An Unsupervised Learning Network for Large-scale LiDAR Point Clouds Registration

Jingbin Liu, Xuanfan Lv, Xiaodong Gong, Yifan Liang, and Juha Hyyppä

***Abstract*—With the continuous development of autonomous driving technologies, the registration of outdoor large-scale LiDAR point clouds has become increasingly important. Unlike indoor small-scale object point clouds, large-scale point clouds have inherent sparsity, abundant outliers, and other limitations. These characteristics often lead to low alignment accuracy and high time consumption when applying existing methods to large-scale point cloud registration. To address these issues, we propose an improved point cloud keypoints extracting method based on rotation compensation and a convolutional end-to-end unsupervised point cloud registration network. The former enables reliable keypoints extraction. The latter further extracts global features from the keypoint point clouds obtained by the former method and learns the overlapping region information between the source and target point clouds using a spatial attention weight encoder, and it can be trained efficiently without pose ground truth. To ensure fast and effective convergence of the network, we introduce a chamfer distance loss based on dynamic overlap rates. We test our method on two outdoor large-scale LIDAR point cloud datasets: PandaSet and KITTI odometry datasets. The results demonstrate excellent and stable performance, when it is applied to either original consecutive frames or the case of simulating large angular variations in real-world scenarios between consecutive frames by randomly transforming the target frame. Moreover, by applying our method's registration results as initial values to the classic ICP, we not only achieve optimal accuracy and robustness but also significantly accelerate the convergence of ICP, enhancing the efficiency of precise registration.**

***Index Terms*—LIDAR point cloud registration, Fast and robust registration, Unsupervised deep learning, Large-scale scene, Autonomous driving positioning, Global localization.**

This work was supported in part by National Key Research Development Program of China with project No. 2021YFB2501102, Wuhan knowledge innovation research program (2022010801010109), Shenzhen Science and Technology Program (JCYJ20210324123611032), Wuhan AI innovation research programme (2023010402040029). *(Corresponding author: Xuanfan Lv.)*

Jingbin Liu is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China, and also with the Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Espoo 02150, Finland (e-mail: jingbin.liu@whu.edu.cn).

Xuanfan Lv, Xiaodong Gong, and Yifan Liang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: lxf0626@whu.edu.cn; gongxiaodong@whu.edu.cn; lyf0312@whu.edu.cn).

Juha Hyyppä is with the Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Espoo 02150, Finland (e-mail: juha.hyyppa@nls.fi).

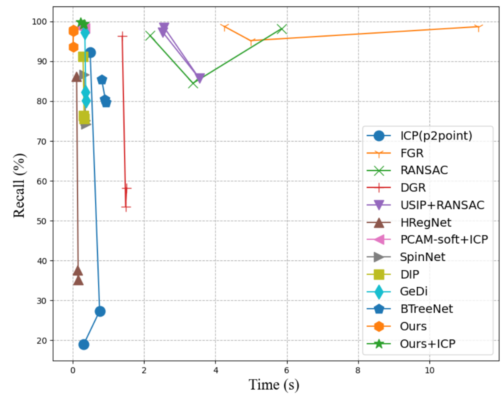


Fig. 1 Comparison of different registration methods based on their computation time and registration recall

I. INTRODUCTION

WITH the maturity and popularity of 3D point cloud acquisition devices such as LiDAR, as well as the general improvement in computer processing power, point clouds have become the primary data format for representing the 3D world and have been widely used. Due to the limited field of view of sensors during the process of scanning the 3D world to obtain point cloud data, registration algorithms are needed to stitch together local point clouds to generate complete 3D scenes. Point cloud registration essentially involves estimating the Euclidean transformation matrix between two frames of scanned point clouds. After obtaining the transformation matrix, point clouds captured in different camera or LiDAR coordinate systems can be converted to the world coordinate system and combined. Point cloud registration is of great significance in fields such as 3D reconstruction [1], localization, and is specifically applied in scenarios such as simultaneous localization and mapping (SLAM) for mobile robots [2], [3], [4] and high-precision map construction for autonomous driving [5], [6], [7], [8], [9].

