ColorPCR: Color Point Cloud Registration with Multi-Stage Geometric-Color Fusion

Juncheng Mu[1] Lin Bie[1] Shaoyi Du[2] Yue Gao[1*]

[1]{BNRist, THUIBCS, School of Software}, Tsinghua University

[2]National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University

{mujc21, biel20}@mails.tsinghua.edu.cn, dushaoyi@xjtu.edu.cn, gaoyue@tsinghua.edu.cn

Abstract

*Point cloud registration is still a challenging and open problem. For example, when the overlap between two point clouds is extremely low, geo-only features may be not sufficient. Therefore, it is important to further explore how to utilize color data in this task. Under such circumstances, we propose ColorPCR for color point cloud registration with multi-stage geometric-color fusion. We design a Hierarchical Color Enhanced Feature Extraction module to extract multi-level geometric-color features, and a GeoColor Superpoint Matching Module to encode transformation-invariant geo-color global context for robust patch correspondences. In this way, both geometric and color data can be used, thus leading to robust performance even under extremely challenging scenarios, such as low overlap between two point clouds. To evaluate the performance of our method, we colorize 3DMatch/3DLoMatch datasets as Color3DMatch/Color3DLoMatch and evaluations on these datasets demonstrate the effectiveness of our proposed method. Our method achieves state-of-the-art registration recall of **97.5%/88.9%** on them.*

1. Introduction

Rigid point cloud registration is a cornerstone in 3D vision and robotics [12, 21], which is fundamental to a series of downstream applications, including 3D semantic segmentation [19, 20] and 3D reconstruction [17]. The objective of registration is to estimate a rigid transformation aligning two overlapping 3D point clouds.

Recently, keypoint-based and correspondence-based methods [3, 6, 10] have shown promising results, which leverage a neural network to estimate the correspondences of the input point clouds and employ a robust estimator, such as random sample consensus (RANSAC [9]), to derive the transformation. Recent methods in the image matching domain have underscored the importance of local neighborhood information [23, 40], inspiring approaches in point cloud registration [21, 34]. These methods perform downsampling to achieve hierarchical points. Then they utilize a transformer to augment the superpoint features, thereby establishing the superpoint correspondences, which can guide the registration of up-sampled points.

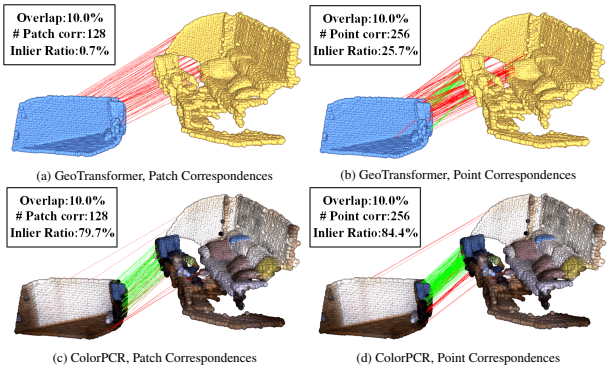

Figure 1. In challenging scenarios with extremely low overlap and geometric distinctiveness, geo-only method fails, while ColorPCR successfully identifies the blue chair.

Although point cloud registration has been investigated for decades, it is still a challenging and open problem dealing with open scenario tasks. When the overlap between two point clouds is extremely low, the registration process may fail in many cases. Geometric features have been deeply explored to deal with this scenario, while the performance is still limited. Therefore, it is important to further explore new representations for robust point cloud registration.

Thanks to recent progress on acquiring color point clouds, it is possible to further investigate how to better explore such color information for registration. Fig. 1 shows examples of how color information is important for registration. As the geometric information is not discriminative,

*Corresponding author

geo-only method cannot find the correct correspondences successfully. However, when color information is available in subfigures (c) and (d), the blue chair in two point clouds is detected. Therefore, it's important to further investigate how to utilize such color data together with geometric information for robust point cloud registration.

To tackle the challenging task of point cloud under a low overlap, we propose ColorPCR for color point cloud registration with multi-stage geometric-color fusion. We design a Hierarchical Color Enhanced Feature Extraction (CEFE) module to extract multi-level features from both geometric and color data, and a GeoColor Superpoint Matching Module to encode transformation-invariant geo-color global context for robust patch correspondences. In this way, both geometric and color data can be used for registration and thus lead to robust performance even when dealing with extremely challenging scenarios, such as low overlap between two point clouds.

We also notice that because there are no publicly available color point cloud datasets, we colorize the widely used 3DMatch/3DLoMatch [12, 37] datasets and generate Color3DMatch/Color3DLoMatch. Evaluations of these datasets demonstrate that our proposed ColorPCR can achieve state-of-the-art performances under all settings. ColorPCR can achieve many gains in performance under low overlap, which demonstrates its effectiveness under extremely challenging scenarios. Moreover, ColorPCR is prior-free and RANSAC-free, so the registration is fast.

The main contributions of this paper are three-fold:

- We propose to jointly use color information and geometric information together for point cloud registration. More specifically, we introduce a feature extraction backbone, named Hierarchical Color Enhanced Feature Extraction (CEFE), which can simultaneously capture both geometric and color information from point clouds.
- We propose a GeoColor SuperPoint Matching Module to encode geometric-color transformation-invariant representations of point clouds. This module achieves further fusion of geometry and color data, leading to accurate superpoint correspondences.
- We generate color point cloud datasets Color3DMatch and Color3DLoMatch. Evaluations have shown superior performance of our proposed method under extremely low overlap settings.

2. Related Work

2.1. Point Cloud Feature Extraction

Inspired by 2D image processing, advancements have been made in 3D point cloud feature extraction. PointNet [19] and PointNet++ [20] pioneered the field of point cloud deep learning. After that, graph convolution [29, 30] and point convolutional kernels [2, 25, 32] are developed. Our CEFE, by integrating color information and KPConv [25], precisely extracts the features.

2.2. Correspondence-based Registration Methods

Recently, Correspondence-based methods [3, 6, 10, 12] have shown promising results in the field of point cloud registration. These methods typically begin with a feature extraction backbone, such as KPConv-FPN [15, 25]. Based on the extracted features, point clouds' correspondences are inferred. Then, a transformation can be estimated with a specific estimator, such as RANSAC or deep robust estimators [4, 7]. Recently, some RANSAC-free methods [21, 34, 35] have adopted a coarse-to-fine approach, achieving state-of-the-art performance. In this work, we follow this approach and enhance it with our powerful CEFE and GeoColor Superpoint Matching Module.

