Heritage Science

RESEARCH

Open Access

Enhancing point cloud registration with transformer: cultural heritage protection of the Terracotta Warriors

Yong Wang[1], Pengbo Zhou[2], Guohua Geng[1*], Li An[1*] and Mingquan Zhou[1]

Abstract

Point cloud registration technology, by precisely aligning repair components with the original artifacts, can accurately record the geometric shape of cultural heritage objects and generate three-dimensional models, thereby providing reliable data support for the digital preservation, virtual exhibition, and restoration of cultural relics. However, traditional point cloud registration methods face challenges when dealing with cultural heritage data, including complex morphological and structural variations, sparsity and irregularity, and cross-dataset generalization. To address these challenges, this paper introduces an innovative method called Enhancing Point Cloud Registration with Transformer (EPCRT). Firstly, we utilize local geometric perception for positional encoding and combine it with a dynamic adjustment mechanism based on local density information and geometric angle encoding, enhancing the flexibility and adaptability of positional encoding to better characterize the complex local morphology and structural variations of artifacts. Additionally, we introduce a convolutional-Transformer hybrid module to facilitate interactive learning of artifact point cloud features, effectively achieving local–global feature fusion and enhancing detail capture capabilities, thus effectively handling the sparsity and irregularity of artifact point cloud data. We conduct extensive evaluations on the 3DMatch, ModelNet, KITTI, and MVP-RG datasets, and validate our method on the Terracotta Warriors cultural heritage dataset. The results demonstrate that our method has significant performance advantages in handling the complexity of morphological and structural variations, sparsity and irregularity of relic data, and cross-dataset generalization.

Keywords Cultural heritage protection, Point cloud registration, Convolutional-Transformer, Local geometric perception

Introduction

Cultural heritage protection has long been a focal point of attention for both the global academic community and conservation circles. As a part of the world cultural heritage, the Terracotta Warriors holds significant historical, artistic, and scientific value and embodies the cultural memory of China's long history. However, due to prolonged natural erosion and human-induced damage, the conservation and restoration of the Terracotta Warriors face enormous challenges. In this context, digital technology and artificial intelligence have gradually become important tools, among which the acquisition and processing of three-dimensional point cloud data show great potential in the 3D reconstruction, morphological analysis, and damage assessment of cultural relics [1–3]. Point cloud registration [4, 5], as an indispensable component of point cloud processing, plays a crucial

*Correspondence:
Guohua Geng
ghgeng@nwu.edu.cn
Li An
202310342@stumail.nwu.edu.cn
[1] School of Information Science and Technology, Northwest University, Xian 710127, China
[2] School of Arts and Communication, Beijing Normal University, Beijing 100875, China

role in cultural heritage protection due to its accuracy and efficiency. The main goal of point cloud registration is to integrate multiple datasets collected from different perspectives or time instances into a globally consistent coordinate system, thereby achieving high-precision 3D reconstruction, object recognition, and scene analysis of cultural relics.

With the continuous advancement of point cloud acquisition technologies and the widespread application of sensors and scanning devices, modern point cloud registration encounters numerous challenges. One challenge is the complex morphological and structural variations, sparsity, and irregularity of the data, which lead to inefficiencies and susceptibility to local noise and overlap rates [6–8] when dealing with large-scale [9], high-density complex cultural relic point clouds using traditional registration methods. Additionally, the differences between different datasets [10–12] pose challenges to point cloud registration as they may contain variations in objects, environments, and sampling methods, resulting in insufficient generalization performance of existing algorithms.

To address these challenges, researchers have proposed numerous innovative point cloud registration methods in recent years [13–15]. These methods cover various aspects ranging from traditional feature-based matching approaches to end-to-end methods. However, traditional feature-based matching methods often rely on handcrafted feature descriptors, leading to unstable performance when dealing with point clouds of different densities and scales. While deep learning-based methods partially address the issue of feature extraction, their generalization performance on large-scale cultural relic data and different datasets remains limited.

Inspired by the successful application of the Transformer architecture [16, 17] in the field of natural language processing, recent research has introduced it into the field of computer vision, aiming to capture wide-range relationships and integrate overall contextual information. Our work aims to apply the Transformer to point cloud registration tasks and proposes a novel method called Enhancing Point Cloud Registration with Transformer (EPCRT). Our method utilizes the Transformer architecture to encompass both local and global geometric features of point cloud data, thereby not only improving the accuracy and efficiency of point cloud registration but also providing new technological means for cultural heritage protection.

This paper brings forward significant contributions across the following dimensions:

Local Geometric Perception Mechanism We introduce an innovative approach for local geometric perception and positional encoding, combining local density information and geometric angle encoding to enhance the flexibility and robustness of positional encoding. This mechanism dynamically adjusts positional encoding information based on local structures, thereby better representing the complex local morphology and structural variations of artifacts.

Convolutional-Transformer Hybrid Module We design a convolutional-Transformer hybrid module to facilitate interactive learning of point cloud features, achieving effective fusion of local and global features. This hybrid module captures the global semantic information of point cloud data while retaining local details, thereby improving registration performance and effectively handling the sparsity and irregularity of artifact point cloud data.

Experimental Validation and Performance Evaluation We conduct extensive experimental validation on multiple standard datasets, including 3DMatch, ModelNet, KITTI, and MVP-RG, and validate it on the Terracotta Warriors cultural heritage dataset. Through benchmarking against cutting-edge methods, we demonstrate the effectiveness and superiority of the proposed approach. Experimental results show that EPCRT exhibits significant performance advantages in handling complex morphological and structural variations, sparsity, irregularity, and generalization across different datasets.

Related work

Deep feature learning Methods In the field of cultural heritage protection, point cloud registration tasks are crucial for accurately reconstructing and safeguarding artifacts, and the application of deep learning in point cloud feature extraction has become increasingly prevalent. To address the inadequate registration accuracy of unsupervised point cloud registration algorithms in cases of partial overlap, Shen et al. [18] proposed a dependable technique for evaluating inliers, enhancing the resilience of unsupervised point cloud registration. This method aims to effectively differentiate inliers and capture geometric differences between source point clouds and pseudo-target. Specifically, the method comprises a Matching Graph Optimization module and an Inlier Assessment module. In the Matching Graph Optimization module, aggregation of matching scores from neighbors improves the estimation of point-to-point matching graphs. This neighborhood information aggregation helps construct discriminative matching graphs, providing high-quality correspondences for generating pseudo-target point clouds. The Inlier Assessment module calculates inlier confidences for each estimated correspondence based on structural differences between source and pseudo-target point cloud. Li et al. [19] proposed a point cloud registration method named QGORE, aiming to achieve efficient point cloud registration while

ensuring outlier removal. QGORE's key idea lies in employing a simple yet effective voting method to estimate upper bounds through geometric consistency. This voting method yields results nearly equivalent to the tightness in traditional GORE methods. Moreover, to enhance computational efficiency, QGORE proposed a single-point RANSAC algorithm that explores "rotation correspondences" to estimate lower bounds, significantly reducing the iterations required by the traditional three-point RANSAC algorithm.

To simplify the ego-motion estimation process by removing most of the complex parts and focusing on the core elements, Vizzo et al. [20] proposed a system based on the point-to-point ICP algorithm, combined with adaptive thresholding for correspondence registration, robust kernel functions, motion compensation methods, and point cloud subsampling strategies. The results indicate that this system performs well under various operating conditions and does not require tuning for specific LiDAR sensors.