As the deep learning achieved great development, many learning-based methods [10], [11], [12], [13], [14] have replaced the classical hand-crafted features [15], [16], [17] to perform point cloud registration tasks more quickly and accurately. However, existing point cloud registration methods have been mostly applied in the stitching of small-scale 3D point clouds, such as indoor or object-level point clouds. When in large-scale point cloud scenes, due to the range errors and noise in the LIDAR scan data, as well as the sparsity of the point cloud, the performance of registration algorithms designed for small-scale scenes are often unsatisfactory, which may bring time-consuming and registration failure problem. In recent years, laser SLAM technologies, with LOAM [18] as a representative example, have been widely applied in fields such as autonomous driving and robot navigation. And registration techniques for large-scale outdoor point clouds are the important component of laser SLAM.

To address the problems mentioned above about the large-scale outdoor point cloud registration, we take inspiration from laser SLAM keypoints extracting algorithms. It is well known that LOAM creatively extracts edge and plane keypoints from large scene scanned point clouds, and it estimates motion by optimizing the distances between edge keypoints and edge lines, as well as between plane keypoints and planes. In this paper, we first proposed an improved keypoints extracting method based on rotation compensation, which can help extract corner and surf keypoints from large-scale outdoor point cloud more robustly. Then, we proposed a convolutional end-to-end unsupervised registration network including a weighted chamfer distance loss based on dynamic overlap ratio, which can help the total registration network train efficiently without the availability of pose ground truth label data. To validate the performance of our network, we performed the experiment on two representative large-scale outdoor LiDAR point cloud datasets, namely KITTI odometry dataset [19] and PandaSet dataset [20]. The results demonstrate that the proposed method performs well in terms of efficiency and in dealing with large rotation variation, and it can provide good initial values for refined registration algorithms, thus surpassing state-of-the-art methods in terms of precision. In summary, our main contributions are threefold:

- We proposed an improved keypoints extracting method based on rotation compensation to address the issue of inaccurate or even failed keypoints extraction when there is an inclination angle between the point cloud frame and the sensor's XOY plane. This method utilizes the idea of point cloud plane segmentation to extract the reference plane of the point cloud frame and applies rotation compensation to the original point cloud, thereby improving the robustness of keypoints extraction.
- Using the robust keypoints extraction, a convolutional end-to-end unsupervised registration network is proposed for large-scale point cloud scenes. It utilizes a Siamese neural network structure and employs a PointNet-like module to extract global features of consecutive point cloud frames' keypoints, which are extracted with the proposed keypoints extracting method based on rotation compensation. And then an attention weight mechanism is designed to enable the network to adaptively learn the overlapping regions of consecutive point cloud frames, thereby achieving fast and effective end-to-end registration of large-scale point cloud scenes.
- Within the proposed convolutional end-to-end unsupervised registration network, a dynamic overlap ratio weighted chamfer distance loss is proposed based on the estimated overlapping regions of large-scale point clouds, so that the registration network is able to converge more quickly and accurately.