2.3. Local Registration with Color

The Iterative Closest Point (ICP) algorithm [5] and its variants [22] have consistently been pivotal in the field of local point cloud registration. Some variants [13, 14, 16] extend this approach by leveraging color information to enhance accuracy. They elevate the registration to a higher-dimensional space, *e.g.*, (x,y,z,r,g,b), to estimate the transformation. Other methods [18] utilize color to optimize the geometric transformation. Inspired by these methods, we leverage color in global registration domain and follow 4DICP [16] to employ the HSV system. Through the GeoColor Fusion Network, we can accomplish efficient registration without disrupting the spatial structure.

2.4. Color-Related Registration Methods

Some methods have implicitly harnessed the information provided by color. They utilize color to detect keypoints [24] or to support 2D-3D multi-modal learning [8, 31, 35, 36, 39]. PEAL [35], for instance, uses color through RGB images, leveraging existing 2D image techniques to detect overlapping regions from images, which are then transferred to a 3D registration network as priors. However, such utilization of color information can lead to information loss due to indirectness. In contrast, ColorPCR directly incorporates color into the multi-stage registration process, resulting in highly specific point-wise features.

3. Method

Given two color point clouds $\mathcal{P} = \{\mathbf{p_i} \in \mathbb{R}^3 | i = 1, ..., N\}$ with $\mathcal{P}_c = \{\mathbf{p_{ci}} \in [0,1]^3 | i = 1, ..., N\}$ and $\mathcal{Q} = \{\mathbf{q_i} \in \mathbb{R}^3 | i = 1, ..., M\}$ with $\mathcal{Q}_c = \{\mathbf{q_{ci}} \in [0,1]^3 | i = 1, ..., M\}$, in which $\mathcal{P}_c$ and $\mathcal{Q}_c$ denote the color of $\mathcal{P}$ and $\mathcal{Q}$ respectively, our objective is to estimate a rigid transformation $T = \{R, t\}$, where $R \in SO(3)$ and $t \in \mathbb{R}^3$. With the

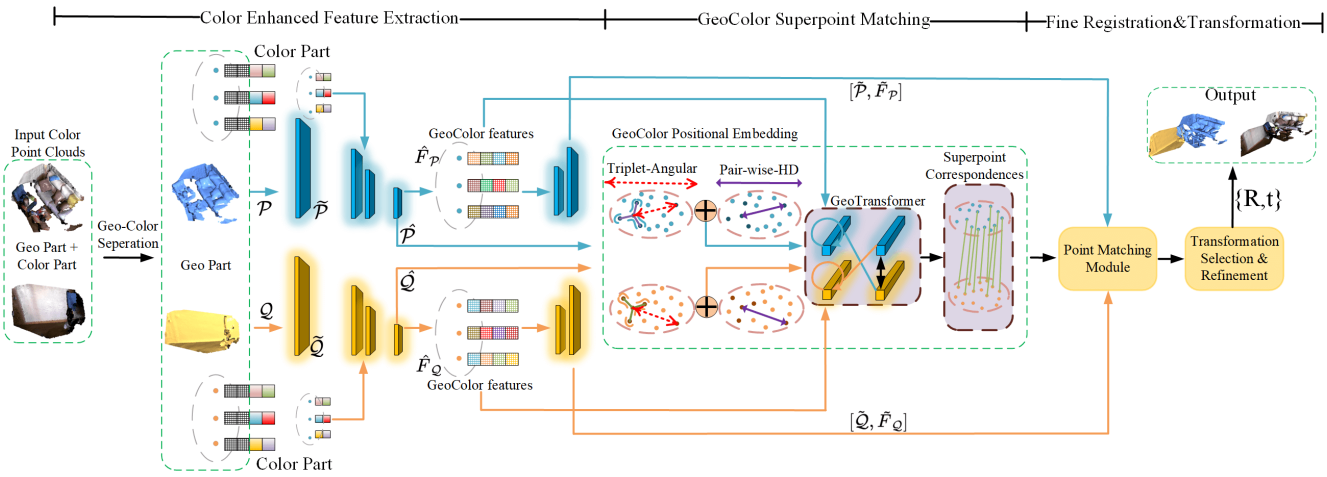

Figure 2. We separate the geometric component from the color point cloud and input it into CEFE, which down-samples these components and leverages the color information to extract geometry-based point features. The GeoColor Superpoint Matching Module encodes superpoints' angles, hue, and distances, and feeds them, along with corresponding points and features, into the GeoTransformer structure. Finally, a Fine Registration process is employed to estimate the most precise transformation.

estimated transformation, we can align the two point clouds. To achieve the most precise alignment, we need to solve the following optimization problem:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{(\mathbf{p}_i^*,\mathbf{q}_i^*)\in\mathcal{C}^*} \|\mathbf{R}\cdot\mathbf{p}_i^* + \mathbf{t} - \mathbf{q}_i^*\|_2^2. \tag{1}$$

In the aforementioned formula, $\mathcal{C}^*$ is the ground-truth correspondence set between the target and source point clouds, representing the overlapping region of the two point clouds. To identify this region, we down-sample the dense points meanwhile performing a geo-color fusion process to generate hierarchical points with highly distinctive features. Based on the patch correspondences derived from these features, the overlapping region can be estimated. We can then proceed to perform a fine registration of the up-sampled points within this region.

3.1. Overview

Our pipeline is illustrated in Fig. 2. We denote the dense points as $\mathcal{P} \in \mathbb{R}^{|\mathcal{P}|\times 3}$ with $\mathcal{P}_c \in [0,1]^{|\mathcal{P}|\times 3}$, the first level down-sampled points as $\tilde{\mathcal{P}} \in \mathbb{R}^{|\tilde{\mathcal{P}}|\times 3}$ with $\tilde{\mathcal{P}}_c \in [0,1]^{|\tilde{\mathcal{P}}|\times 3}$, and the superpoints (*i.e.* the last level down-sampled points) as $\hat{\mathcal{P}} \in \mathbb{R}^{|\hat{\mathcal{P}}|\times 3}$ with $\hat{\mathcal{P}}_c \in [0,1]^{|\hat{\mathcal{P}}|\times 3}$, in which $|\hat{\mathcal{P}}| < |\tilde{\mathcal{P}}| < |\mathcal{P}|$. The corresponding notations for $\mathcal{Q}$ are similar. Then, we employ the Color Enhanced Feature Extraction (CEFE) module (Sec. 3.2) to extract point-wise features from both the dense points and down-sampled points. Throughout this process, color information is harnessed to guide the extraction of features based on the local geometric structure. Specifically, we utilize color information to enhance feature discriminability without disrupting the spatial structural information of the point cloud.

