correspondence estimation [20], while photometric inconsistency hinders the modeling of feature consistency between input and rendered frames, making it challenging for the model to obtain discriminative features, especially in scenes with reflections and rendering artifacts [52]. Despite the excellent registration performance of PointMBF [52], it still suffers from low overlap and photometric inconsistency, which result in a large number of correspondence outliers and erroneous rendered and registration results, as illustrated in Fig. 1.

We observe that inliers must be located within overlapping regions, whereas outliers can be filtered out early using overlapping constraint. Although identifying inliers and overlaps is a chicken-and-egg situation, it can be efficiently resolved through an iterative process. Meanwhile, recent advances in 3D Gaussian Splatting [24] have made significant strides in novel view synthesis [4, 54], 3D generation [6, 50], and SLAM [23, 26, 46]. This technique not only represents geometry in a smooth and continuously differentiable manner but also preserves view-varying texture information, which is adaptive to photometric changes. Additionally, 3D points convert easily into 3D Gaussians, making this approach highly suitable for RGB-D registration.

To this end, we propose OG-UPCR, an innovative unsupervised point cloud registration framework designed to address low overlap and photometric inconsistency, as shown in Fig. 1. Specifically, we propose the **O**verlapping **C**onstraint for **I**nliers **D**etection module (OCID) based on the observation that inliers must lie within overlapping regions. This module leverages overlapping constraints to detect inliers, addressing issues arising from low overlap. Additionally, we design the **G**aussian **R**endering for **P**hotometric **A**daptation module (GRPA), which converts 3D points into 3D Gaussians and rasterizes them into image space, leveraging the rich geometric and texture information contained in the Gaussians. Based on this, nonlinear optimization is applied to each pixel to enhance photometric adaptation and enable the model to extract discriminative features. Moreover, this process refines the transformation and improves overall registration.

The OG-UPCR is thoroughly evaluated on the standard datasets, including 3DMatch [53], ScanNet [9], and the challenging ScanNet1500 [35], which contain paired views with low overlap and inconsistent photometric information, achieving state-of-the-art performance. Additionally, we conduct extensive ablation experiments on individual modules to demonstrate the effectiveness of each component.

In summary, our contributions are as follows:

- We propose OG-UPCR, an unsupervised registration framework that combines overlapping constraint with gaussian splatting, enhancing the accuracy and robustness of the overall registration model.
- We propose a correspondence optimization strategy us-

ing iterative random overlapping region prediction. This module selects faithful inliers of correspondences within the overlapping region, mitigating the impact of low overlap and enhancing the accuracy of geometric fitting.

- We propose a photometric adaptation module for paired views that employs Gaussian splatting to independently render the Gaussians of both views. This module models the photometric consistency of features across frames, and refines the final transformation.

2. Related Work

2.1. Point Cloud Registration

Early point cloud registration methods [21, 33, 34, 39] rely on hand-crafted feature descriptors, which are sensitive to noise and not robust to changes in the number and density of point clouds. Recently, deep learning feature descriptors have demonstrated impressive performance and can achieve better registration results. Research based on deep learning can be divided into supervised [2, 20, 29, 37, 41, 51] and unsupervised [12, 13, 28, 36, 45, 52] methods. However, supervised methods require real poses as supervision signals, making them difficult to apply practically. When encountering new scenes with domain gaps, they can only rely on the trained weights to infer the registration results, which often affects their performance. Unsupervised methods can be further divided into geometric-based and RGB-D based methods according to the input data. Among geometric-based methods, RIENet [36] proposes a reliable inlier evaluation method based on neighborhood consensus. UDReg [28] operates by learning the posterior probability distribution of Gaussian Mixture Models from point clouds and then predicting distribution-level correspondences. Since the input contains only geometric information, these methods struggle with geometric weak regions (such as floors, ceilings, and walls) and repeated structures in real scenes. RGB-D based methods are able to handle real scenes, such as LLT [45] employing a linear attention module to fuse geometric and texture features. PointMBF [52] proposes a multi-scale bidirectional fusion network to fully utilize the complementary information of RGB-D. Nevertheless, these methods all employ a conventional differentiable renderer to render 3D points directly, making them susceptible to low overlap and photometric inconsistency. In contrast, we propose the OCID module based on overlapping constraint to help obtain more reliable correspondences. Additionally, we design the GRPA module to reduce the impact of photometric inconsistency, thereby enhancing registration accuracy.

2.2. Correspondence Optimization Strategy

The handling of correspondence outliers can be effectively addressed using methods such as RANSAC. Although RANSAC [15] and its variants [3, 31] are widely