II. Related Work

A. Classical Point Cloud Registration

Traditional point cloud registration methods can be roughly divided into optimization-based methods and handcrafted feature-based methods. Among them, the Iterative Closest Point (ICP) [21] is the most classic optimization-based point cloud registration method. Given source point cloud P and target point cloud Q, first obtain the correspondence between points using the nearest neighbor rule. For example, the point p_i in P corresponds to the nearest neighbor point q_i in Q. The initial transformation matrix T is usually set to the identity matrix I. Then, iteratively optimize the distance error to estimate the transformation matrix T. The iteration will be terminated when the required error is below a threshold or the maximum number of iterations is reached. However, ICP is not a global registration algorithm. Its accuracy relies on the quality of the initial registration guess, and it often gets trapped in local optima. Many ICP variants [22], [23], [24] have been proposed to handle the problems existed in ICP. Go-ICP (Generalized-Iterative Closest Point) [24] utilizes the branch-and-bound method in the SE(3) space. It subdivides the initial space into smaller subspaces using an octree data structure and eliminates unfavorable subspaces through the branch-and-bound technique. It continues subdividing the subspaces that meet the threshold conditions to find the globally optimal transformation. While this method addresses the issue of local minima, it remains sensitive to initialization. On the other hand, GICP [23] combines the iterative closest point (ICP) algorithm and the point-to-plane ICP algorithm into a probabilistic framework, where both point-to-point and point-to-plane methods become special cases. This method establishes Gaussian distributions on pairs of points. Consequently, for any rigid transformation T, the distances between the corresponding points follow a Gaussian distribution. The optimal transformation matrix is then determined as the one that maximizes the Gaussian probability of the distances between the corresponding points after the transformation. MVGICP [22] calculates the mean and variance within voxels at different scales, from large to small, during iterations. It incorporates these values into the GICP model and utilizes the Gauss-Newton method to obtain the transformation matrix. The iteration continues with smaller voxels. Larger voxels allow for a more global coarse registration of point clouds, while smaller voxels further improve the accuracy of the registration results. Additionally, the method eliminates the need for nearest neighbor search, significantly enhancing computational efficiency. The Normal Distributions Transform (NDT) [25] is another classic optimization-based point cloud registration method. It characterizes point cloud data using mathematical distribution properties. By voxelizing the target point cloud and calculating the mean and variance of points within each voxel, a probability density function describing the distribution of points within the voxel can be obtained. The optimal transformation matrix is the one that maximizes the overall likelihood of the source point cloud within its corresponding probability density function.

Methods based on handcrafted features [26], [27], [28], [29] do not utilize all the point cloud data, which may contain outlier points and noise. These methods extract individual or combined effective features from the point cloud, such as points, lines, surfaces, normal vectors, and curvature. They then achieve fast correspondence matching based on custom feature descriptors. Once the point-to-point correspondences are determined, the transformation matrix can be calculated without the need for iterative methods using the RANSAC algorithm [30]. Despite the widespread application of these classical registration methods, they still face challenges such as long computation time and poor robustness when dealing with large-scale LIDAR point cloud registration problems.

B. Learning-based Point Cloud Registration

The field of learning-based point cloud registration encompasses two main divisions: feature-learning-based and end-to-end-based methods. Feature-learning-based approaches focus on leveraging deep features to estimate accurate correspondences, allowing for one-step optimization (e.g., SVD or RANSAC) to estimate the transformation without iterative processes [31]. For example, PPFNet [10] calculates point pair features (PPF) from the local neighborhoods of sampled points and uses them as inputs to the network. It employs multiple PointNet [32] networks to fuse local features at different scales and global features, which are then encoded using MLPs to obtain the final features. PPFNet leverages global context awareness and feature encoding to enhance rotation invariance and robustness against noise. However, calculating PPF features requires a significant amount of nearest neighbor annotation data, and the establishment of local reference frames relies on estimated normal vectors, making it sensitive to noise. On the other hand, end-to-end-learning-based methods employ neural networks to directly transform two input point clouds into a corresponding transformation matrix [31]. End-to-end networks integrate processing modules of various steps into a single network, which requires a large amount of memory and is more suitable for registration tasks with small datasets. For example, DCP (Deep Closest Point) [12] uses DGCNN (Dynamic Graph Convolutional Neural Network) [33] to learn an embedding module that maps input point clouds to a high-dimensional space. It utilizes a transformer module to encode the contextual information of the point cloud and outputs the predicted rigid transformation matrix using a differentiable SVD (Singular Value Decomposition) layer.