While the CEFE module effectively extracts the features of hierarchical points, it is not enough to encode the color and geometric structure of the global context. To address this, we propose GeoColor Superpoint Matching Module (Sec. 3.3). It procures geo-color global information and further enhances the superpoint features, thereby accomplishing precise superpoint registration.

Following these steps, we follow GeoTransformer [21] to perform fine matching, and ultimately select and refine the estimated transformation to get the final one.

3.2. Color Enhanced Feature Extraction

Hierarchical Feature Extraction. Our Color Enhanced Feature Extraction Module (CEFE) employs the strategy proposed by FPN [15] to carry out hierarchical feature extraction. The point cloud is down-sampled to N_s levels (in accordance with previous works [12, 21, 34], we set $N_s = 4$). The primary registration process does not involve the original dense point cloud for two main reasons: (1) the dense points are excessively clustered, thereby failing to provide refined information, and (2) dense points inevitably contain color noise, which significantly impacts the registration process.

The first three feature extraction levels involve the deliberate injection of color information to enhance the features. We abstain from color enhancement during the fourth level, aiming to ensure stability during the superpoint feature extraction. The subsequent two levels merely involve the concatenation of features [15]. The hierarchical feature extraction process is illustrated in Fig. 3.

Color Enhanced Point Convolution. We have designed a color-enhanced point convolution process for hierarchical

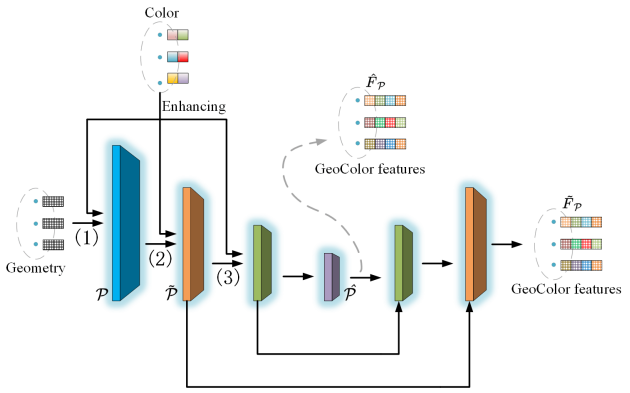

Figure 3. The initial three levels of the CEFE module take the points and corresponding features as input, enhancing the features based on the color information of the point cloud. With the introduction of color, the geo-color feature is progressively enhanced at each level and is propagated to $\tilde{\mathcal{P}}$, which are subsequently used to estimate correspondences.

feature extraction. For any given level-n point $x \in \mathbf{X}^n$, its color-enhanced feature f_e can be computed with the following recursive formula:

$$f_e = G_1(\mathbf{C}^n, \mathbf{X}^n, \mathbf{F}_r, g)(x), \tag{2}$$

where $\mathbf{X}^n \in \mathbb{R}^{\mathcal{N}_n \times 3}$ with $\mathbf{C}^n \in [0,1]^{\mathcal{N}_n \times 3}$ represents the color point set of level n, G_1 is a convolution function to be described later, g is the kernel function and $\mathbf{F_r} \in \mathbb{R}^{\mathcal{N}_n \times d_r}$ is the rough feature set of $\mathbf{X}^n$, extracted from the points of the previous level. $\mathbf{F}_r$ can be computed as:

$$\mathbf{F}_r^n = \{G_2(\mathbf{X}^{n-1}, \mathbf{F}^{n-1}, g)(x_i^n) | x_i^n \in \mathbf{X}^n\}. \tag{3}$$

Here, $\mathbf{X}^{n-1} \in \mathbb{R}^{\mathcal{N}_{n-1} \times 3}$ and $\mathbf{F}^{n-1} \in \mathbb{R}^{\mathcal{N}_{n-1} \times d_{n-1}}$ are the point set and feature set of level $n-1$. G_2 is also a convolution function, with the only difference between G_1 and G_2 being that G_1 incorporates a color enhancement process. By combining Eq. (2) and Eq. (3), we have successfully introduced color into the point features. In summary, from the previous level point features, we employ the G_2 convolution to obtain coarse point features. Afterward, we introduce color using the color-enhanced G_1 convolution, adjusting the blended features based on the point's geometric information. To achieve a comprehensive geo-color fusion, we further perform G_2 convolution again, resulting in highly distinctive point features. Fig. 4 illustrates the detailed execution process of the convolution. Note that in the first level of feature extraction, due to the absence of the previous level, the extraction process commences directly with the G_1 convolution.

The specific calculations of the convolution are as follows. Given a point set $\mathbf{X}$ with its color set $\mathbf{C}$ and its feature

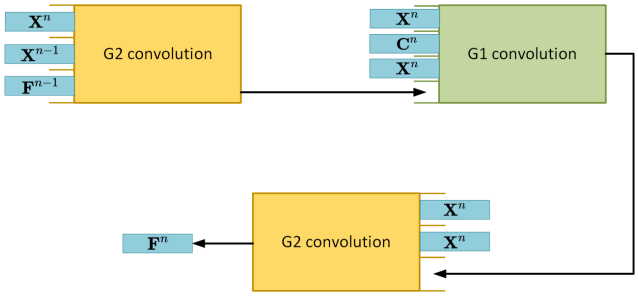

Figure 4. The joint execution of G1 and G2 convolutions is carried out for feature extraction. Here, $\mathbf{X}^n$ and $\mathbf{C}^n$ represent the color points of level n, and $\mathbf{X}^{n-1}$ and $\mathbf{F}^{n-1}$ represents the points and features of the previous level respectively. The first G2 aims to obtain rough features from the previous-level points. This is followed by the execution of G1 to enhance the geometric features with color information. Finally, G2 is executed again to facilitate further integration of color and geometric information.

set $\mathbf{F}$, we have:

$$G_1(\mathbf{C}, \mathbf{X}, \mathbf{F}, g)(x) = \sum_{x_i \in \mathcal{N}x \subset \mathbf{X}} g(x_i - x) f_i^c \tag{4}$$

and

$$G_2(\mathbf{X}, \mathbf{F}, g)(x) = \sum_{x_i \in \mathcal{N}x \subset \mathbf{X}} g(x_i - x) f_i. \tag{5}$$