To address the challenge of reconstructing 3D models of artifacts with limited samples and avoiding overfitting, Zhu et al. [21] proposed a transfer learning-based method to recover the 3D shape of artifact faces from a single old photograph. This method utilizes UV position maps to represent the 3D shape and employs a convolutional neural network to reconstruct the UV position map from the 2D image.

End-to-end Methods To enhance the robustness of point cloud registration algorithms, Zhang et al. [22] proposed an end-to-end learning approach to learn partial permutation matrices. This approach addresses the shortcomings of existing hard assignment methods in handling outliers and avoids misleading correspondences that can arise in soft matching methods. The algorithm introduces a registration framework called the Soft-to-Hard Matching Procedure (S2H matching process). This process consists of two steps: the S-step and the H-step. In the S-step, soft matching matrices, which represent the matching probabilities between corresponding points rather than hard assignments, are learned using techniques like graph signal processing. Then, in the H-step, partial permutation matrices are obtained by projecting and clipping the soft matching matrices, achieving hard assignment and avoiding misleading correspondences.

To address the challenges of partial overlap and different datasets, Tan et al. [23] proposed a framework named MCLNet that leverages multi-level consistency algorithms. MCLNet first trims points outside the overlapping region using point-level consistency. It introduces a multi-scale attention module to ensure consistency learning at various levels, thereby establishing dependable correspondences. This module captures features at different scales, improving the handling of local feature matching in point cloud registration. To further enhance accuracy, the authors propose consistency learning to alleviate the adverse effects of non-coincident points. This method helps manage non-overlapping points in point clouds, preventing them from adversely affecting the matching results and making the overall framework more robust and reliable. Wang et al. [24] proposed a registration method named Neighborhood Multi-compound Transformer (NMCT). Firstly, they introduced Neighborhood Position Encoding, which enhances the ability to extract relevant local feature information and local coordinate information by selecting spatial points using the nearest neighbor method. Secondly, they employed the Multi-compound Transformer as the interaction module for point cloud information, consisting of both spatial and temporal transformers. The combination of these two stages enables NMCT to better handle the complexity and diversity of point cloud data. The algorithm was extensively tested on multiple datasets, demonstrating excellent generalization and robustness.

Transformer Methods In the past few years, there has been notable advancement in point cloud registration techniques leveraging Transformer learning. To seamlessly integrate geometric and visual data from disparate modalities, Wang et al. [25] introduced a Geometric-Aware Visual Feature Extractor. This method gradually fuses geometric and visual information of RGB and depth data using a multi-scale local linear transformation. The depth data's geometric attributes function akin to convolution kernels, reshaping the visual characteristics of RGB data. This process places the generated visual-geometric features in a normalized feature space, mitigating visual differences caused by geometric variations and obtaining more reliable correspondences.

To address the issue of handling the relationships between point clouds in continuous scans during 3D point cloud registration, Zaman et al. [26] proposed a method that uses a continual graph network architecture with an attention mechanism. This approach improves the registration of current point cloud pairs by leveraging the learned associations from previous point cloud pairs, thereby enhancing the expressiveness of the point clouds. The results show that this method significantly improves correspondence performance, registration performance, and generalization ability.

To enhance the performance of registration within expansive 3D environments, Han et al. [27] introduced a model used on Hough voting for rejecting outlier correspondences. This approach utilizes an overlap-based correspondence calculation method and extracts depth geometric features to enhance registration performance under low overlap ratios. Transform parameters are

represented in 6D Hough space using triplet voting to address ambiguity issues during the matching process. Similarity values are employed as features for each vote to reduce ambiguity during training. The algorithm combines fully convolutional geometric feature networks and Transformer attention mechanisms to reduce noise during the voting process. Finally, a binning method is used to determine consensus on correspondences and predict the final transformation parameters. This method demonstrates superior performance on both indoor and outdoor datasets.

Inspired by the successful results achieved by feature learning-based methods, Transformer-based learning approaches, and end-to-end techniques, particularly in addressing challenges such as complex morphological and structural variations, sparsity, irregularity, generalization across different datasets, and outdoor large-scale scenes, we introduce a Transformer-based end-to-end approach for point cloud registration.

Method

In alignment with the architectures of D3feat [28] and Predator [29], the end-to-end algorithm we propose utilizes an encoder-decoder network with a hierarchical structure. Additionally, we employ RANSAC to estimate rigid transformations, as depicted in Fig. 1.

Problem setting

For two sets of points denoted as source $P = \{p_i \in \mathbb{R}^3 \mid i = 1, 2, \ldots, N\}$ and target $Q = \{q_i \in \mathbb{R}^3 \mid i = 1, 2, \ldots, M\}$, both residing in three-dimensional space, where N and M denote the number of points. Point cloud registration endeavors to align them through an unknown 3D rigid transformation $RT = \{R, T\}$, which comprises rotation $R \in SO(3)$ and translation $T \in \mathbb{R}^3$. This transformation aims to minimize the disparities between corresponding points in the source and target clouds, achieving optimal alignment. The formula is as follows:

$$\min_{R,T} \sum\nolimits_{(p_i,q_i)\in\vartheta} \|R \cdot p_i + T - q_i\|_2^2 \tag{1}$$

Here, ϑ symbolizes the ground truth correspondences between points in P and Q. The notation $\|\bullet\|$ signifies the Euclidean distance.

Encoder-decoder

Encoder To process the denser original point cloud P and Q, each in $\mathbb{R}^{N\times 3}$ and $\mathbb{R}^{M\times 3}$, we employ the KPConv module as our foundation. This module, comprising a sequence of residual units and strided convolutions, facilitates downsampling, thereby reducing the number of keypoints to P' and Q', each in $\mathbb{R}^{N'\times 3}$ and $\mathbb{R}^{M'\times 3}$. Additionally, we adopt a shared encoding mechanism to extract pertinent features, yielding $F'_{P'}$ and $F'_{Q'}$, each in $\mathbb{R}^{N'\times D}$ and $\mathbb{R}^{M'\times D}$, where D denotes the feature dimension.

Decoder The decoder module follows a conventional design, featuring a 3-layer network structure. It incorporates upsampling, linear transformation operations, and skip connections as its primary components.