C. Registration for Large-scale Outdoor LIDAR Point Clouds

For large-scale outdoor LiDAR point cloud registration tasks, there are several challenges that need to be addressed. These challenges include handling outliers, noise, and distortion in individual point clouds, dealing with low overlap between consecutive frames, and coping with the high computational costs due to the large volume of point cloud data. Most existing learning-based registration methods are designed for small-scale indoor point clouds or object point clouds. However, in recent years, with the continuous development of technologies such as autonomous driving, deep learning methods for large-scale outdoor LiDAR point cloud registration tasks have also emerged. For example, DeepVCP [34] implements a deep virtual correspondences point method. The network first uses PointNet++ [35] to extract semantic features of the points. USIP [36] and RSKDD-Net [37] are two different point cloud keypoint detectors that can be used for registration tasks. USIP leverages a feature proposal network to learn stable keypoints from the input 3D point cloud and its transformed counterpart. It proposes a probabilistic chamfer loss to optimize the distances between keypoints of the input point cloud pairs. RSKDD-Net utilizes a random sampling extension group strategy to expand the receptive field of each sampled point for clustering neighboring points. Then, an attention mechanism is used to aggregate the positions and features of the neighboring points to obtain the keypoints. DDRNet [38] and HRegNet [39] are network architectures that have emerged in the past two years specifically for large-scale scene point cloud registration tasks. DDRNet utilizes a local-spatially aware encoder to gather posture information comprising local and spatial features. It also incorporates an attentional weighting module, enabling the network to adaptively prioritize overlapping areas. HRegNet conducts registration on hierarchically extracted keypoints and descriptors. By combining dependable features from deeper layers with accurate position information from shallower layers, the framework achieves registration that is both robust and precise. However, both of the above methods require the use of ground truth trajectory data to train the network in a supervised manner, which limits their application in real-world scenarios.

III. Methodology

In this section, we proposed an improved keypoints extracting method based on rotation compensation and a convolutional end-to-end unsupervised registration network for large-scale laser point cloud registration problems. The proposed approach, as illustrated in Fig. 1, extracts keypoints from the original point cloud using the improved keypoints extracting method, and extracts global features from the source keypoints cloud denoted as P_k^S and the target keypoints cloud denoted as P_k^T in the registration network. Then, an attentional weighting module and a symmetric MLPs encoder are used to predict the pose $T_{S \to T} \in SE(3)$ for point cloud registration. The pose is obtained by optimizing the weighted chamfer distance based on dynamic overlap ratio between the original non-ground point clouds denoted as P_q^S and P_q^T. We describe the detail of our network loss function in the last part of this section.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3417415

A. An Improved Keypoints Extracting Method Based on Rotation Compensation

As for large-scale LIDAR point cloud, directly extracting features from the original point cloud using structures similar to PointNet [32] is unreliable and can result in loss of local information due to the sparsity of the original point cloud. Existing methods often choose to sample the original point cloud [35], [36]. After obtaining the sampled points, they extract corresponding local regions of the point cloud by finding neighboring points for each sampled point and recover global features from all the local geometric batches. Although this method of extracting global features from local geometric patches of the point cloud is effective, it has certain issues:

- Low computational efficiency: Existing approaches mostly use FPS (Farthest Point Sampling) algorithm as the sampling method, which has a time complexity of $O(N^2)$.
- Increased network complexity: Extracting multi-scale features from each local geometric batches requires using multiple layers of PointNet extractors, which increases the overall complexity of the network.

In order to address the mentioned issues, we draw inspiration from the keypoints extracting method in LOAM [18] and directly extracts global features from the keypoints cloud P_k. However, the original surf points and corner points extracting method in LOAM fails or produces inaccurate results when there is an inclination between the point cloud frame and the sensor's XOY plane ϕ_{XOY}. To tackle this, we proposed an improved keypoints extracting method based on rotation compensation. The overall algorithm workflow is illustrated in Fig. 2. First, the ground reference plane ϕ_R is extracted from the original point cloud P_i using a plane segmentation algorithm based on RANSAC, which is shown in Algorithm. 1.