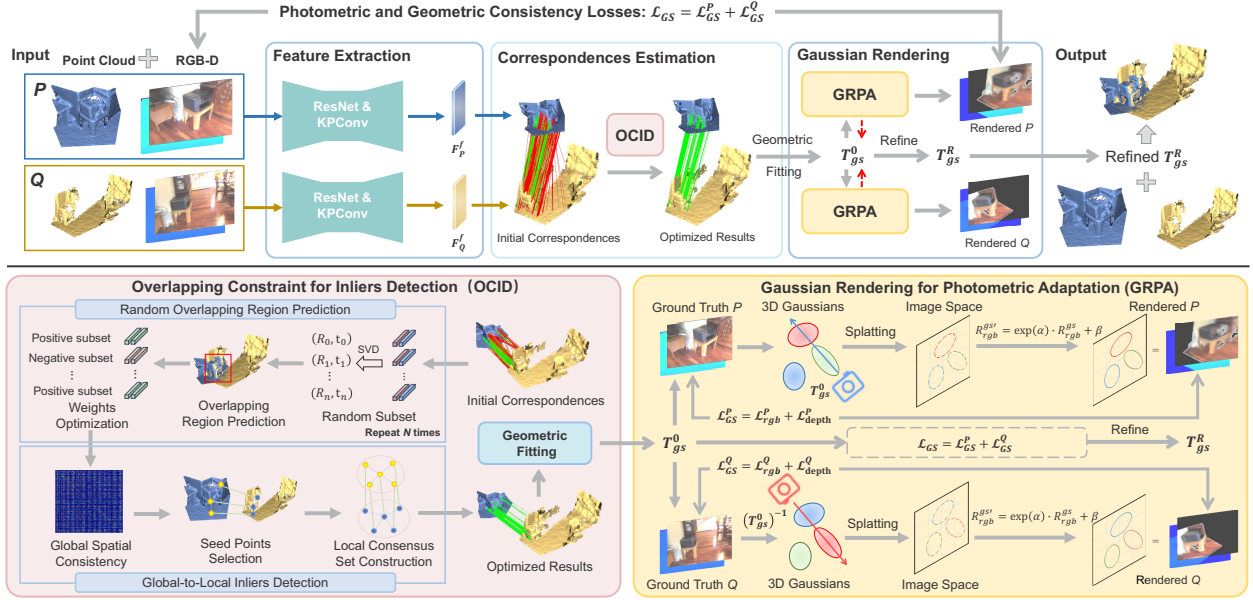


Figure 2. The overview of OG-UPCR. The backbone first extracts the features F_P^f and F_Q^f from the point clouds P and Q . The initial correspondences are estimated by calculating Lowe’s ratio in the feature space, which is then optimized by the OCID module using the overlapping constraint. Subsequently, the rigid transformation is determined through geometric fitting. Finally, based on this transformation, the Gaussian cross-rendering results generated by the GRPA module from the paired views are used to calculate the loss. This process not only refines the final result but also aids the model in acquiring distinct features.

recognized algorithms, their performance remains sensitive to outliers, often requiring numerous iterations to produce relatively accurate results. SC2PCR [5] and MAC [55] are also effective in managing correspondence outliers, but both rely on relative spatial constraint and fail to account for the fact that correct correspondences should be confined to the overlapping region from global spatial constraint. This oversight leads to inefficiency and inaccuracy, as excessive redundant point pairs are utilized. In contrast, we propose an overlapping constraint module based on global spatial to select reliable correspondences, thereby improving the efficiency and accuracy of the registration process.

2.3. Gaussian Splatting

As a novel scene representation and efficient differentiable renderer, 3D Gaussian Splatting [24] has shown its significant advantages in novel view synthesis [4, 48, 54], 3D generation [6, 50] and SLAM [23, 26, 46]. SplatAM [23] introduces an online tracking and mapping system based on the underlying 3D Gaussian representation, without requiring the real pose. However, it relies on dense RGB-D frame inputs, making it difficult to obtain accurate poses with sparse input. PixelSplat [6] proposes that the radiance field parameterized by 3D Gaussian primitives can be learned and reconstructed from sparse views. NoPoSplat [48] integrates PnP [18] and 3DGS [24] to estimate poses from sparse views, achieving superior novel

view synthesis. Nevertheless, these methods estimate poses from an image-based perspective, resulting in the loss of 3D information. In contrast, the proposed GRPA module derives a more precise geometric transformation from the point cloud perspective, producing rasterized outputs independent of true pose data. This approach exhibits robust generalization and achieves accurate registration even with sparse inputs.

3. Method

3.1. Problem Statement and Overview

Problem Statement. Given point clouds $P \in \mathbb{R}^{N \times 3}$ and $Q \in \mathbb{R}^{N \times 3}$, the purpose of the registration task is to obtain a transformation matrix $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$, that aligns P and Q in the same coordinate system. This transformation matrix consists of a rotation matrix $\mathbf{R} \in \mathbf{SO}(3)$ and a translation vector $\mathbf{t} \in \mathbb{R}^3$. The correspondence set $C = \{\mathbf{p}_i, \mathbf{q}_j | \mathbf{p}_i \in P, \mathbf{q}_j \in Q\}$ is estimated by feature extraction and matching. The transformation can be determined by:

$$\arg \min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^n \|\mathbf{R} \cdot \mathbf{p}_i + \mathbf{t} - \mathbf{q}_j\|_2^2, \quad (1)$$

where $\|\cdot\|_2^2$ denotes the Euclidean norm, n is the number of correspondences that can be found in paired point clouds.

Overview. In this paper, we propose an OCID module and a GRPA module to address challenges associated with low overlap and photometric inconsistency. The OG-UPCR processes paired point clouds and their corresponding RGB-D frames as input, following the PointMBF [52] to extract feature descriptors. Using these descriptors, we compute an initial correspondence set by measuring distances in the feature space. The OCID module then filters this initial set based on an overlapping region constraint, selecting reliable seed points for geometric fitting, which produces a coarse transformation, denoted as \mathbf{T}_{gs}^0 , as described in Section 3.2. This transformation serves as the initial pose for the GRPA module, where 3D points are first converted into 3D Gaussians and then projected onto an image using a Gaussian splatting renderer, which is subsequently refined to obtain the final result, \mathbf{T}_{gs}^R , as detailed in Section 3.3. Additionally, the entire registration framework is trained in an end-to-end unsupervised manner using geometric and photometric consistency losses between input and rendered frames, as discussed in UR&R [13]. An overview of OG-UPCR is provided in Fig. 2.

Feature Extraction. Building upon prior research, we employ KPConv [38] and ResNet [19] to extract multi-scale information from the point clouds and images. Subsequently, a multi-scale bidirectional fusion method [52] generates distinctive fused feature descriptors $F_P^f \in \mathbb{R}^{N \times C_{3D}}$ and $F_Q^f \in \mathbb{R}^{N \times C_{3D}}$ for point clouds P and Q .

3.2. Correspondence Estimation

Initial Correspondence Set Acquisition. After obtaining the feature descriptors of the point clouds, we follow previous method [13, 45, 52] to obtain the initial correspondence set based on Lowe’s ratio [25]. For a point \mathbf{p}_i , its Lowe’s ratio can be computed as:

$$r_i = \frac{D(\mathbf{p}_i, \mathbf{q}_j^{1nn})}{D(\mathbf{p}_i, \mathbf{q}_j^{2nn})}, \quad (2)$$

where $D(\cdot)$ represents the Euclidean distance in the feature space and q_j^{knn} is the k -th similarity point in the point cloud Q . We assign weights to each correspondence using the formula $w = 1 - r$. Next, we apply the top- k algorithm to select the weighted correspondences, resulting in a correspondence set from $C_{P \rightarrow Q}$. We choose an equal number of correspondences from $C_{Q \rightarrow P}$ in a similar manner to create the initial correspondence set:

$$C_{\text{initial}} = \{(\mathbf{p}, \mathbf{q}, w)_m : 0 \leq m < 2k\}, \quad (3)$$

where k denotes the top k weights. The resulting correspondence C_{initial} will be input into the OCID module for further optimization.