Here, g is the kernel function, f_i is the feature of x_i, and f_i^c is the feature incorporated with color, which we will describe later. We follow the approach of [11, 25] to use a radius neighborhood $\mathcal{N}x = \{x_i \in \mathbf{X} | \, \|x_i - x\| \leq r, r \in \mathbb{R}\}$ to perform convolution, which ensures robustness in cases with varying densities. For the sake of a more concise form, we use y_i to represent $x_i - x$. Those patterns closely resemble the general point convolution, except for f_i^c, which is calculated by:

$$f_i^c = \text{concat}([fi, c_i/\sigma_c]). \tag{6}$$

Here, $c_i = \{h_i, s_i, v_i\}$ is the color of point x_i in the HSV color space, and σ_c is a hyperparameter which controls the strength of incorporating color (typically, we set $\sigma_c = 1.0$). Then, we follow [25] to perform the subsequent calculation. Function g operates within the ball $\mathcal{B}_r^3 = \{y \in \mathbb{R}^3 \mid \|y\| \leq r\}$ with a radius r. We implement it with weight matrices and kernel points that can fill the neighborhood. We denote the kernel points as $\{\tilde{x}_k \mid k < K\} \subset \mathcal{B}_r^3$ and use $\{\mathbf{W}_k \mid k < K\} \subset \mathbb{R}^{D_{\text{in}} \times D_{\text{out}}}$ to represent the corresponding weight matrix, where K is the number of kernel points, D_{in} and D_{out} are the input and output feature dimensions. Then, for any $y_i = x_i - x \in \mathcal{B}_r^3$, we have:

$$g(y_i) = \sum_{k<K} h(y_i, \tilde{x}_k) \mathbf{W}_k, \tag{7}$$

where h measures the proximity between y_i and the $\tilde{x}_k$, which should have a larger value when they are closer. It can be computed as:

$$h(y_i, \tilde{x}_k) = max(0, 1 - \frac{\|y_i - \tilde{x}_k\|}{\sigma}). \quad (8)$$

Here, σ is exactly the hyperparameter described in KPConv [25] which should be adjusted according to input density.

3.3. GeoColor Superpoint Matching

In the CEFE module, we generate hierarchical points with distinctive features. These features significantly improve registration performance. However, the intra-point-cloud geometric structure has not been fully integrated into the registration network. Previous works [12, 21, 34] have leveraged transformers to encode global contextual information, which has proven critical. However, these methods do not encompass the information pertaining to color, leading to the loss of structural information. To address these issues and achieve deep geometric-color fusion, we propose a novel GeoColor Superpoint Matching Module that can encode transformation-invariant structure. In this module, we implement GeoColor Positional Embedding to acquire geo-color fusion encoding, followed by the utilization of GeoTransformer [21] to extract color-guided superpoint features in accordance with the structure of the point cloud.

Due to uncertainties such as camera shooting angles and lighting conditions, color sampling often contains noise. To mitigate the interference of noise in the depiction of the global context, it is imperative to segregate the components that are easily influenced by noise to the greatest extent feasible. Therefore, we use the HSV color space, allowing the separation of the most stable hue (for relevant experiments, please refer to Sec. 4.8), which participates in the embedding process to achieve robust structural encoding.

Given two superpoints $p, q \in \hat{\mathcal{P}}$ with their hues h_p, h_q, our embedding process can be computed by the following two parts.

(1) Pair-wise Hue-Distance(HD) Embedding. We denote the hue difference between p and q as $\Delta\hat{\mathcal{H}} = \|h_p - h_q\|_2$ and the three-dimensional Euclidean distance as $d = \|p - q\|_2$. The Pair-wise HD Embedding is calculated with:

$$r^{HD} = F_s(\frac{\Delta\hat{\mathcal{H}} \cdot d}{\sigma_{HD}})\mathbf{W}^{HD}. \quad (9)$$

Here, F_s is a sinusoidal function proposed by [27] and σ_{HD} is a hyperparameter that can tune the sensitivity of HD variations. $\mathbf{W}^{HD} \in \mathbb{R}^{d_t \times d_t}$ is the projection matrices for HD embedding, where d_t is the output dim of F_s.

(2) Triplet-wise Angular Embedding. We follow [21] to compute angular embedding with triplets of superpoints. First, assuming p is the selected *embedding point*, we denote the set of its k (typically $k = 3$) nearest neighbor points as $\mathcal{K}$. For $k_i \in \mathcal{K}$, we calculate the angle $a_k = \angle kpq$. By denoting the angular embedding r^A, we have:

$$r^A = \max_{k_i \in \mathcal{K}} F_s(\frac{a_k}{\sigma_A}) \times \mathbf{W}^A \quad (10)$$

Here, σ_A is a hyperparameter that can tune the sensitivity of angular variations, and $\mathbf{W}^A \in \mathbb{R}^{d_t \times d_t}$ is the projection matrices for angular embedding.

To combine the two types of embeddings, the final structure embedding r can be obtained by simply adding r^{HD} and r^A. Along with the embeddings, superpoints carrying distinctive features are introduced into GeoTransformer [21] for color-guided global contextual information extraction. Then, the final superpoint features are used to calculate the Gaussian Correlation Matrix and the top k items are the ultimately selected superpoint correspondences.

4. Experiments

In this section, we evaluate ColorPCR on our indoor datasets Color3DMatch and Color3DLoMatch, in which each point is equipped with an RGB value.

4.1. Color3DMatch & Color3DLoMatch

Color3DMatch (C3DM) and Color3DLoMatch (C3DLM) are derived from 3DMatch [37]. 3DMatch is composed of 62 scenes, among which 46 are used for training, 8 for validation, and 8 for testing. Most of the scenes have their corresponding RGB-D data, but a small fraction only have depth data. We follow previous work's data splits [12] for effective performance evaluation while only removing the depth-only scene [26]. Specifically, only a small set of training data is removed and all the validate and test data are kept the same. The point cloud pairs in C3DM have an overlap of more than 0.3, while those in C3DLM have a lower overlap of 0.1-0.3. To validate the performance of the registration model in cases of extremely low overlap and geometric distinctiveness, we provide additional data divisions, dividing the overlap into the ranges of 0.1-0.15, 0.15-0.2, 0.2-0.25, and 0.25-0.3 respectively.