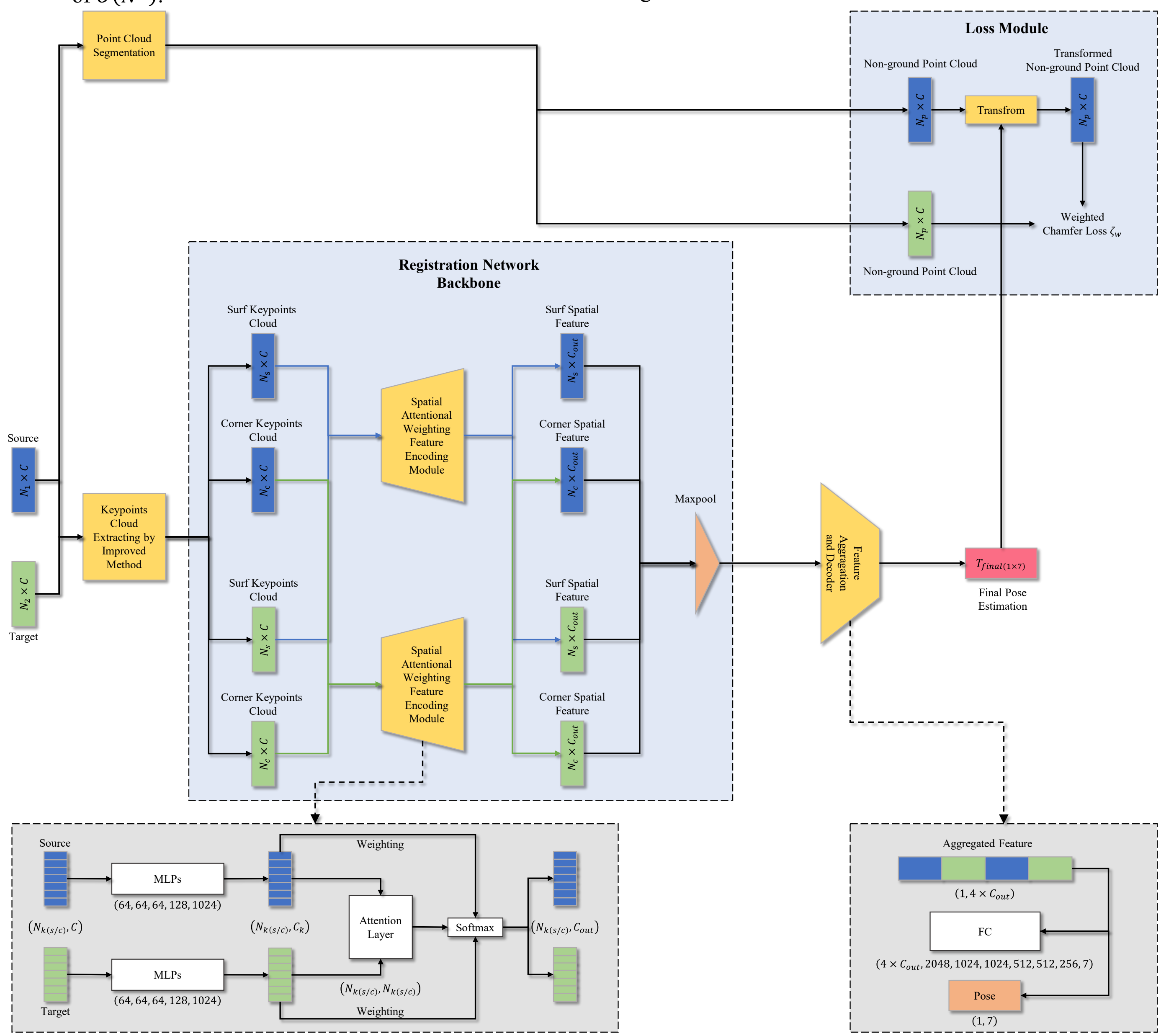


Fig. 2 Architecture of proposed registration network

Algorithm. 1 Point Cloud Plane Segmentation

Input: Initial point cloud: $P^i_{N\times 3}$
Down leafsize: L_s
Output: The inplane point cloud: $P^n_{N_n\times 3}$
The outplane point cloud: $P^m_{N_m\times 3}$
The segment plane coefficients: ξ^i
// Perform voxel downsampling on the original point cloud.
1: $P^d_{N_d\times 3}$ = VoxelGridDownsample ($P^i_{N\times 3}$);
// Random sampling inliers to estimate the model of segment plane.
2: $Q^s_{N_s\times 3}$ = RandomSample ($P^d_{N_d\times 3}$);
// Calculate the initial coefficients of segment plane.
3: ξ^i_s = Model ();
// Substitute all points from $P^d_{N_d\times 3}$ into the model and count the number N_i of inliers.
4: N_i = CalInliers ($P^d_{N_d\times 3}$)
// When the change in the number of inliers exceeds the threshold ε, iterate and update the optimal plane model with the current inliers, and update the number of inliers.
5: While ($\Delta N_i > \varepsilon$) do
6: ξ^i_s = Model ();