Overlapping Constraint for Inliers Detection. This module is designed to detect inliers from the initial correspondence set, facilitating efficient registration in scenarios with

low overlap. Existing works [5, 55] directly remove outliers based on the consistency graph but neglect the importance of the overlapping region constraint, leading to significant computational expenses. As shown in Fig. 2, the module consists of two components, which are detailed below.

Random Overlapping Region Prediction. As is well known, Weighted SVD [8] produces an approximate geometric transformation by performing singular value decomposition on the weights of the correspondence set. The accuracy of this transformation will be affected by the inliers in the correspondence, meaning that the higher the weight of the inliers, the more accurate the transformation will be. Motivated by this and inspired by RANSAC [15], we first randomly select a partial subset from the initial correspondence:

$$C_n = \{(\mathbf{p}, \mathbf{q}, w)_n : 0 \leq n < 2rk\} \in C_{\text{initial}}, \quad (4)$$

where r denotes the proportion of the subset in the initial correspondence, which is set to 0.2. Next, we apply Weighted SVD to obtain the transformation \mathbf{T}_s of this subset. Unlike similar previous work [13, 45, 52], we do not directly leverage the common L2 distance to assess whether \mathbf{T}_s is correct, which is prone to fall into local optima. Instead, we propose a novel approach based on overlapping region prediction. Using \mathbf{T}_s , we predict the overlapping area of the point cloud and evaluate its success by the overlap ratio O_p , which is defined as:

$$P' = \mathbf{T}_s \odot P, \quad (5)$$

$$O_p = \frac{1}{N} \sum_{n=1}^{N_{P'}} \mathbb{I}[\|\mathbf{p}'_n - \mathbf{q}_n\|_2 < \tau], \quad (6)$$

where $\mathbb{I}[\cdot]$ is the Iverson bracket, \odot represents the rigid transformation, $N_{P'}$ is the number of points of the point cloud P' and τ is set to 5cm. If O_p exceeds the threshold O_{thr} , the prediction is considered successful. This implies that most of the correspondences within the selected subset are inliers, and the subset is classified as a positive subset, receiving a higher weight in the initial correspondence set. To meet these conditions, we multiply the weight of the points in the subset of successful predictions by a reward factor, thereby increasing their proportion in the initial correspondence. Conversely, if the prediction fails, it indicates that the selected subset contains a high percentage of outliers. In this case, the corresponding weights in those subsets are multiplied by a penalty factor, which is defined as follows:

$$w_n = \begin{cases} w_n \cdot w_r, & O_p \geq O_{thr} \\ w_n \cdot w_p, & O_p < O_{thr} \end{cases}, \quad (7)$$

where w_r and w_p denote the reward factor and penalty factor, and $w_r \cdot w_p = 1$ to ensure the balance change of a correspondence when it is selected into the subset many times.

The above prediction process is repeated N times. After completing the random overlapping region prediction, the weights of the individual correspondences within the initial correspondence set are significantly optimized, with inliers receiving higher weights and the weights of outliers being reduced. To enhance the efficiency of the algorithm, we also use the top-k algorithm to select the appropriate correspondence. The optimized set of correspondences can be defined as follows:

$$C_f = \left\{ (\mathbf{p}, \mathbf{q}, w)_f : 0 \leq m < 2k_f \right\}, \quad (8)$$

where k_f means that the top-k algorithm is used to select items with larger weights to speed up the algorithm. We also provide a pseudocode and simple flowchart in Supplementary Material. 2.

Global-to-Local Inliers Detection. The optimized correspondence set C_f already exhibits a high inlier ratio. However, one issue remains: the selected correspondences are concentrated in specific local regions rather than uniformly distributed. This concentration hinders the subsequent geometric fitting. To address this challenge, we propose a global-to-local inlier detection module based on spatial consistency, as shown in Fig. 2. First, we compute the second order spatial compatibility (SC^2) [5] of the optimized correspondence set globally, which can be defined as:

$$SC_{ij}^2 = C_{ij} \cdot \sum_{k=1}^N C_{ik} \cdot C_{kj}, \quad (9)$$

$$C_{ij} = \begin{cases} 1, & d_{ij} \leq d_{thr} \\ 0, & d_{ij} > d_{thr} \end{cases}, \quad (10)$$

$$d_{ij} = \left| \|\mathbf{p}_i - \mathbf{p}_j\|_2 - \|\mathbf{q}_i - \mathbf{q}_j\|_2 \right|, \quad (11)$$

where $\mathbf{p}_i, \mathbf{p}_j$ and $\mathbf{q}_i, \mathbf{q}_j$ denote two sets of correspondence points in C_f , and d_{thr} is set to 10cm. Subsequently, the optimized weights w_f are employed to find the seed points with the maximum confidence score within their neighborhood, defined by a radius R , ensuring that the correspondences are uniformly distributed. To ensure that the selected seed points are reliable inliers, we measure the correspondence of the seed point and its top- K neighbors based on the constructed SC^2 metric to further prune potential outliers. Finally, these seed points and their neighboring points form multiple consensus sets.

Geometric Fitting. To project images of the transformed point clouds P and Q using a Gaussian renderer, it is necessary to estimate their geometric transformation. This estimation is achieved through geometric fitting. Following [5], we estimate the geometric transformation according to each consensus set. Finally, we choose the transformation with the highest number of inliers as the result \mathbf{T}_{gs}^0 .

3.3. Gaussian Rendering

This module is designed to help the model adapt to inconsistent photometric information between the input and rendered frames. Previous approaches [13, 45, 52] rely on traditional differentiable renderer [32] to directly project 3D points into the image space, which makes their models susceptible to significant viewpoint changes and limits the ability to extract more discriminative features. To overcome these challenges, we draw inspiration from [22, 26] and introduce the 3D Gaussian splatting technique. Additionally, we propose a photometric adaptation mechanism for each pixel to resolve the inconsistencies caused by viewpoint changes, thereby enabling the model to learn more discriminative features for registration.