Transformer

Local Geometric Perception Mechanism (LGP): In traditional Transformer models, positional encoding is typically implemented in a fixed manner, such as sine/cosine positional encoding. However, for point cloud data, where the number of points is variable, traditional positional encoding methods are not suitable. Therefore, we introduce a Local Geometric Perception Mechanism, which dynamically adjusts positional encoding information by integrating local density information and

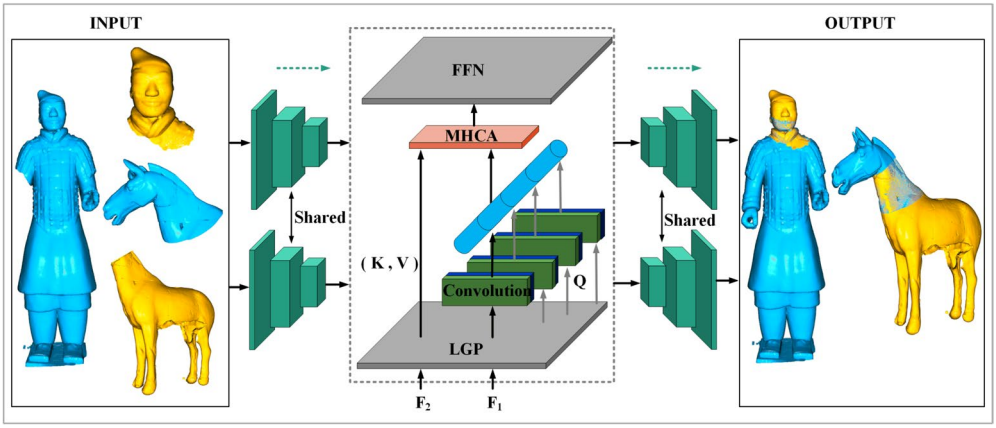

Fig. 1 Network architecture of EPCRT. *LGP* Local Geometric Perception, *MHCA* Multi-head Cross Attention, *FFN* Feedforward Neural Network

geometric angle encoding. The local density information allows the adjustment of positional encoding parameters based on the actual density, enabling the model to better adapt to point clouds of different densities. Additionally, we incorporate geometric angle encoding into the Local Geometric Perception Mechanism to enhance the model's performance by capturing angle information of points in the point cloud. The Local Geometric Perception Mechanism enables the model to better understand the spatial structure and achieve improved performance in point cloud processing tasks.

Local Density Information: Obtaining local density information to adjust the parameters of dynamic positional encoding. Local density information is determined by counting the points located in the immediate vicinity of each point. By defining the local neighborhood using a spherical region, the local density information of each point can be obtained. This information is then used to adjust the magnitude of dynamic positional encoding, allowing it to better adapt to the local structure. The calculation formula is as follows:

$$\psi_i = \sum_{j=1}^{N} \left(K \parallel p_i - p_j \parallel \right) \tag{2}$$

Here, N denotes the aggregate number of points within the point cloud, with p_i indicating the coordinates of the $i-th$ point, and $K(\cdot)$ stands as the kernel function, defined as follows:

$$K(r) = e^{-\frac{r^2}{2\sigma^2}} \tag{3}$$

where, r represents the distance between points,σ serves as the standard deviation of $K(\cdot)$, regulating the extent of the local neighborhood.

Geometric Angle Encoding: Integrating geometric angle position information into dynamic positional encoding, so that positional encoding not only dynamically adjusts its magnitude based on local density information but also fine-tunes positional encoding according to angle information. This enables better capturing of the local structure and directional information. The calculation formula is as follows:

$$\theta_i = arccos\left(f_i \cdot \beta\right) \tag{4}$$

$$\eta_i = [sin(\theta_i), cos(\theta_i)] \tag{5}$$

where, f_i denotes the normal vector of the $i-th$ point, β is the reference direction, θ_i stands as angle between normal vector and the reference direction, and η_i represents the angle information of the $i-th$ point.

Dynamic Fusion Position Encoding: Integrating local density information ψ_i and angle information η_i into positional encoding, dynamically adjusting the magnitude of positional encoding to be correlated with local density. Specifically, points with higher local density will have smaller positional encoding values, while points with lower local density will have larger positional encoding values. Simultaneously, attention is paid to directional information within the point cloud. This way, dynamic positional encoding better adapts to the local structure and enhances the performance of the model in registration tasks. The computation is expressed by the following formula:

$$\begin{aligned} \overset{LGP}{i} = & sin\left(\frac{pos_i}{10000^{2\times d/D\times\psi_i}} + \eta_i\right) \\ & + cos\left(\frac{pos_i}{10000^{2\times d/D\times\psi i}} + \eta_i\right) \end{aligned} \tag{6}$$

where, point cloud data points are represented as $p_i = \left(x_i, y_i, z_i\right)$, pos_i represents the position of the $i-th$ point, $pos_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$, d represent the dimensions of positional encoding, D represent the dimensions of point cloud data, and α_i^{LGP} represents local geometric positional information.

Convolutional-Transformer Network: Traditional point cloud registration methods typically employ iterative local search strategies to achieve registration processes, but they lack in global correlation and feature learning. To optimize the efficiency and accuracy, we introduce a Convolutional-Transformer network.

Firstly, we employ convolutional operations to extract features from the input data, aiming to capture local structural information. This helps reduce the dimensionality of the point cloud data and extract useful feature information. Next, we feed the features extracted by convolutional operations into a Transformer model. The Transformer model achieves global correlation and feature learning among the point cloud data through its cross-attention mechanism. With the multi-head attention mechanism, the Transformer is able to simultaneously consider different aspects of the point cloud data, thereby enhancing the accuracy and robustness.

we define $F'_{P'} = (x_1^{P'}, x_2^{P'} \cdots x_{N'}^{P'})$ and $F'_{Q'} = (x_1^{Q'}, x_2^{Q'} \cdots x_{N'}^{Q'})$ as the input $MHAttn(F'_{P'}, F'_{Q'}, F'_{Q'})$ in the i-th layer, and $Z' = (z_1^{P',Q'}, z_2^{P',Q'} \cdots z_{N'}^{P',Q'})$ as the resulting matrix. The expression is given by:

$$\overset{P',Q'}{i} = \sum_{j=1}^{N'} soft\max\left(\alpha_{i,j}^{Cross-}\right) x_j^{Q'} W^{V,Q'} \tag{7}$$

where, $\alpha_{i,j}^{Cross-}$ represents the unnormalized weight coefficient, characterized as follows:

$$\begin{aligned} {}^{Cross-}_{i,j} = & \frac{1}{\sqrt{d_{head}}} \left(Conv\left(x_i^{P'}\right) W^{Q,P'} + \alpha_i^{LGP} \right) \\ & \left(x_j^{Q'} W^{K,Q'} \right)^T \end{aligned} \tag{8}$$

Finally, the output defines the matching relationship between point clouds, as follows:

$$F_i = MLP\left(cat\left[F'_{P'}, z_i^{P',Q'} \right] \right) \tag{9}$$

Loss function

Our proposed network EPCRT is built upon end-to-end training and supervised using ground truth data. The loss function is as follows:

Feature Loss In line with the methodologies of D3Feat [28] and Predator [29], we employ a circle loss function to assess feature divergence and regulate point-wise feature descriptors in the training. It is defined as follows:

$$\mathcal{L}_{FL}^{P} = \frac{1}{N_P} \sum_{i=1}^{N_P} \log \left[1 + \sum_{j \in \varepsilon_p} e^{c\beta_p^j (d_i^j - \Delta p)} \bullet \sum_{k \in \varepsilon_n} e^{\lambda \beta_p^k (\Delta n - d_i^k)} \right] \tag{10}$$

Here, d_i^j denotes the Euclidean distance between features, $d_i^j = \left\| f_{p_i} - f_{q_j} \right\|_2$. ε_p and ε_n represent the matching and non-matching points of P_{RS} (randomly sampled points from the source point cloud), corresponding to positive and negative regions, respectively. Δp and Δn denote positive and negative areas respectively, and λ is a predefined parameter. Similarly, the feature loss $\mathcal{L}_{FL}^{P}$ for the target point cloud is calculated analogously. The total feature loss is expressed as $\mathcal{L}_{FL} = \frac{1}{2}(\mathcal{L}_{FL}^{P} + \mathcal{L}_{FL}^{Q})$.