7: Optimization (N_i = CalInliersNum ($P^d_{N_d\times 3}$));
8: $\xi^i = \xi^i_s$;
// Substitute ξ^i to calculate $P^n_{N_n\times 3}$ and $P^m_{N_m\times 3}$
9: $P^n_{N_n\times 3}$, $P^m_{N_m\times 3}$ = ModelApply ();
10: return $P^n_{N_n\times 3}$, $P^m_{N_m\times 3}$, ξ^i

Then, the angle θ_N between the normal vector $\vec{n}_R$ of ϕ_R and the normal vector $\vec{n}_{XOY}$ of ϕ_{XOY} is calculated, along with the cross-product vector $\vec{n}_M$. This allows for the computation of the compensation matrix $T_{\phi_R\to\phi_{XOY}} \in SE(3)$, which represents the transformation from ϕ_R to ϕ_{XOY}. The point cloud P^i is then transformed by applying the compensation transformation to obtain $P^{i'}$, and the keypoints cloud $P^{i'}_k$ is extracted from $P^{i'}$. Finally, the inverse transformation $T'_{\phi_R\to\phi_{XOY}}$ is applied to the transformed $P^{i'}_k$ to obtain the final keypoints cloud P^i_k. Our approach not only reduces the time complexity of feature point sampling to $O(N)$ but also decreases the overall complexity of the network, improving training efficiency.