Gaussian Rendering for Photometric Adaptation. Specifically, each point can be represented by a Gaussian [22] with eight parameter values: its point position $\mu_{3D} \in \mathbb{R}^3$, RGB color $\mathbf{c} \in \mathbb{R}^3$, radius r , and opacity $o \in [0, 1]$. Each Gaussian affects a point in 3D space $\mathbf{x} \in \mathbb{R}^3$ as follows:

$$g(\mathbf{x}) = o \exp\left(-\frac{\|\mathbf{x} - \mu_{3D}\|^2}{2r_{3D}^2}\right), \quad (12)$$

each pixel ρ in the differentiable rendered RGB image can then be rendered as follows:

$$R_{rgb}^{gs}(\mathbb{G}, \mathbf{T}) = \sum_{i=1}^n c_i g_i(\rho) \prod_{j=1}^{i-1} (1 - g_j(\rho)), \quad (13)$$

where $g(\rho_a)$ can be obtained from Eq. (12) and (13), but with μ_{3D} and r_{3D} of the splatted 2D Gaussians in pixel space:

$$\mu_{2D} = \mathbf{K}\mathbf{T} \frac{\mu_{3D}}{d}, r_{2D} = \frac{f}{d}, \quad (14)$$

where \mathbf{K} is the camera intrinsic matrix, \mathbf{T} is the extrinsic matrix, f is the focal length of the camera, and d is the depth. To further enhance the adaptability to photometric changes, we optimize the nonlinear photometric adaptation of each pixel:

$$R_{rgb}^{gs'} = \exp(\alpha) \cdot R_{rgb}^{gs} + \beta, \quad (15)$$

where α and β are both learnable parameters. A similar differentiable rendered depth image can be expressed as follows:

$$R_{depth}^{gs}(\mathbb{G}, \mathbf{T}) = \sum_{i=1}^n d_i g_i(\rho_a) \prod_{j=1}^{i-1} (1 - g_j(\rho_a)). \quad (16)$$

This module refines transformations by rendering and comparing them to minimize the discrepancy between the rendered object and the actual observations in the query image. The loss function can be defined as:

$$L_{gs}^P = L_{rgb}^P + L_{depth}^P, \quad (17)$$

$$L_{rgb}^P = \left\| R_{rgb}^{gs'} (\mathbb{G}^Q, \mathbf{T}_{gs}^0) - I^P \right\|_1, \quad (18)$$

$$L_{depth}^P = \left\| R_{depth}^{gs} (\mathbb{G}^Q, \mathbf{T}_{gs}^0) - D^P \right\|_1, \quad (19)$$

where \mathbb{G}^Q denote Gaussians associated with the point cloud Q , I^P and D^P correspond to RGB image and depth image. L_{gs}^Q can be obtained in the same way. For the point cloud Q , we set the extrinsic matrix corresponding to its Gaussian as \mathbf{T}_{gs}^0 , the rendering viewpoint of the camera as \mathbf{T}_0 (\mathbf{T}_0 is the identity matrix), and the opposite for the point cloud P . The input initial transformation \mathbf{T}_{gs}^0 can be refined into the final result \mathbf{T}_{gs}^R output after GRPA module.

3.4. Loss function

As in previous works [13, 45, 52], the loss function consists of photometric, depth, and correspondence consistency losses without the need for ground truth. It can be defined as follows:

$$L = \lambda_{gs} L_{gs} + \lambda_{corr} L_{corr}, \quad (20)$$

where L_{gs} represents the photometric and depth consistencies loss as defined in Eq. (17), Eq. (18) and Eq. (19), λ_{gs} and λ_{corr} control the magnitude of these losses ensuring they are within an appropriate order of magnitude. L_{corr} represents the correspondence consistency loss of C_f , which can be donated as:

$$L_{corr} = \sum_{(p,q,w)_f \in C_f} w_f (T_{gs}^R \odot p_f - q_f)^2 / 2k_f, \quad (21)$$

where \odot represents the rigid transformation.

4. Experiments

4.1. Experimental Settings

Dataset. Following [52], OG-UPCR has been evaluated on several indoor real-world scene datasets, including 3DMatch [53], ScanNet [9] and ScanNet1500 [35]. Among them, 3DMatch and ScanNet generate view pairs by sampling an interval of 20 frames, resulting in 3DMatch containing 122k/1.5k/1.5k view pairs and ScanNet containing 1594k/12.6k/26k view pairs for training/validation/test, respectively. ScanNet1500 contains 1500 view pairs with an average interval of 480.8 frames between views, which makes it more challenging for the registration task.

Metrics. Following the convention [12, 13, 45, 52], we evaluate OG-UPCR by three evaluation metrics: rotation error, translation error, and chamfering error. In analysis, we report the mean and median values of these metrics and assess their accuracy across various thresholds.

Competitors. We compare OG-UPCR with other methods, which can be primarily divided into three categories: (1) Conventional methods, including SIFT [25], SuperPoint [11] and FCGF [7]. (2) Supervised methods, including 3DMVReg [16], REGTR [49], BUFFER [1], and

PARENet [47]. (3) Unsupervised methods, including UR&R [13], BYOC [12], LLT [45], PointMBF [52], and CCAP [52]. For the supervised methods, we verify their generalization on ScanNet using the officially provided weights trained on 3DMatch.

4.2. Evaluation on ScanNet

We first evaluate our OG-UPCR on the widely used ScanNet, with training weights from 3DMatch and ScanNet, respectively. The former validates the generalization performance comparison with other methods, and the latter evaluates the conventional cases.

Trained on 3DMatch. As shown in Table 1, the generalization experimental results trained on 3DMatch demonstrate that our method achieves the state-of-the-art registration results compared with all other methods, both unsupervised and supervised. This indicates that our method has excellent adaptability to unseen scenes, balancing both registration accuracy and generalization.

Trained on ScanNet. When there is no domain gap between the training set and the test set, our method also achieves the best registration results, as shown in Table 1. Even when the baseline is almost saturated, our method still shows significant accuracy improvements across each evaluation metric, with average reductions in rotation/translation/chamfer distance errors of 40.0%/34.5%/41.3%, respectively. This demonstrates that our proposed method also delivers excellent registration performance in conventional cases.