Overlap Loss For supervised training, we employ a binary cross-entropy loss function, expressed as:

$$\mathcal{L}_{OL}^{P} = \frac{1}{N} \sum_{i=1}^{N} O_{pi}^{label} \log\left(O_{pi}\right) + \left(1 - O_{pi}^{label}\right) \log\left(1 - O_{pi}\right) \tag{11}$$

where, $O_{p_i}^{label}$ represents the overlapping mark at point p_i of ground truth, characterized as follows:

$$O_{pi}^{label} = \begin{cases} 1, & \left\| T_{P,Q}^{GT}(pi) - NN\left(T_{P,Q}^{GT}(pi), Q\right) \right\| < \tau_1 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where $T_{P,Q}^{GT}$ signifies the ground truth rigid transformation, and NN denotes the nearest neighbor. τ_1 serves as the threshold for overlap determination. Likewise, the overlap loss $\mathcal{L}_{OL}^{Q}$ for the target point cloud is computed in a similar manner. The overall overlap loss is formulated as $\mathcal{L}_{OL} = \frac{1}{2}(\mathcal{L}_{OL}^{P} + \mathcal{L}_{OL}^{Q})$.

In summary, the overall loss function is $\mathcal{L} = \mathcal{L}_{FL} + \mathcal{L}_{OL}$.

Experiments

Dataset and evaluation metrics

To assess the efficacy of EPCRT in handling issues such as complex structural variations, sparsity, irregularity, and large-scale scenes, we conducted extensive experiments on various datasets, including real indoor scenes from 3DMatch [30] and 3DLoMatch [29], synthetic datasets ModelNet [31] and ModelLoNet, incomplete synthetic dataset Multi-View Partial [32], and outdoor large-scale odometry KITTI [33] dataset.

3DMatch The 3DMatch dataset comprises depth images from 62 different scenes sourced from datasets like 7-Scenes and SUN3D. 3DLoMatch is a dataset generated from the 3DMatch dataset. Notably, the overlap ratios for 3DMatch and 3DLoMatch datasets are greater than 30% and between 10% to 30%, respectively.

ModelNet The ModelNet dataset is based on the ModelNet40 dataset, a computer-aided design (CAD) synthetic dataset containing 12,311 models. ModelLoNet is a dataset generated from the ModelNet dataset. The overlap ratios for ModelNet and ModelLoNet datasets are 73.5% and 53.6%, respectively.

MVP-RG The MVP-RG dataset is derived from a synthetic and partially incomplete Multi-View Partial

(MVP) point cloud dataset [34]. It consists of 7,600 pairs models.

Odometry KITTI The Odometry KITTI dataset comprises data captured from city, rural, and highway scenes using the Velodyne HDL-64E S3 LiDAR scanner. There are 11 large scenes.

Evaluation Metrics In line with the approaches of Predator [29], REGTR [35], and GMCNet [32], we evaluated the datasets using Relative Rotation Error (RRE) and Relative Translation Error (RTE). Additionally, Registration Recall (RR), Modified Chamfer Distance (CD), and Root Mean Square Error (RMSE) were employed for evaluating specific datasets. The definitions are outlined below:

$$RTE = \left\| t - t^{GT} \right\|_2 \tag{13}$$

$$RRE = \arccos\left(\frac{trace\left(R^T R^{GT}\right) - 1}{2} \right) \tag{14}$$

$$\begin{aligned} CD(P,Q) = & \frac{1}{|P|} \sum_{p \in P} \min_{q \in Q_{raw}} \left\| T_{P,Q}^{GT}(p) - q \right\|_2^2 \\ & + \frac{1}{|Q|} \sum_{q \in Q} \min_{p \in P_{raw}} \left\| q - T_{P,Q}^{GT}(p) \right\|_2^2 \end{aligned} \tag{15}$$

$$RMSE = \sqrt{\frac{1}{\left|C_{ij}^{GT}\right|} \sum_{(p,q) \in C_{ij}^{GT}} \left\| T_{P,Q}^{GT}(p) - q \right\|_2^2} \tag{16}$$

where, R^{GT} and t^{GT} represent the ground truth error of rotation and translation. C_{ij}^{GT} denotes the collection of ground truth correspondences.

To distinguish Eq. (16), we specify the RMSE for the MVP-RG dataset as follows:

$$\mathcal{L}_{RMSE} = \frac{1}{N} \sum_{i=1}^{N} \left\| T^{GT}(p_i) - T(p_i) \right\|_2 \tag{17}$$

3DMatch and 3DLoMatch

To validate registration performance of EPCRT under low overlap, we adopted the training method from Predator and conducted evaluations on the 3DMatch and 3DLoMatch datasets.

Additionally, we compared EPCRT with other cutting-edge techniques, including FCGF [36], Predator [29], OMNet [37], REGTR [35], GeoTrans [38], RoReg [40], UDPReg [39], MAC [41], and RIGA [42]. Figure 2 shows the registration visualization of low overlap datasets.

Fig. 2 Registration visualization on 3DMatch, 3DLoMatch

As depicted in Table 1, our proposed algorithm not only outperforms other algorithms in terms of the three registration metrics on the sparsity datasets, but it also exhibits lower parameter count and average processing time. In the comparison of the Registration Recall (RR) metric with MAC, UDPReg, RoReg, and GeoTrans algorithms under 3DLoMatch, our proposed algorithm demonstrates improvements of 16.3%, 11.8%, 4.9%, and 2.1% respectively.

ModelNet and ModelLoNet40

To further validate registration performance of EPCRT, we extended the training phase with the Predator and subsequently performed assessments on both the ModelNet and ModelLoNet datasets. Additionally, we compared the EPCRT algorithm against other cutting-edge techniques, including PointNetLK [43], DCP [44], RPM-Net [45], Predator[29], OMNet [37], REGTR [35], UDPReg[39], and HECPG [46]. Figure 3 shows the

Table 1 Performance on 3DMatch and 3DLoMatch datasets

Method	3DMatch			3DLoMatch			Param.(M)	Time(s)
	RR(%)	RRE(°)	RTE(m)	RR(%)	RRE(°)	RTE(m)		
FCGF[36]	85.1	1.949	0.066	40.1	3.147	0.100	8.76	0.16
D3Feat[28]	81.6	2.161	0.067	37.2	3.361	0.103	24.3	0.40
OMNet[37]	90.5	4.166	0.105	8.40	7.299	0.151	–	–
Predator[29]	89.0	2.029	0.064	59.8	3.048	0.093	7.43	0.54
REGTR[35]	92.0	1.567	0.049	64.8	2.827	0.077	–	–
GeoTrans[38]	92.0	1.808	0.063	74.0	2.934	0.089	9.83	0.23
UDPReg[39]	91.4	1.642	0.064	64.3	2.951	0.086	–	–
RoReg[40]	93.2	1.840	0.063	71.2	3.090	0.093	–	–
MAC[41]	93.7	1.890	0.062	59.8	3.500	0.098	–	–
RIGA[42]	89.3	1.798	0.056	65.1	3.016	0.089	–	–
Our	**94.3**	**1.497**	**0.041**	**76.1**	**2.765**	**0.068**	7.5	0.16