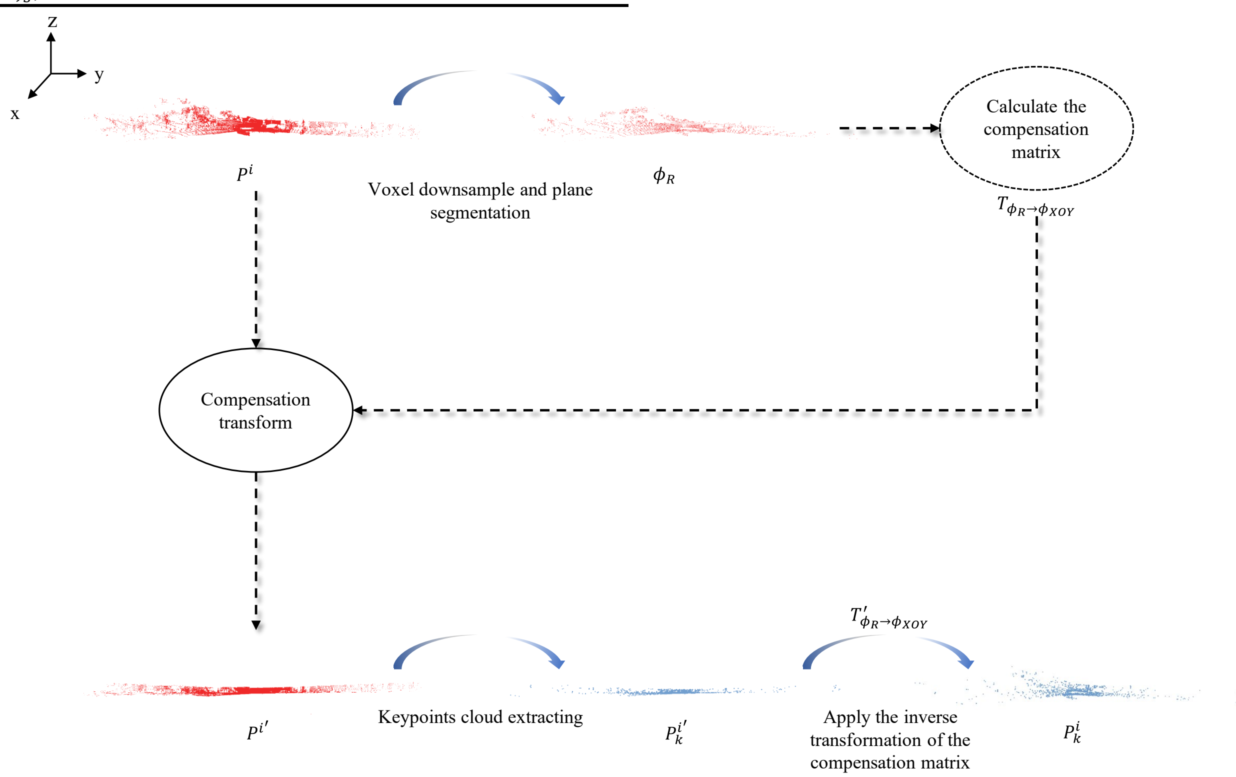


Fig. 3 The workflow of keypoints cloud extracting method based on rotation compensation

B. Convolutional End-to-end Unsupervised Registration Network

The convolutional end-to-end unsupervised registration network for large-scale laser point clouds, illustrated in Fig. 1, can be decomposed into two modules: a feature encoder module with spatial attentional weighting mechanism and a Siamese architecture decoder. As for the loss function, we proposed a weighted chamfer distance loss based on dynamic overlap ratio.

1) Spatial attentional weighting feature encoder

After extracting keypoints from the source point cloud P^s and the target point cloud P^t, the corresponding surf point cloud P_s and corner point cloud P_c can be separately extracted. Given that LOAM estimates motion by optimizing the distances between corner points and edge lines as well as between surf points and planes, we choose to extract features from the P_s and P_c of both the source P^s and target P^t using a spatial attention weight encoder with a Siamese-like architecture.

The encoder consists of two parts: a multi-layered perceptions (MLPs) and an attention layer. The MLPs is similar to the PointNet architecture having size 64, 64, 64, 128, 1024. Both the $P^s_{k(s/c)}$ (surf or corner points cloud from source P^s) and $P^t_{k(s/c)}$ (surf or corner points cloud from target P^t) are input to the MLPs, then we can get the corresponding keypoints global features as Eq. 1.

$$F^{g^{s/t}}_{k(s/c)} = MLPs_{[3,64,64,64,128,1024]}\left(P^{s/t}_{k(s/c)}\right) \tag{1}$$

After the keypoints global feature embedding, We designed an attention layer that adaptively learns the correlations between features, allowing the network to focus more on the similar regions between the keypoint point clouds. For the source P^s and the target P^t, their corresponding extracted surf point cloud and corner point cloud already have a good level of matching. However, the inclusion of the attention layer further alleviates the problem of mismatch caused by the sparsity of point clouds. We used a linear function to a query vector and key vector denoted as Q and K. Then we computed the attention scores S by Eq. 2.