4.3. Evaluation on ScanNet1500

Following previous work [52], we conduct experiments on the challenging ScanNet1500 dataset to further demonstrate the effectiveness of our method. As shown in Table 2, our method outperforms all other unsupervised and supervised methods on the most stringent accuracy metric, and performs only slightly worse than the supervised method with the best generalization performance on the most permissive accuracy metric. The ScanNet1500 includes view pairs with low overlap and significant viewpoint variations, which pose substantial challenges for existing unsupervised methods. These conditions hinder the performance of previous methods, as geometric and photometric consistency between view pairs is greatly affected. In contrast, our method maintains strong performance by identifying reliable correspondences through overlapping constraint and further optimizing the model using the Gaussian Splatting technique.

4.4. Ablation Studies

Analysis of Each Component. We conduct a comprehensive ablation study for each component. The OCID module comprises two key components: **R**andom **O**verlapping region prediction (RO) and **G**lobal-to-**L**ocal inliers detection

Table 1. Pairwise Registration accuracies and errors on ScanNet. Pose Sup means training with ground truth. The best result is shown in bold and the second best is underlined.

Method	Train set	Pose Sup	Rotation(°)					Translation(cm)					Chamfer(mm)					Time(s)
			Accuracy		Error			Accuracy		Error			Accuracy		Error			
			5	10	45	Mean	Med	5	10	25	Mean	Med	1	5	10	Mean	Med	
SIFT	-		55.2	75.7	89.2	18.6	44.3	17.7	44.5	79.8	26.5	11.2	38.1	70.6	78.3	42.6	1.7	-
SuperPoint	-		65.5	86.9	96.6	8.9	3.6	21.2	51.7	88.0	16.1	9.7	45.7	81.1	88.2	19.2	1.2	-
FCGF	-	✓	70.2	87.7	96.2	9.5	3.3	27.5	58.3	82.9	23.6	8.3	52.0	78.0	83.7	24.4	0.9	0.129
3DMVReg	3DMatch	✓	87.7	93.2	97.0	6.0	1.2	69.0	83.1	91.8	11.7	2.9	78.9	89.2	91.8	10.2	0.2	-
REGTR	3DMatch	✓	86.0	93.9	98.6	4.4	1.6	61.4	80.3	91.4	14.4	3.8	80.9	90.9	93.6	13.5	0.2	-
BUFFER	3DMatch	✓	95.9	98.4	99.2	2.4	0.9	79.6	92.3	96.9	5.7	2.4	93.6	97.3	98.6	4.2	0.2	0.177
PARENet	3DMatch	✓	97.5	98.3	99.5	1.9	0.7	85.9	95.7	97.0	3.9	1.8	94.2	97.9	98.6	4.2	0.2	0.165
UR&R	3DMatch	✓	92.3	95.3	98.2	3.8	0.8	77.6	89.4	95.5	7.8	2.3	86.1	94.0	95.6	6.7	0.1	0.070
UR&R	3DMatch		87.6	93.1	98.3	4.3	1.0	69.2	84.0	93.8	9.5	2.8	79.7	91.3	94.0	7.2	0.2	0.071
BYOC	3DMatch		66.5	85.2	97.8	7.4	3.3	30.7	57.6	88.9	16.0	8.2	54.1	82.8	89.5	9.5	0.9	-
LLT	3DMatch		93.4	96.5	98.8	2.5	0.8	76.9	90.2	96.7	5.5	2.2	86.4	95.1	96.8	4.6	0.1	0.125
PointMBF	3DMatch		94.6	97.0	98.7	3.0	0.8	81.0	92.0	97.1	6.2	2.1	91.3	96.6	97.4	4.9	0.1	0.079
CCAP	3DMatch		95.6	98.1	-	2.4	0.7	81.5	92.3	-	3.7	1.9	-	-	-	-	-	-
Ours	3DMatch		97.9	98.9	99.5	1.8	0.7	87.7	96.2	98.6	3.8	1.8	95.5	98.4	98.8	3.3	0.1	0.136
UR&R	ScanNet		92.7	95.8	98.5	3.4	0.8	77.2	89.6	96.1	7.3	2.3	86.0	94.6	96.1	5.9	0.1	0.071
BYOC	ScanNet		86.5	95.2	99.1	3.8	1.7	56.4	80.6	96.3	8.7	4.3	78.1	93.9	96.4	5.6	0.3	-
LLT	ScanNet		95.5	97.6	99.1	2.5	0.8	80.4	92.2	97.6	5.5	2.2	88.9	96.4	97.6	4.6	0.1	0.125
PointMBF	ScanNet		96.0	97.6	98.9	2.5	0.7	83.9	93.8	97.7	5.6	1.9	92.8	97.3	97.9	4.7	0.1	0.079
CCAP	ScanNet		97.1	98.2	-	1.9	0.6	85.9	93.6	-	3.9	1.8	-	-	-	-	-	-
Ours	ScanNet		98.3	99.2	99.7	1.5	0.7	88.8	96.7	99.0	3.3	1.7	96.1	98.8	99.1	2.8	0.1	0.136

Table 2. Pairwise Registration accuracies and errors on ScanNet1500. The performance of our method is comparable to the supervised BUFFER method and significantly surpasses other unsupervised methods.