The best results are in bold

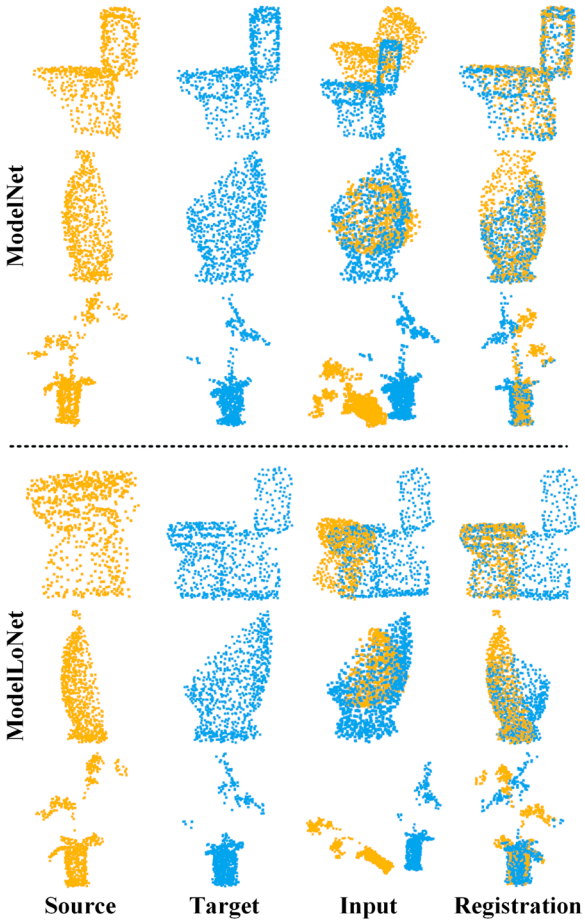

Fig. 3 Registration visualization on ModelNet, ModelLoNet

registration visualization of the ModelNet and ModelLoNet datasets, respectively.

From Table 2, it is evident that our EPCRT achieves superior registration outcomes compared to other algorithms on the ModelNet and ModelLoNet datasets. While our proposed algorithm slightly lags behind the UDPReg algorithm in terms of the Relative Translation Error (RTE) metric, overall, our proposed algorithm exhibits a clear advantage in handling registration tasks.

MVP-RG

To confirm the registration performance of our EPCRT algorithm in incomplete and irregularity models, we trained it using the Predator method and conducted evaluations. Additionally, we compared the proposed algorithm against other cutting-edge techniques, including DCP [44], RPM-Net [45], GMCNet [32], IDAM [47], Predator [29], and DSMNet [48]. Figure 4 shows the registration visualization of the MVP-RG dataset.

From Table 3, it is apparent that our EPCRT algorithm achieves superior results on MVP-RG dataset compared to other algorithms. Through performance comparison with other algorithms, our proposed algorithm demonstrates a clear advantage in handling point cloud registration tasks under incomplete and irregularity scenarios.

Outdoor dataset: odometry KITTI

To confirm the registration performance of EPCRT algorithm in large-scale scenes, we trained it using the Predator method and conducted evaluations on Odometry

Table 2 Performance on ModelNet and ModelLoNet datasets

Method	ModelNet			ModelLoNet		
	CD	RRE(°)	RTE(m)	CD	RRE(°)	RTE(m)
PointNetLK[43]	0.0235	29.725	0.297	0.0367	48.567	0.507
DCP[44]	0.0117	11.975	0.171	0.0268	16.501	0.3
RPM-Net[45]	0.00085	1.712	0.018	0.005	7.342	0.124
OMNet[37]	0.0015	2.947	0.032	0.0074	6.517	0.129
Predator[29]	0.00089	1.739	0.019	0.0083	5.235	0.132
REGTR[35]	0.00078	1.473	0.014	0.0037	3.93	0.087
UDPReg[39]	0.0306	1.331	**0.011**	0.0416	3.578	**0.069**
HECPG [46]	–	1.472	0.016	–	3.371	0.089
Our	**0.0006**	**1.132**	0.013	**0.0035**	**1.585**	0.124

The best results are in bold

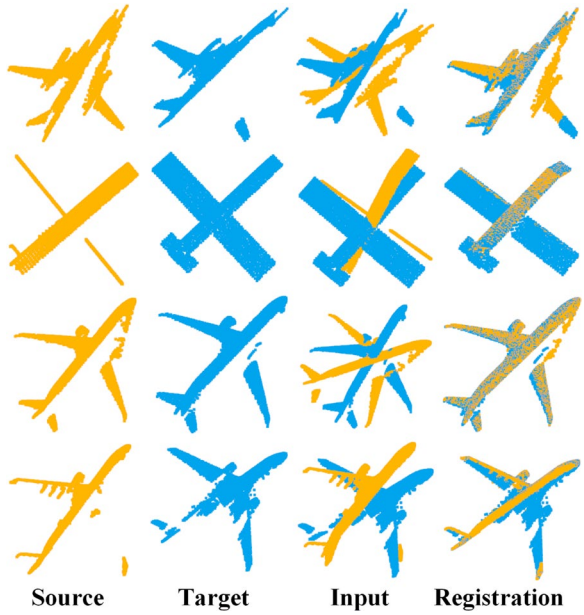

Fig. 4 Registration visualization on MVP-RG

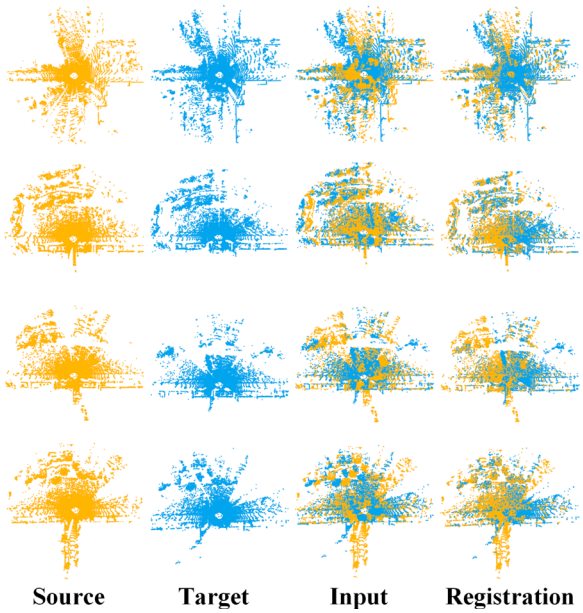

Fig. 5 Registration visualization on Odometry KITTI

Table 3 Evaluation results on MVP-RG dataset

Method	RRE(°)	RTE(m)	$\mathcal{L}_{RMSE}$(%)
DCP [44]	30.37	0.273	0.634
RPM-Net [45]	22.20	0.174	0.327
IDAM [47]	24.35	0.280	0.344
Predator [29]	10.58	0.067	0.125
DSMNet [48]	14.17	0.158	–
GMCNet [32]	16.57	0.174	0.246
Our	**7.33**	**0.061**	**0.022**

The best results are in bold

KITTI dataset. Additionally, we compared the proposed algorithm against other cutting-edge techniques, including FCGF [36], D3Feat[28], Predator [29], SpinNet [49], HRegNet [50], GeoTrans [38], SHM_{DGR}[22], GeDi [51], MAC[41], SC^2 -PCR++ [52] and RIGA [42]. Figure 5 shows the registration visualization of large-scale scenes dataset.