4.2. Metrics

We primarily evaluate ColorPCR's performance with three metrics as used in prior works [3, 12, 21]: (1) Inlier Ratio (IR), which is the fraction of putative correspondences whose residuals are below a certain threshold (0.1m) under the ground-truth transformation, (2) Feature Matching Recall (FMR), the fraction of point cloud pairs whose inlier ratio is above a certain threshold (5%), and (3) Registration Recall (RR), the fraction of point cloud pairs whose transformation error is smaller than a certain threshold (RMSE$< 0.2m$). To make our experiments more

#Samples	Color3DMatch					Color3DLoMatch				
	5000	2500	1000	500	250	5000	2500	1000	500	250
Feature Matching Recall (%)↑										
PerfectMatch [10]	95.0	94.3	92.9	90.1	82.9	63.6	61.7	53.6	45.2	34.2
FCGF [6]	97.4	97.3	97.0	96.7	96.6	76.6	75.4	74.2	71.7	67.3
D3Feat [3]	95.6	95.4	94.5	94.1	93.1	67.3	66.7	67.0	66.7	66.5
SpinNet [1]	97.6	97.2	96.8	95.5	94.3	75.3	74.9	72.5	70.0	63.6
Predator [12]	96.6	96.6	96.5	96.3	96.5	78.6	77.4	76.3	75.7	75.3
YOHO [28]	98.2	97.6	97.5	97.7	96.0	79.4	78.1	76.3	73.8	69.1
CoFiNet [34]	98.1	98.3	98.1	98.2	98.3	83.1	83.5	83.3	83.1	82.6
GeoTransformer [21]	97.9	97.9	97.9	97.9	97.6	88.3	88.6	88.8	88.6	88.3
PEAL [35]	99.0	99.0	99.1	99.1	98.8	91.7	92.4	92.5	92.9	92.7
ColorPCR (ours)	**99.5**	**99.5**	**99.5**	**99.5**	**99.5**	**96.5**	**96.5**	**97.0**	**97.0**	**96.7**
Inlier Ratio (%)↑										
PerfectMatch [10]	36.0	32.5	26.4	21.5	16.4	11.4	10.1	8.0	6.4	4.8
FCGF [6]	56.8	54.1	48.7	42.5	34.1	21.4	20.0	17.2	14.8	11.6
D3Feat [3]	39.0	38.8	40.4	41.5	41.8	13.2	13.1	14.0	14.6	15.0
SpinNet [1]	47.5	44.7	39.4	33.9	27.6	20.5	19.0	16.3	13.8	11.1
Predator [12]	58.0	58.4	57.1	54.1	49.3	26.7	28.1	28.3	27.5	25.8
YOHO [28]	64.4	60.7	55.7	46.4	41.2	25.9	23.3	22.6	18.2	15.0
CoFiNet [34]	49.8	51.2	51.9	52.2	52.2	24.4	25.9	26.7	26.8	26.9
GeoTransformer [21]	71.9	75.2	76.0	82.2	85.1	43.5	45.3	46.2	52.9	57.7
PEAL [35]	72.4	79.1	84.1	86.1	87.3	45.0	50.9	57.4	60.3	62.2
ColorPCR (ours)	**75.0**	**80.5**	**84.7**	**86.5**	**87.8**	**51.2**	**56.6**	**63.1**	**66.0**	**68.0**
Registration Recall (%)↑										
PerfectMatch [10]	78.4	76.2	71.4	67.6	50.8	33.0	29.0	23.3	17.0	11.0
FCGF [6]	85.1	84.7	83.3	81.6	71.4	40.1	41.7	38.2	35.4	26.8
D3Feat [3]	81.6	84.5	83.4	82.4	77.9	37.2	42.7	46.9	43.8	39.1
SpinNet [1]	88.6	86.6	85.5	83.5	70.2	59.8	54.9	48.3	39.8	26.8
Predator [12]	89.0	89.9	90.6	88.5	86.6	59.8	61.2	62.4	60.8	58.1
YOHO [28]	90.8	90.3	89.1	88.6	84.5	65.2	65.5	63.2	56.5	48.0
CoFiNet [34]	89.3	88.9	88.4	87.4	87.0	67.5	66.2	64.2	63.1	61.0
GeoTransformer [21]	92.0	91.8	91.8	91.4	91.2	75.0	74.8	74.2	74.1	73.5
PEAL [35]	94.6	93.7	93.7	93.9	93.4	81.7	81.2	80.8	80.4	80.1
GeoTransformer+MAC [38]	95.7	95.7	95.2	95.3	94.6	78.9	78.7	78.2	77.7	76.6
ColorPCR (ours)	**96.7**	**96.5**	**97.0**	**96.4**	**96.5**	**88.9**	**88.5**	**88.1**	**86.5**	**85.0**

Table 1. Evaluation results on Color3DMatch and Color3DLoMatch. #Samples in the table represents the number of correspondences selected by RANSAC.

comprehensive, we also evaluate the Relative Rotation Error (RRE) and Relative Translation Error (RTE).

4.3. RANSAC Estimator Results

We collect the recent state-of-the-art methods: PerfectMatch [10], FCGF [6], D3feat [3], SpinNet [1], YOHO [28], CofiNet [34], GeoTransformer [21], PEAL [35], and MAC [38] to compare our performance with them. Most of them use RANSAC as an estimator, so we follow them and report the results with different numbers of sampled correspondences in Tab. 1. For *Feature Matching Recall* (FMR), our method improves to 99.5% on C3DM and 96.5%-97.0% on C3DLM. For *Inlier Ratio* (IR), ColorPCR improves to 87.8% on C3DM and 68.0% on C3DLM at most. For *Registration Recall* (RR), ColorPCR improves to 97.0% on C3DM and 88.9% on C3DLM at most.

4.4. ColorPCR with Different Overlaps

Most previous methods solely focused on the geometric structure of point clouds, which resulted in the loss of color information in point cloud registration. In simple cases with high point cloud overlap and geometric distinctiveness, they can perform relatively well. However, in challenging scenarios with extremely low overlap and low geometric distinctiveness, these methods struggle to achieve accurate registration. While guided by color, ColorPCR facilitates the registration process under challenging conditions. Fig. 5 illustrates the performance of ColorPCR and some geometric-only comparison methods under different overlaps. With the decrease in overlap, the disparity between ColorPCR and geo-only methods becomes more pronounced. At an overlap of 10%-15%, ColorPCR shows an improvement of approximately **40%** compared to Predator, around **30%** compared to Cofinet, and roughly **25%** compared to GeoTransformer.