Method	Train set	Pose Sup	Rotation(°)				Translation(cm)				Chamfer(mm)			
			Accuracy		Error		Accuracy		Error		Accuracy		Error	
			5	10	45	Med	5	10	25	Med	1	5	10	Med
LLT	3DMatch		36.1	49.9	79.9	10.0	17.2	29.9	46.6	29.6	23.5	40.4	48.3	12.1
PointMBF	3DMatch		51.1	60.8	82.9	4.7	31.5	44.2	59.3	13.8	44.8	57.5	63.3	1.7
BUFFER	3DMatch	✓	74.9	82.5	90.9	1.8	47.4	66.5	81.0	5.0	66.5	79.5	84.2	0.3
PARENet	3DMatch	✓	65.6	71.9	88.6	3.1	50.1	71.7	83.3	4.7	61.7	77.9	82.7	0.3
Ours	3DMatch		76.9	83.4	89.3	1.6	49.5	68.4	80.9	5.1	67.9	80.2	83.4	0.3
LLT	ScanNet		43.7	54.6	80.3	7.5	24.0	36.1	52.9	20.6	31.7	47.8	53.4	6.6
PointMBF	ScanNet		55.9	65.9	84.9	3.4	37.1	49.9	63.1	10.0	48.5	62.1	67.2	1.2
Ours	ScanNet		81.3	85.9	91.2	1.5	53.5	72.1	83.3	4.3	72.4	83.3	85.9	0.3

(GL). For the GRPA module, we use a conventional differentiable renderer [32] instead. As shown in Table 3, the results reveal the following: 1) The baseline performs worst on both datasets when none of the proposed components are used. 2) Adding the GRPA module improves registration performance on both datasets, indicating that the GRPA refines transformation. 3) The complete OCID module significantly improves registration, with a notable increase in performance on ScanNet1500: 22.2% in rotation accuracy ($< 5^\circ$), 14.6% in translation accuracy (< 5 cm), and 20.4% in chamfering accuracy (< 1 mm). This shows the OCID module’s effectiveness in low-overlap scenes. 4) Additionally, using RO and GL separately results in improved registration performance for the entire framework. While the performance improvements on ScanNet are similar for both methods, RO achieves greater performance gains on ScanNet1500, demonstrating that the inclusion of overlapping constraint is particularly effective in scenarios with low overlap. 5) Finally, the complete framework with all com-

ponents shows the best performance on both ScanNet and ScanNet1500, highlighting its strong adaptability.

Table 3. Ablation studies on OG-UPCR. The model is trained on ScanNet. The best result is shown in bold.

OCID		GRPA	Test set	Rotation(°)		Translation(cm)		Chamfer(mm)	
RO	GL			Accuracy < 5	Error Med	Accuracy < 5	Error Med	Accuracy < 1	Error Med
×	×	×	ScanNet	96.0	0.8	83.9	1.9	92.8	0.1
×	×	✓		97.4	0.8	84.9	1.9	94.6	0.1
✓	×	×		97.7	0.7	85.7	1.8	95.1	0.1
×	✓	×		97.5	0.8	85.5	1.9	95.0	0.1
✓	×	×		97.2	0.8	84.7	1.9	93.9	0.1
✓	✓	✓		98.3	0.7	88.8	1.7	96.1	0.1
×	×	×	ScanNet1500	56.3	3.4	36.5	10.0	48.5	1.2
×	×	✓		65.1	2.1	42.3	6.8	57.1	0.5
✓	×	×		78.5	1.6	51.1	4.9	68.9	0.3
×	✓	×		73.9	1.9	46.1	5.7	65.5	0.4
✓	×	×		74.7	1.6	46.7	4.8	66.4	0.3
✓	✓	✓		81.3	1.5	53.5	4.3	72.4	0.3

Validation of the OCID for Inliers Detection. To evaluate the effect of the OCID module on the detection of inliers,

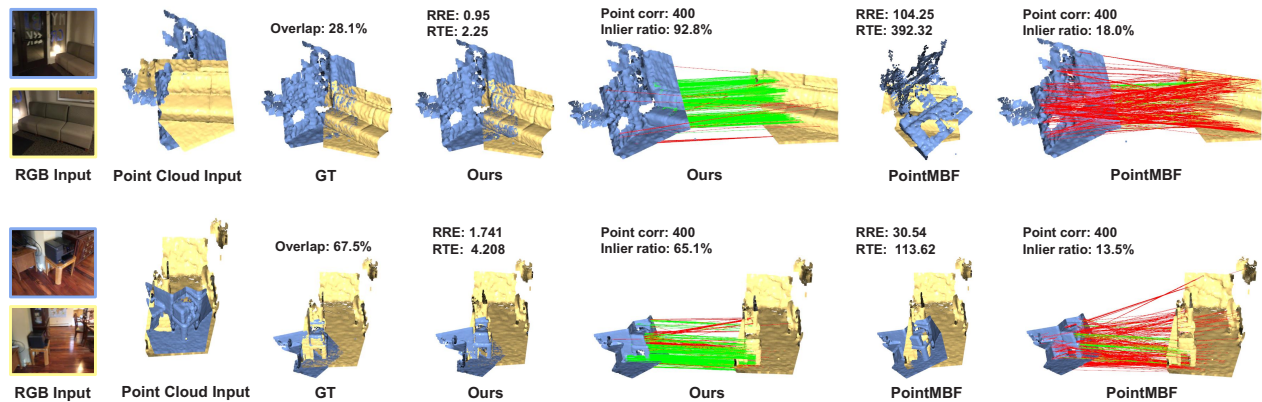


Figure 3. The visualization of point cloud registration results for PointMBF and the proposed OG-UPCR is presented. The results in the first row depict registration case with low overlap, while the results in the second row illustrate registration case with inconsistent photometric information caused by significant viewpoint variation.