From Table 4, we can see that our EPCRT algorithm achieves superior registration results on the KITTI Odometry dataset compared to other algorithms. Through performance comparison with other algorithms,

Table 4 Evaluation results on Odometry KITTI dataset

Method	RTE(cm)	RRE(°)	RR(%)
FCGF [36]	9.5	0.30	96.6
D3Feat [28]	7.2	0.30	**99.8**
SpinNet [49]	9.9	0.47	99.1
Predator [29]	6.8	0.27	**99.8**
HRegNet [50]	12	0.29	99.7
GeoTrans [38]	7.4	0.27	**99.8**
GeDi [51]	7.5	0.33	**99.8**
SHM_{DGR}[22]	9.3	0.28	97.6
MAC [41]	8.4	0.40	99.5
SC^2 -PCR++ [52]	7.1	0.32	99.6
RIGA [42]	13.5	0.45	99.1
Our	**6.7**	**0.25**	**99.8**

The best results are in bold

our algorithm demonstrates a clear advantage in handling point cloud registration tasks in large-scale scene.

Cultural heritage dataset

To evaluate the registration performance of the proposed algorithm in cultural heritage datasets, we first validate it using the dataset of the Terracotta Warriors in the Mausoleum of the First Qin Emperor collected by Northwest University, as shown in the Figs. 6 and 7. Additionally, we compared the proposed algorithm against other cutting-edge techniques, including Predator [29], as shown in Table 5.

From the Figs. 6, 7 and Table 5, it can be seen that there are good registration results in the head, feet, leg, and arms of the Terracotta Warriors.

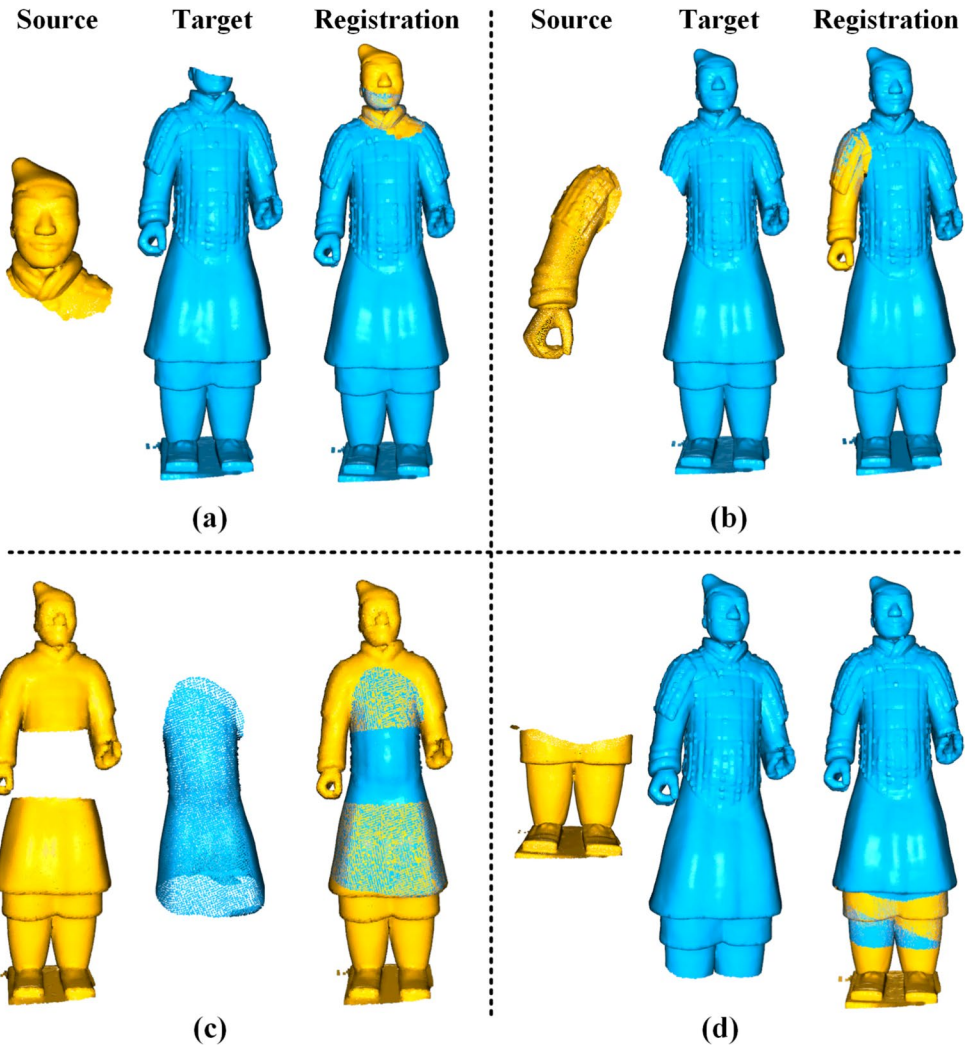

Fig. 6 Registration visualization of 3DMatch → Terracotta Warriors data. (**a** stands for head registration; **b** stands for arm registration; **c** stands for body registration); **d** stands for feet registration)

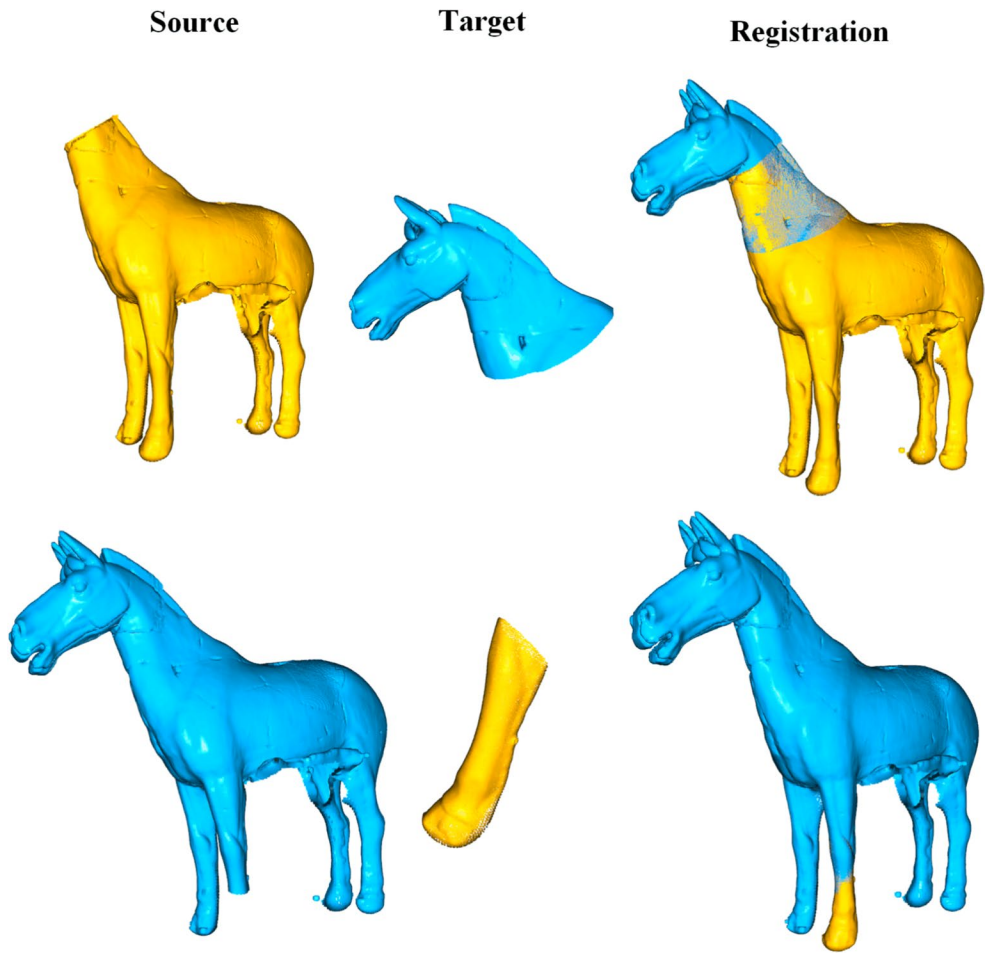

Fig. 7 Registration visualization of 3DMatch → Terracotta Warriors data

Table 5 Evaluation results on Terracotta Warriors dataset

Method	RRE(°)	RTE(cm)	RR(%)
Predator [29]	1.56	10.74	65.91
Our	**0.97**	**7.35**	**88.64**

The best results are in bold

Ablation study

To validate the impact of individual module selection within EPCRT model, we conducted ablation study using the 3DMatch and 3DLoMatch datasets.