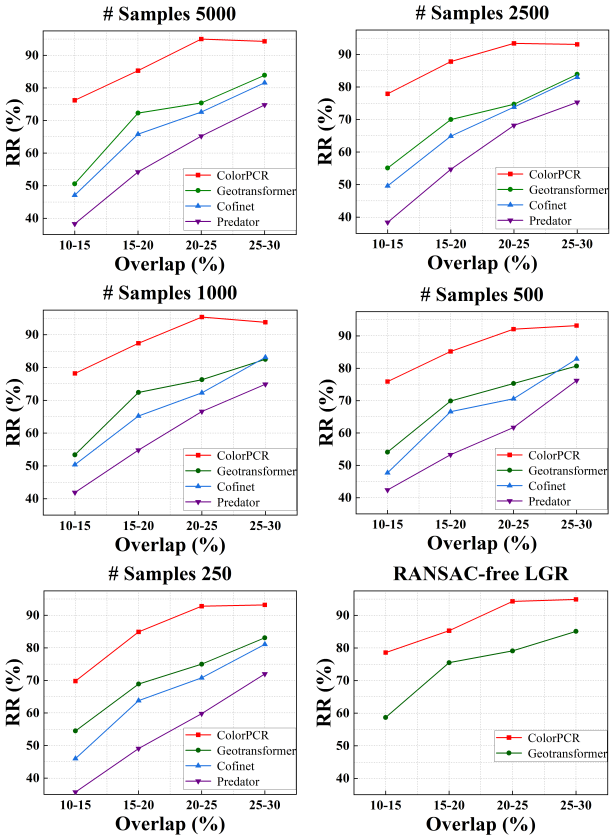

Figure 5. Comparing ColorPCR and some geometric-only methods with different overlaps.

4.5. RANSAC-free Estimator Results

We list the recent RANSAC-free methods' registration results in Tab. 2, as well as compare them with their RANSAC version. In the table, the parentheses after ColorPCR denote the phases of color enhancement being adopted (see Sec. 4.8). On Color3DMatch, ColorPCR (1) with RANSAC as an estimator achieves the highest RR, followed by it with LGR. Due to the high discriminability of point features, we can use LGR to achieve the second-best results on Color3DMatch. On Color3DLoMatch, our ColorPCR (1,2,3) with RANSAC performs best, followed by it with LGR. The results demonstrate that our ColorPCR achieves more accurate registration than comparison methods without the time-consuming RANSAC estimator.

Model	Estimator	Samples	RR (%)	
			C3DM	C3DLM
CoFiNet [34]	RANSAC-50k	5000	89.3	67.5
GeoTransformer [21]	RANSAC-50k	5000	92.0	75.0
ColorPCR (1,2,3)	RANSAC-50k	5000	96.7	**88.9**
ColorPCR (1)	RANSAC-50k	5000	**97.5**	88.2
CoFiNet [34]	LGR	all	87.6	64.8
GeoTransformer [21]	LGR	all	91.5	74.0
ColorPCR (1,2,3)	LGR	all	96.5	**88.3**
ColorPCR (1)	LGR	all	**97.3**	87.1

Table 2. Registration results with RANSAC and LGR on C3DM and C3DLM. The parentheses after ColorPCR denote the phases of color enhancement being adopted.

	Estimator	C3DM		C3DLM	
		RRE (°)	RTE (m)	RRE (°)	RTE (m)
Predator [12]	RANSAC-50k	2.029	0.064	3.048	0.093
CoFiNet [34]	RANSAC-50k	2.002	0.064	3.271	0.090
GeoTransformer [21]	RANSAC-free	1.772	0.061	2.849	0.088
REGTR [33]	RANSAC-free	1.567	0.049	2.827	0.077
PEAL [35]	RANSAC-free	1.748	0.062	2.788	0.087
ColorPCR (ours)	RANSAC-free	**1.492**	**0.048**	**2.581**	**0.075**

Table 3. Relative Rotation Errors (RRE) and Relative Translation Errors (RTE) on C3DM and C3DLM benchmarks.

4.6. RRE and RTE

We compare RRE and RTE with the recent RANSAC and RANSAC-free state of the arts [12, 21, 33–35] in Tab. 3. On both Color3DMatch and Color3DLoMatch, our ColorPCR achieves the best performance.

4.7. CEFE Is Agnostic to Different Methods

To prove the universality and effectiveness of Hierarchical Color Enhanced Feature Extraction (CEFE, see Sec. 3.2), we replaced the backbone network of some methods with CEFE. As illustrated in Tab. 4, applying CEFE to these network structures significantly improves registration performance. Specifically, we evaluate CEFE on three methods [12, 21, 34]. With Predator, RR is improved by 4.9% on C3DM and 13.9% on C3DLM at most. With CoFiNet, the improvement is 7.1% on C3DM and 13.0% on C3DLM at most. GeoTransformer with CEFE achieves the best *RR*, with an improvement of 4.9% on C3DM and 12.3% on C3DLM at most. These experiments demonstrate the strong universality of CEFE, as it can effectively extract precise point-wise geometric-color fusion features from color point clouds.

4.8. Ablation Studies

In this section, we conduct ablation studies on the different components of our ColorPCR to confirm the effectiveness of each component individually. All our ablation experiments are conducted with the LGR estimator.

Tab. 5 lists five pertinent experiments. The term *Enhance levels* denotes the phases in which CEFE performs color enhancement, while *Positional embedding* outlines the embedding techniques employed in the experiment.

We first ablate the GeoColor Superpoint Matching. As inferred from experiments (a)(b)(e), under equivalent CEFE (1) condition, geo-color embedding with hue results in a registration recall of (97.3%, 87.1%). The performance of embedding without color introduction on C3DM diminished by 1.1%, while incorporating all components of HSV into the embedding process precipitated a substantial decline in registration recall, which demonstrates the instability of saturation and value parts due to factors such as noise. Consequently, utilizing hue in geo-color positional embedding is the most stable. We conduct further investigation with the Patch Inlier Ratio (PIR) under different overlaps (Tab. 6). The universal improvement of PIR compared with geometric-only superpoint matching demonstrates its significant role. Owing to the precision of superpoint features, ColorPCR can accomplish more robust superpoint correspondences for fine matching process.