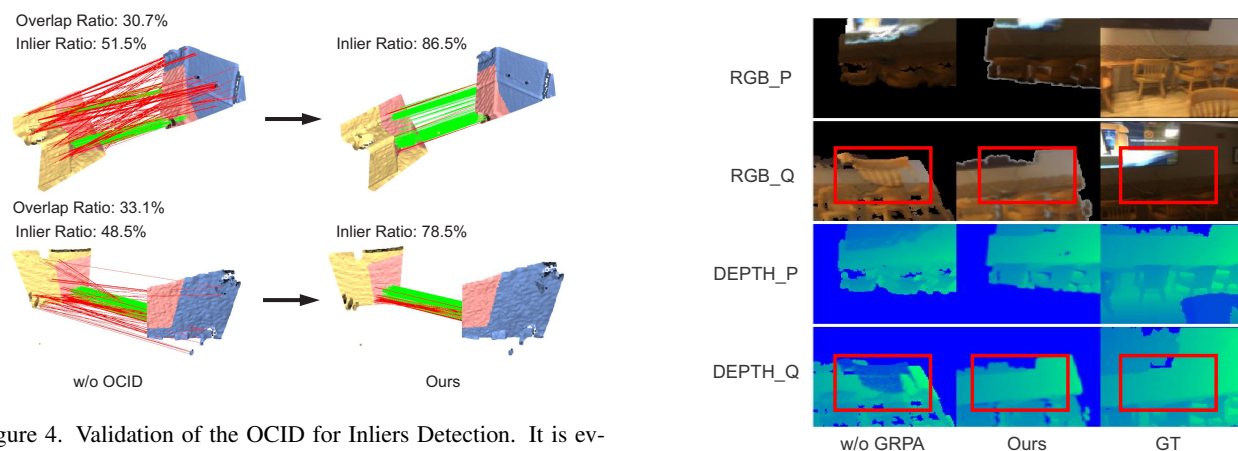


Figure 4. Validation of the OCID for Inliers Detection. It is evident that after incorporating the OCID, the correspondences are constrained to the overlapping regions (highlighted in pink), which simplifies geometric fitting and improves registration accuracy.

we visually compared the results before and after incorporating it. As shown in Fig. 4, after integrating the OCID module, the proportion of inliers is significantly increased in scenarios with low overlap, which facilitates subsequent geometric fitting and enhances the overall performance of the registration model.

Validation of the GRPA for Photometric Adaptation. To explore the impact of the GRPA, we replaced it with a traditional differentiable renderer and compared the rendered results. As shown in Fig. 5, the results demonstrate that incorporating the GRPA significantly reduces artifacts and enhances inter-frame consistency. This improvement is essential for ensuring consistent feature descriptors, as further discussed in Supplementary Material 4.4.

Qualitative Results. The qualitative results, compared with PointMBF, are shown in Fig. 3. OG-UPCR achieves superior registration results and identifies more accurate cor-

Figure 5. Validation of the GRPA for Photometric Adaptation. Comparison of the rendered results reveals that after applying the GRPA module, the artifacts are significantly reduced, leading to improved consistency between the input and rendered frames.

respondences, demonstrating enhanced adaptability to low overlap (first row). Additionally, our method performs better under inconsistent photometric conditions in the scene (second row), indicating greater robustness.

5. Conclusion

We introduce OG-UPCR, a framework for unsupervised point cloud registration. The method consists of two primary modules: OCID and GRVA. The OCID module enables the model to adapt to low-overlap scenes while maintaining a balance between accuracy and efficiency. The GRVA module refines the initial geometric transformation, enhancing the model’s robustness to inconsistent photometric information. Extensive experiments on public datasets demonstrate the effectiveness of the proposed method.

Acknowledgement. This work is supported by the National Key R&D Program of China (No. 2023YFB3307202), Science and Technology Innovation Leading Talent Project of Special Support Plan for High-level Talents of Zhejiang Province (Grant No. 2022R52053).

References

- [1] Sheng Ao, Qingyong Hu, Hanyun Wang, Kai Xu, and Yulan Guo. Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration. In *CVPR*, pages 1255–1264, 2023. 6
- [2] Sheng Ao, Qingyong Hu, Hanyun Wang, Kai Xu, and Yulan Guo. Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration. In *CVPR*, pages 1255–1264, 2023. 2
- [3] Daniel Barath, Jiri Matas, and Jana Noskova. Magsac: marginalizing sample consensus. In *CVPR*, pages 10197–10205, 2019. 2
- [4] David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. pixelsplat: 3D gaussian splats from image pairs for scalable generalizable 3D reconstruction. In *CVPR*, pages 19457–19467, 2024. 2, 3
- [5] Zhi Chen, Kun Sun, Fan Yang, and Wenbing Tao. Sc2pcr: A second order spatial compatibility for efficient and robust point cloud registration. In *CVPR*, pages 13221–13231, 2022. 3, 4, 5
- [6] Zilong Chen, Feng Wang, Yikai Wang, and Huaping Liu. Text-to-3D using gaussian splatting. In *CVPR*, pages 21401–21412, 2024. 2, 3
- [7] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *ICCV*, pages 8958–8966, 2019. 6
- [8] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *CVPR*, pages 2514–2523, 2020. 4
- [9] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3D reconstructions of indoor scenes. In *CVPR*, pages 5828–5839, 2017. 2, 6
- [10] Zheng Dang, Lizhou Wang, Yu Guo, and Mathieu Salzmann. Learning-based point cloud registration for 6D object pose estimation in the real world. In *ECCV*, pages 19–37. Springer, 2022. 1
- [11] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *CVPR*, pages 224–236, 2018. 6
- [12] Mohamed El Banani and Justin Johnson. Bootstrap your own correspondences. In *ICCV*, pages 6433–6442, 2021. 1, 2, 6
- [13] Mohamed El Banani, Luya Gao, and Justin Johnson. Unsupervisedr&r: Unsupervised point cloud registration via differentiable rendering. In *CVPR*, pages 7129–7139, 2021. 2, 4, 5, 6
- [14] Mohamed El Banani, Ignacio Rocco, David Novotny, Andrea Vedaldi, Natalia Neverova, Justin Johnson, and Ben Graham. Self-supervised correspondence estimation via multiview registration. In *WACV*, pages 1216–1225, 2023. 1
- [15] 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. 2, 4
- [16] Zan Gojcic, Caifa Zhou, Jan D Wegner, Leonidas J Guibas, and Tolga Birdal. Learning multiview 3D point cloud registration. In *CVPR*, pages 1759–1769, 2020. 6
- [17] Ning Guo, Baohua Zhang, Jun Zhou, Ketian Zhan, and Shuang Lai. Pose estimation and adaptable grasp configuration with point cloud registration and geometry understanding for fruit grasp planning. *Computers and Electronics in Agriculture*, 179:105818, 2020. 1
- [18] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 3
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 4
- [20] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3D point clouds with low overlap. In *CVPR*, pages 4267–4276, 2021. 2
- [21] Andrew E Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE TPAMI*, 21(5):433–449, 1999. 2
- [22] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat track & map 3D gaussians for dense RGB-D slam. In *CVPR*, pages 21357–21366, 2024. 5
- [23] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat track & map 3D gaussians for dense RGB-D slam. In *CVPR*, pages 21357–21366, 2024. 2, 3
- [24] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM TOG*, 42(4):1–14, 2023. 1, 2, 3
- [25] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004. 4, 6
- [26] Hidenobu Matsuki, Riku Murai, Paul HJ Kelly, and Andrew J Davison. Gaussian splatting slam. In *CVPR*, pages 18039–18048, 2024. 2, 3, 5
- [27] 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 *CVPR*, pages 13611–13620, 2023. 1
- [28] 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 *CVPR*, pages 13611–13620, 2023. 2
- [29] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *CVPR*, pages 11143–11152, 2022. 2