From Table 6, it is evident that on top of the original encoder-decoder(Base) architecture, all four individual

Table 6 Ablation of different modules

Method	3DMatch			3DLoMatch		
	RR(%)	RRE(°)	RTE(m)	RR(%)	RRE(°)	RTE(m)
Base	93.2	2.155	0.071	71.1	3.071	0.090
+LDI+C-TNet	93.8	1.854	0.063	73.3	2.878	0.078
+GAE+C-TNet	93.6	1.863	0.066	73.7	2.976	0.085
+LGP+Cross	94.0	1.610	0.059	74.9	1.869	0.0.74
+LGP+C-TNet	**94.3**	**1.497**	**0.041**	**76.1**	**2.765**	**0.068**

The best results are in bold

modules(Local Density Information: LDI; Geometric Angle Encoding: GAE) exhibit certain performance improvements. Among these, the proposed Local Geometric Perception Mechanism (LGP) and the Convolutional-Transformer Network(C-TNet) show significant enhancements in performance. Furthermore, the combined utilization of the proposed modules yields the best overall performance.

Conclusion

This paper proposes a Transformer-based registration method, named Enhancing Point Cloud Registration (EPCRT), aiming to address challenges in cultural heritage protection, such as complex structural variations, sparsity, irregularity, and generalization across different datasets. By introducing dynamic adjustment mechanisms and convolutional-Transformer hybrid modules, our approach can flexibly capture both local and global geometric features and achieve effective feature fusion and interactive learning. Through extensive experimental evaluations on multiple benchmark datasets, we showcase the effectiveness and superiority of EPCRT. Experimental results show that EPCRT exhibits significant performance advantages in handling complex structural variations, sparse and irregular scenes, and generalization to different datasets. Compared to traditional methods, EPCRT can align point clouds more accurately and demonstrate better generalization across different datasets, which is crucial for the accuracy and reliability of cultural heritage protection. Future research directions include further optimizing the performance of the EPCRT method, particularly in enhancing its effectiveness in handling the internal structure of complex cultural heritage data.

Acknowledgements
We thank the National and Local Joint Engineering Research Center for Cultural Heritage Digitization for providing the Terracotta Warriors data.

Author contributions
Yong Wang: Conceptualization, Methodology, Resources, Writing, original draft preparation, Writing, review and editing. Pengbo Zhou: Writing, review and editing, Conceptualization, Visualization. Guohua Geng: Conceptualization, Methodology, Writing, review and editing. Li An: Methodology, Writing, review and editing, Visualization. Mingquan Zhou: Conceptualization, Methodology, original draft preparation.

Funding
This research was funded by the National Natural Science Foundation of China: 62271393. National Science and Technology Support Program：2023YFF0906504. Key Laboratory Project of the Ministry of Culture and Tourism: crrt2021K01, 1222000812. Xi'an Science and Technology Plan Project: 2024JH-CXSF-0014. National key research and development plan: 2020YFC1523301, 2020YFC1523303.

Availability of data and materials
The data will be available upon reasonable request.

Declarations

Ethics approval and consent to participate
Written informed consent has been obtained from the School of Information Science and Technology of Northwest University and all authors for this article, and consent has been obtained for the data used.

Competing interests
The authors declare that they have no competing interests.

Received: 19 June 2024 Accepted: 17 August 2024
Published online: 29 August 2024

References

1. Liu S, Bin Mamat MJ. Application of 3D laser scanning technology for mapping and accuracy assessment of the point cloud model for the great achievement palace heritage building. Herit Sci. 2024;12(1):153.
2. Charatan D, Li SL, Tagliasacchi A, Sitzmann V. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024; p. 19457–19467.
3. Guo Y, Li Y, Ren D, Zhang X, Li J, Pu L. et al. LiDAR-Net: A Real-scanned 3D Point Cloud Dataset for Indoor Scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024. p. 21989–21999.
4. Slimani K, Achard C, Tamadazte B. RoCNet++: triangle-based descriptor for accurate and robust point cloud registration. Pattern Recognit. 2024;147: 110108.
5. Kim J, Kim J, Paik S, Kim H. Point cloud registration considering safety nets during scaffold installation using sensor fusion and deep learning. Autom Constr. 2024;159: 105277.
6. Wu Q, Wang J, Zhang Y, Dong H, Yi C. Accelerating point cloud registration with low overlap using graphs and sparse convolutions. IEEE Trans Multimed. 2023. https://doi.org/10.1109/TMM.2023.3283881.
7. Wu Y, Zhang Y, Ma W, Gong M, Fan X, Zhang M, et al. RORNet: partial-to-partial registration network with reliable overlapping representations. IEEE Trans Neural Netw Learn Syst. 2023. https://doi.org/10.1109/TNNLS.2023.3286943.
8. Arnold E, Mozaffari S, Dianati M. Fast and robust registration of partially overlapping point clouds. IEEE Robot Autom Lett. 2021;7(2):1502–9.
9. Lu F, Chen G, Liu Y, Zhan Y, Li Z, Tao D, et al. Sparse-to-dense matching network for large-scale lidar point cloud registration. IEEE Trans Pattern Anal Mach Intell. 2023;45(9):11270–82. https://doi.org/10.1109/TPAMI.2023.3265531.
10. Qin Z, Yu H, Wang C, Peng Y, Xu K. Deep graph-based spatial consistency for robust non-rigid point cloud registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023; p. 5394–5403.
11. Hu H, Hou Y, Ding Y, Pan G, Chen M, Ge X. V2PNet: voxel-to-point feature propagation and fusion that improves feature representation for point cloud registration. IEEE J Select Top Appl Earth Obs Remote Sens. 2023;16:5077–88. https://doi.org/10.1109/JSTARS.2023.3278830.
12. Wang Y, Zhou P, Geng G, An L, Liu Y. CCAG: end-to-end point cloud registration. IEEE Robot Autom Lett. 2024;9(1):435–42. https://doi.org/10.1109/LRA.2023.3331666.
13. Monji-Azad S, Hesser J, Löw N. A review of non-rigid transformations and learning-based 3D point cloud registration methods. ISPRS J Photogramm Remote Sens. 2023;196:58–72.
14. Liu S, Wang T, Zhang Y, Zhou R, Li L, Dai C, et al. Deep semantic graph matching for large-scale outdoor point cloud registration. IEEE Trans Geosci Remote Sens. 2024;62:1–12. https://doi.org/10.1109/TGRS.2024.3355707.
15. Li X, Liu G, Sun S, Li B, Yi W. Rethinking scene representation: a saliency-driven hierarchical multi-scale resampling for RGB-D scene point cloud in robotic applications. Expert Syst Appl. 2024;243: 122881.