Then we ablate the levels in CEFE that perform color enhancement. Experiments (e)(f)(g) respectively evaluated the situations of color enhancement during the first level,

#Samples	Color3DMatch					Color3DLoMatch				
	5000	2500	1000	500	250	5000	2500	1000	500	250
Predator [12]	89.0	89.9	90.6	88.5	86.6	59.8	61.2	62.4	60.8	58.1
CoFiNet [34]	89.3	88.9	88.4	87.4	87.0	67.5	66.2	64.2	63.1	61.0
GeoTransformer [21]	92.0	91.8	91.8	91.4	91.2	75.0	74.8	74.2	74.1	73.5
Predator (CEFE)	92.0	93.8	92.2	93.3	91.5	71.6	75.1	73.0	72.8	70.2
	3.0↑	3.9↑	1.6↑	4.8↑	4.9↑	11.8↑	13.9↑	10.6↑	12.0↑	12.1↑
CoFiNet (CEFE)	93.0	94.0	94.3	94.5	92.9	76.6	77.1	76.4	75.2	74.0
	3.7↑	5.1↑	5.9↑	7.1↑	5.9↑	9.1↑	10.9↑	12.2↑	12.1↑	13.0↑
GeoTransformer (CEFE)	**95.1**	**95.3**	**96.7**	**95.9**	**95.6**	**87.3**	**87.0**	**86.4**	**85.7**	**83.9**
	3.1↑	3.5↑	4.9↑	4.5↑	4.4↑	12.3↑	12.2↑	12.2↑	11.6↑	10.4↑

Table 4. Evaluating the universality of CEFE. The value in the table is Registration Recall (RR).

Experiments	Enhance levels	Positional embedding	RR (%)	
			C3DM	C3DLM
(a)	(1)	geo-only	96.2	87.1
(b)	(1)	geo-color (hsv)	96.3	86.2
(c)	(1)(2)	geo-color (hsv)	96.6	86.9
(d)	(1)(2)(3)	geo-color (hsv)	96.5	86.9
(e)	(1)	geo-color (h)	**97.3**	87.1
(f)	(1)(2)	geo-color (h)	97.1	86.5
(g)	(1)(2)(3)	geo-color (h)	96.5	**88.3**

Table 5. Performance differences in network structures with various methods of color introduction and positional embedding

Dataset	C3DM	C3DLM	0.1-0.15	0.15-0.2	0.2-0.25	0.25-0.3
GeoTransformer [21]	86.1	53.6	36.8	53.9	62.0	70.6
ColorPCR (geo-only)	88.2	62.0	45.3	59.6	67.6	74.2
ColorPCR	**89.2**	**62.7**	**45.8**	**60.9**	**68.4**	**75.3**
	1.0↑	0.7↑	0.5↑	1.3↑	1.2↑	1.1↑

Table 6. Registration results for ColorPCR and GeoTransformer. The data in the table is the Patch Inlier Ratio (PIR).

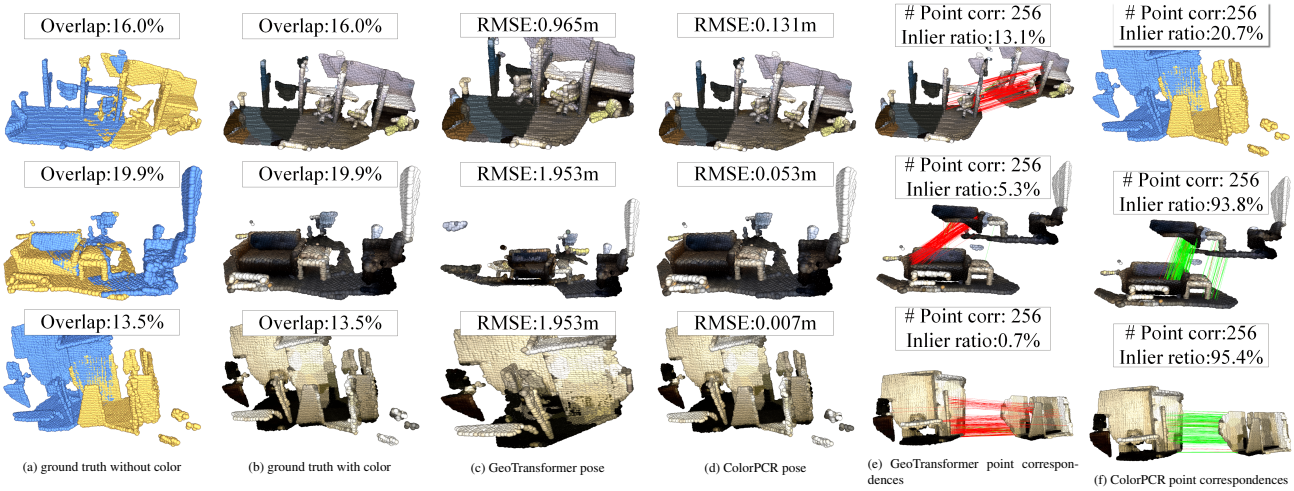

Figure 6. Registration performance with ColorPCR and GeoTransformer. Each row visualizes a challenging registration case with extremely low overlap and geometric distinctiveness where GeoTransformer is entirely ineffective. ColorPCR identifies the color differences of the floor (Row one), the white chair (Row two), and the subtle color differences in the wall (Row three). By fully leveraging color information, ColorPCR achieves robust and effective registration, even in challenging scenarios.

Overlap	0.1-0.15	0.15-0.2	0.2-0.25	0.25-0.3
Patch Inlier Ratio (%)↑				
ColorPCR (1)	45.8	60.9	68.4	75.3
ColorPCR (1,2,3)	**47.6**	**62.7**	**71.7**	**78.9**
Feature Matching Recall (%)↑				
ColorPCR (1)	92.7	**96.8**	**99.4**	97.5
ColorPCR (1,2,3)	**93.5**	96.5	98.6	**98.9**
Inlier Ratio (%)↑				
ColorPCR (1)	38.7	49.5	54.5	59.8
ColorPCR (1,2,3)	**40.0**	**50.8**	**56.8**	**62.4**
Registration Recall (%)↑				
ColorPCR (1)	76.4	86.0	**94.5**	94.9
ColorPCR (1,2,3)	**79.4**	**87.9**	**94.5**	**96.3**

Table 7. The registration results of CEFE (1) and CEFE (1,2,3) under different low overlaps.

the first two levels, and first three levels. The experimental outcomes indicate that CEFE (1) exhibits superior performance under conditions of high overlap. However, under challenging scenarios characterized by low overlap, CEFE (1,2,3) demonstrates better results, suggesting a greater necessity for color information guidance in such situations.

A more detailed investigation was conducted into the performance of ColorPCR (1) and ColorPCR (1,2,3) under different overlaps (Tab. 7). ColorPCR (1,2,3) significantly outperforms ColorPCR (1), with nearly all metrics being considerably higher. The difference in PIR ranges from 1.8% to 3.6%, indicating a substantial disparity even at the superpoint matching phase. In contrast to high overlap registration scenarios where geometric features require more attention, in challenging cases, the registration necessitates a greater emphasis on color guidance.