- [30] Wang Qingshan and Zhang Jun. Point cloud registration algorithm based on combination of ndt and plicp. In *2019 15th International Conference on Computational Intelligence and Security (CIS)*, pages 132–136. IEEE, 2019. 1
- [31] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: A universal framework for random sample consensus. *IEEE TPAMI*, 35(8):2022–2038, 2012. 2
- [32] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3D deep learning with pytorch3d. *ArXiv preprint arXiv:2007.08501*, 2020. 1, 5, 7
- [33] Radu Bogdan Rusu, Nico Blodow, Zoltan Csaba Marton, and Michael Beetz. Aligning point cloud views using persistent feature histograms. In *IROS*, pages 3384–3391. IEEE, 2008. 2
- [34] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3D registration. In *ICRA*, pages 3212–3217. IEEE, 2009. 2
- [35] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, pages 4938–4947, 2020. 2, 6
- [36] Yaqi Shen, Le Hui, Haobo Jiang, Jin Xie, and Jian Yang. Reliable inlier evaluation for unsupervised point cloud registration. In *AAAI*, pages 2198–2206, 2022. 2
- [37] Yejun Shou, Xiaoyao Wei, Haocheng Wang, Qian Zheng, Shuai Li, Zhijie Xu, Gang Pan, and Yanlong Cao. Point cloud registration based on adaptively fused multimodal features. *IEEE Robotics and Automation Letters*, 2025. 2
- [38] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *ICCV*, pages 6411–6420, 2019. 4
- [39] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique shape context for 3D data description. In *Proceedings of the ACM workshop on 3D object retrieval*, pages 57–62, 2010. 2
- [40] Siddharth Tourani, Jayaram Reddy, Sarvesh Thakur, K Madhava Krishna, Muhammad Haris Khan, and N Dinesh Reddy. Leveraging cycle-consistent anchor points for self-supervised RGB-D registration. In *ICRA*, pages 2737–2744. IEEE, 2024. 1
- [41] Haiping Wang, Yuan Liu, Qingyong Hu, Bing Wang, Jianguo Chen, Zhen Dong, Yulan Guo, Wenping Wang, and Bisheng Yang. Roreg: Pairwise point cloud registration with oriented descriptors and local rotations. *IEEE TPAMI*, 2023. 2
- [42] Haocheng Wang, Yanlong Cao, Yejun Shou, Lingfeng Shen, Xiaoyao Wei, Zhijie Xu, and Kai Ren. is-map: Neural implicit mapping and positioning for structural environments. In *ACCV*, pages 747–763, 2024. 1
- [43] Haocheng Wang, Yanlong Cao, Xiaoyao Wei, Yejun Shou, Lingfeng Shen, Zhijie Xu, and Kai Ren. Structerf-slam: Neural implicit representation slam for structural environments. *Computers & Graphics*, page 103893, 2024. 1
- [44] Yao Wang, Mingxing Zhang, Meng Li, Hongyan Cui, and Xiaogang Chen. Development of a humanoid robot control system based on ar-bci and slam navigation. *Cognitive Neurodynamics*, 18(5):2857–2870, 2024. 1
- [45] Ziming Wang, Xiaoliang Huo, Zhenghao Chen, Jing Zhang, Lu Sheng, and Dong Xu. Improving RGB-D point cloud registration by learning multi-scale local linear transformation. In *ECCV*, pages 175–191. Springer, 2022. 1, 2, 4, 5, 6
- [46] Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. Gs-slam: Dense visual slam with 3D gaussian splatting. In *CVPR*, pages 19595–19604, 2024. 2, 3
- [47] Runzhao Yao, Shaoyi Du, Wenting Cui, Canhui Tang, and Chengwu Yang. Pare-net: Position-aware rotation-equivariant networks for robust point cloud registration. In *ECCV*, pages 287–303. Springer, 2024. 6
- [48] Botao Ye, Sifei Liu, Haofei Xu, Xueting Li, Marc Pollefeys, Ming-Hsuan Yang, and Songyou Peng. No pose, no problem: Surprisingly simple 3D gaussian splats from sparse unposed images. *arXiv preprint arXiv:2410.24207*, 2024. 3
- [49] Zi Jian Yew and Gim Hee Lee. Regtr: End-to-end point cloud correspondences with transformers. In *CVPR*, pages 6677–6686, 2022. 6
- [50] Taoran Yi, Jiemin Fang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, and Xinggang Wang. Gaussian-dreamer: Fast generation from text to 3D gaussian splatting with point cloud priors. *arXiv preprint arXiv:2310.08529*, 2023. 2, 3
- [51] 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 *CVPR*, pages 17702–17711, 2023. 2
- [52] 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 *ICCV*, pages 17694–17705, 2023. 1, 2, 4, 5, 6
- [53] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from RGB-D reconstructions. In *CVPR*, pages 1802–1811, 2017. 2, 6
- [54] Kai Zhang, Sai Bi, Hao Tan, Yuanbo Xiangli, Nanxuan Zhao, Kalyan Sunkavalli, and Zexiang Xu. Gs-lrm: Large reconstruction model for 3D gaussian splatting. *arXiv preprint arXiv:2404.19702*, 2024. 2, 3
- [55] Xiyu Zhang, Jiaqi Yang, Shikun Zhang, and Yanning Zhang. 3D registration with maximal cliques. In *CVPR*, pages 17745–17754, 2023. 3, 4
- [56] Yuchao Zheng, Yujie Li, Shuo Yang, and Huimin Lu. Global-pbnet: A novel point cloud registration for autonomous driving. *IEEE TITS*, 23(11):22312–22319, 2022. 1