16. Hassani A, Walton S, Li J, Li S, Shi H. Neighborhood attention transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023; p. 6185–6194.
17. Xia Z, Pan X, Song S, Li LE, Huang G. Vision transformer with deformable attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022; p. 4794–4803.
18. Shen Y, Hui L, Jiang H, Xie J, Yang J. Reliable inlier evaluation for unsupervised point cloud registration. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36; 2022; p. 2198–2206.
19. Li J, Shi P, Hu Q, Zhang Y. QGORE: quadratic-time guaranteed outlier removal for point cloud registration. IEEE Trans Pattern Anal Mach Intell. 2023;45(9):11136–51.
20. Vizzo I, Guadagnino T, Mersch B, Wiesmann L, Behley J, Stachniss C. Kiss-icp: in defense of point-to-point icp-simple, accurate, and robust registration if done the right way. IEEE Robot Autom Lett. 2023;8(2):1029–36.
21. Zhu J, Fang B, Chen T, Yang H. Face repairing based on transfer learning method with fewer training samples: application to a terracotta warrior with facial cracks and a Buddha with a broken nose. Herit Sci. 2024;12(1):186.
22. Zhang Z, Sun J, Dai Y, Zhou D, Song X, He M. End-to-end learning the partial permutation matrix for robust 3D point cloud registration. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36; 2022; p. 3399–3407.
23. Tan B, Qin H, Zhang X, Wang Y, Xiang T, Chen B. Using multi-level consistency learning for partial-to-partial point cloud registration. IEEE Trans Vis Comput Graph. 2023. https://doi.org/10.1109/TVCG.2023.3280171.
24. Wang Y, Zhou P, Geng G, An L, Li K, Li R. Neighborhood multi-compound transformer for point cloud registration. IEEE Trans Circ Syst Video Technol. 2024. https://doi.org/10.1109/TCSVT.2024.3383071.
25. Wang Z, Huo X, Chen Z, Zhang J, Sheng L, Xu D. Improving rgb-d point cloud registration by learning multi-scale local linear transformation. Berlin: Springer; 2022. p. 175–91.
26. Zaman A, Yangyu F, Ayub MS, Irfan M, Guoyun L, Shiya L. CMDGAT: knowledge extraction and retention based continual graph attention network for point cloud registration. Expert Syst Appl. 2023;214: 119098.
27. Han J, Shin M, Paik J. Robust point cloud registration using Hough voting-based correspondence outlier rejection. Eng Appl Artif Intell. 2024;133: 107985.
28. Bai X, Luo Z, Zhou L, Fu H, Quan L, Tai CL. D3feat: Joint learning of dense detection and description of 3d local features. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020; p. 6359–6367.
29. Huang S, Gojcic Z, Usvyatsov M, Wieser A, Schindler K. Predator: Registration of 3d point clouds with low overlap. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition; 2021; p. 4267–4276.
30. Zeng A, Song S, Nießner M, Fisher M, Xiao J, Funkhouser T. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017; p. 1802–1811.
31. Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, et al. 3d shapenets: A deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015; p. 1912–1920.
32. Pan L, Cai Z, Liu Z. Robust partial-to-partial point cloud registration in a full range. IEEE Robot Autom Lett. 2024;9(3):2861–8. https://doi.org/10.1109/LRA.2024.3360858.
33. Geiger A, Lenz P, Urtasun R, Are we ready for autonomous driving? the kitti vision benchmark suite. In,. IEEE conference on computer vision and pattern recognition. IEEE. 2012;2012:3354–61.
34. Pan L, Chen X, Cai Z, Zhang J, Zhao H, Yi S. et al. Variational relational point completion network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021; p. 8524–8533.
35. Yew ZJ, Lee GH. Regtr: End-to-end point cloud correspondences with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022; p. 6677–6686.
36. Choy C, Park J, Koltun V. Fully convolutional geometric features. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019; p. 8958–8966.
37. Xu H, Liu S, Wang G, Liu G, Zeng B. Omnet: Learning overlapping mask for partial-to-partial point cloud registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021; p. 3132–3141.
38. Qin Z, Yu H, Wang C, Guo Y, Peng Y, Xu K. Geometric transformer for fast and robust point cloud registration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022; p. 11143–11152.
39. Mei G, Tang H, Huang X, Wang W, Liu J, Zhang J, et al. Unsupervised Deep Probabilistic Approach for Partial Point Cloud Registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023; p. 13611–13620.
40. Wang H, Liu Y, Hu Q, Wang B, Chen J, Dong Z, et al. Roreg: pairwise point cloud registration with oriented descriptors and local rotations. IEEE Trans Pattern Anal Mach Intell. 2023;45(8):10376–93.
41. Zhang X, Yang J, Zhang S, Zhang Y. 3D Registration with Maximal Cliques. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023; p. 17745–17754.
42. Yu H, Hou J, Qin Z, Saleh M, Shugurov I, Wang K, et al. RIGA: rotation-invariant and globally-aware descriptors for point cloud registration. IEEE Trans Pattern Anal Mach Intell. 2024. https://doi.org/10.1109/TPAMI.2023.3349199.
43. Aoki Y, Goforth H, Srivatsan RA, Lucey S. Pointnetlk: Robust & efficient point cloud registration using pointnet. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019; p. 7163–7172.
44. Wang Y, Solomon JM. Deep closest point: Learning representations for point cloud registration. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019; p. 3523–3532.
45. Yew ZJ, Lee GH. Rpm-net: Robust point matching using learned features. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020; p. 11824–11833.
46. Xie Y, Zhu J, Li S, Hu N, Shi P. HECPG: hyperbolic embedding and confident patch-guided network for point cloud matching. IEEE Trans Geosci Remote Sens. 2024;62:1–12. https://doi.org/10.1109/TGRS.2024.3370591.
47. Li J, Zhang C, Xu Z, Zhou H, Zhang C. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration. Berlin: Springer; 2020. p. 378–94.
48. Qiu C, Wang Z, Lin X, Zang Y, Wang C, Liu W. DSMNet: deep high-precision 3D surface modeling from sparse point cloud frames. IEEE Geosci Remote Sens Lett. 2023. https://doi.org/10.1109/LGRS.2023.3306940.
49. Ao S, Hu Q, Yang B, Markham A, Guo Y. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021; p. 11753–11762.
50. Lu F, Chen G, Liu Y, Zhang L, Qu S, Liu S, et al. Hregnet: A hierarchical network for large-scale outdoor lidar point cloud registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021; p. 16014–16023.
51. Poiesi F, Boscaini D. Learning general and distinctive 3D local deep descriptors for point cloud registration. IEEE Trans Pattern Anal Mach Intell. 2022;45(3):3979–85.
52. Chen Z, Sun K, Yang F, Guo L, Tao W. $SC^2$2-PCR++: rethinking the generation and selection for efficient and robust point cloud registration. IEEE Trans Pattern Anal Mach Intell. 2023;45(10):12358–76.

Publisher's Note