4.9. Qualitative Results

Fig. 6 visualizes the performance gap in registration between ColorPCR and GeoTransformer in scenarios with extremely low overlap and geometric distinctiveness. GeoTransformer produces highly scattered correspondence estimates. However, guided by color information, ColorPCR successfully extracts consistent correspondences, achieving accurate registration with a higher inlier ratio.

5. Conclusion

We propose ColorPCR for color point cloud registration. Through Color Enhanced Feature Extraction, we obtain precise multi-level point features. With the GeoColor Superpoint Matching Module, we encode geo-color global context, enabling accurate superpoint registration. ColorPCR achieves robust registration even in challenging scenarios with extremely low overlap. We also generate two color point cloud datasets, *i.e.*, Color3DMatch/Color3DLoMatch. Evaluations on them have demonstrated the effectiveness of our proposed method.

Acknowledgements This work was supported by National Natural Science Funds of China (No. 62088102, 62021002), Beijing Natural Science Foundation (No. 4222025).

References

[1] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. SpinNet: Learning a general surface descriptor for 3D point cloud registration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 11753–11762, 2021. 6

[2] Matan Atzmon, Haggai Maron, and Yaron Lipman. Point convolutional neural networks by extension operators. *arXiv preprint arXiv:1803.10091*, 2018. 2

[3] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3D local features. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6359–6367, 2020. 1, 2, 5, 6

[4] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. PointDSC: Robust point cloud registration using deep spatial consistency. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 15859–15869, 2021. 2

[5] Paul Besl and Neil McKay. Method for registration of 3-d shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, pages 586–606. Spie, 1992. 2

[6] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *Int. Conf. Comput. Vis.*, pages 8958–8966, 2019. 1, 2, 6

[7] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2514–2523, 2020. 2

[8] Mohamed El Banani and Justin Johnson. Bootstrap your own correspondences. In *Int. Conf. Comput. Vis.*, pages 6433–6442, 2021. 2

[9] Martin Fischler and Robert 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. 1

[10] Zan Gojcic, Caifa Zhou, Jan Wegner, and Andreas Wieser. The perfect match: 3D point cloud matching with smoothed densities. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5545–5554, 2019. 1, 2, 6

[11] Pedro Hermosilla, Tobias Ritschel, Pere-Pau Vázquez, Àlvar Vinacua, and Timo Ropinski. Monte carlo convolution for learning on non-uniformly sampled point clouds. *ACM Transactions on Graphics (TOG)*, 37(6):1–12, 2018. 4

[12] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3D point clouds with low overlap. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4267–4276, 2021. 1, 2, 3, 5, 6, 7

[13] Ji Hoon Joung, Kwang Ho An, Jung Won Kang, Myung Jin Chung, and Wonpil Yu. 3D environment reconstruction using modified color ICP algorithm by fusion of a camera and a 3D laser range finder. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3082–3088. IEEE, 2009. 2

[14] Michael Korn, Martin Holzkothen, and Josef Pauli. Color supported generalized-icp. In *International Conference on Computer Vision Theory and Applications*, pages 592–599. IEEE, 2014. 2

[15] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2117–2125, 2017. 2, 3

[16] Hao Men, Biruk Gebre, and Kishore Pochiraju. Color point cloud registration with 4D ICP algorithm. In *IEEE International Conference on Robotics and Automation*, pages 1511–1516. IEEE, 2011. 2

[17] Ajmal Mian, Mohammed Bennamoun, and Robyn Owens. Automatic correspondence for 3D modeling: An extensive review. *International Journal of Shape Modeling*, 11(02): 253–291, 2005. 1

[18] Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Colored point cloud registration revisited. In *Int. Conf. Comput. Vis.*, pages 143–152, 2017. 2

[19] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 652–660, 2017. 1, 2

[20] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017. 1, 2

[21] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 11143–11152, 2022. 1, 2, 3, 5, 6, 7

[22] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pages 145–152. IEEE, 2001. 2

[23] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8922–8931, 2021. 1

[24] Hanzhe Teng, Dimitrios Chatziparaschis, Xinyue Kan, Amit K Roy-Chowdhury, and Konstantinos Karydis. Centroid distance keypoint detector for colored point clouds. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1196–1205, 2023. 2

[25] Hugues Thomas, Charles Ruizhongtai Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. KpConv: Flexible and deformable convolution for point clouds. In *Int. Conf. Comput. Vis.*, pages 6411–6420, 2019. 2, 4, 5

[26] Julien Valentin, Angela Dai, Matthias Nießner, Pushmeet Kohli, Philip Torr, Shahram Izadi, and Cem Keskin. Learning to navigate the energy landscape. In *International Conference on 3D Vision*, pages 323–332. IEEE, 2016. 5

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017. 5

[28] Haiping Wang, Yuan Liu, Zhen Dong, and Wenping Wang. You only hypothesize once: Point cloud registration with rotation-equivariant descriptors. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1630–1641, 2022. 6

[29] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 10296–10305, 2019. 2

[30] 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 (tog)*, 38(5):1–12, 2019. 2

[31] 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 *Eur. Conf. Comput. Vis.*, pages 175–191. Springer, 2022. 2

[32] Wenxuan Wu, Zhongang Qi, and Li Fuxin. PointConv: Deep convolutional networks on 3D point clouds. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9621–9630, 2019. 2

[33] Zi Jian Yew and Gim Hee Lee. REGTR: End-to-end point cloud correspondences with transformers. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6677–6686, 2022. 7

[34] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. CofiNet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884, 2021. 1, 2, 3, 5, 6, 7

[35] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17702–17711, 2023. 2, 6, 7

[36] 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 *Int. Conf. Comput. Vis.*, pages 17694–17705, 2023. 2

[37] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1802–1811, 2017. 2, 5

[38] Xiyu Zhang, Jiaqi Yang, Shikun Zhang, and Yanning Zhang. 3D registration with maximal cliques. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17745–17754, 2023. 6

[39] Yu Zhang, Junle Yu, Xiaolin Huang, Wenhui Zhou, and Ji Hou. PCR-CG: Point cloud registration via deep explicit color and geometry. In *Eur. Conf. Comput. Vis.*, pages 443–459. Springer, 2022. 2

[40] Qunjie Zhou, Torsten Sattler, and Laura Leal-Taixe. Patch2Pix: Epipolar-guided pixel-level correspondences. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4669–4678, 2021. 1