$$S_{ij} = \left[\frac{Q_i \times K_j}{row(Q)}\right]_{i\in N_{k(s/c)}, j\in N_{k(s/c)}} \tag{2}$$

where $row(Q)$ denotes the size of query vector Q, $N_{k(s/c)}$ represents the size of keypoints cloud, and both the Q and K are learnable parameters which make the network put more attention to the similar regions of keypoints cloud. In order to get the attention weight matrix denoted as W, we used the softmax function to normalize attention scores S as Eq. 3.

$$W = softmax(S) \tag{3}$$

Finally, we used the attention weight matrix to modify the features computed by $MLPs$, formulated as Eq. 4.

$$\begin{cases} F^{g^s}_{k(s/c)} = F^{g^s}_{k(s/c)} + \left(W \times F^{g^t}_{k(s/c)}\right) \\ F^{g^t}_{k(s/c)} = F^{g^t}_{k(s/c)} + \left(W \times F^{g^s}_{k(s/c)}\right) \end{cases} \tag{4}$$

where the residual term $W \times F^{g^{s/t}}_{k(s/c)}$ refers to the prior learned weighting region between the source keypoints cloud and the target keypoints cloud.

2) Siamese architecture decoder

For the purpose of aggregating all the keypoints global features, we chose to put them to a symmetric max-pooling function and make them concatenated, and then we got the aggregated global features. In order to predict the final pose, we designed the FC layers, which has six hidden layers, 2048, 1024, 1024, 512, 512, 256, and an output layer of size 7 whose parameters represent the predicted pose $T_{final(1\times7)}$. The first four values of $T_{final(1\times7)}$ denote the rotation quaternion $q \in \mathbb{R}^4, q^Tq = 1$, and last three denote the translation vector $t \in \mathbb{R}^3$.

3) Dynamic overlap ratio weighted chamfer distance loss

For supervised learning, the accuracy of registration network predictions can be optimized by setting the loss function as the L2 norm of the difference between the predicted pose and the ground truth. However, in real-world scenarios, acquiring accurate ground truth for network training is often challenging and complex. Therefore, for our unsupervised network, we have chosen the chamfer distance used for quality evaluation of point cloud reconstruction as the loss function. It aims to minimize the distance between all matched points in the non-ground point clouds denoted as P^S_q and P^T_q from the source P^s and target P^t, as the ground points represent the majority of unstable keypoints [40]. To ensure more robust and faster convergence during backpropagation optimization, we multiply the traditional chamfer distance loss by a correction coefficient based on the dynamic overlap ratio of the point clouds, formulated as Eq. 5.

$$\begin{cases} Loss\left(P^{S_{trans}}_q, P^t_q\right) = \Upsilon \cdot min_{\varphi: P^{s_{trans}}_q \to P^t_q} \dfrac{1}{size\left(P^{S_{trans}}_q\right)} \sum_{x\in P^{S_{trans}}_q} \|x-\varphi(x)\|^2 \\ \Upsilon = \dfrac{size\left(P^{S_{trans}}_q\right)}{num_{x\in P^{s_{trans}}_q}\left(x < mean(\|x-\varphi(x)\|^2) + std(\|x-\varphi(x)\|^2)\right)} \end{cases} \tag{5}$$

Where $P^{S_{trans}}_q$ is the non-ground cloud from source P^s and transformed by the predicted pose $T_{final(1\times7)}$. The φ function finds the corresponding points between $P^{S_{trans}}_q$ and P^T_q based on nearest neighbor point rule. And the coefficient γ represents the reciprocal of the ratio between the number of nearest neighbor matching distances that are smaller than the mean of the nearest neighbor matching distances plus the standard error and the total number of keypoints.

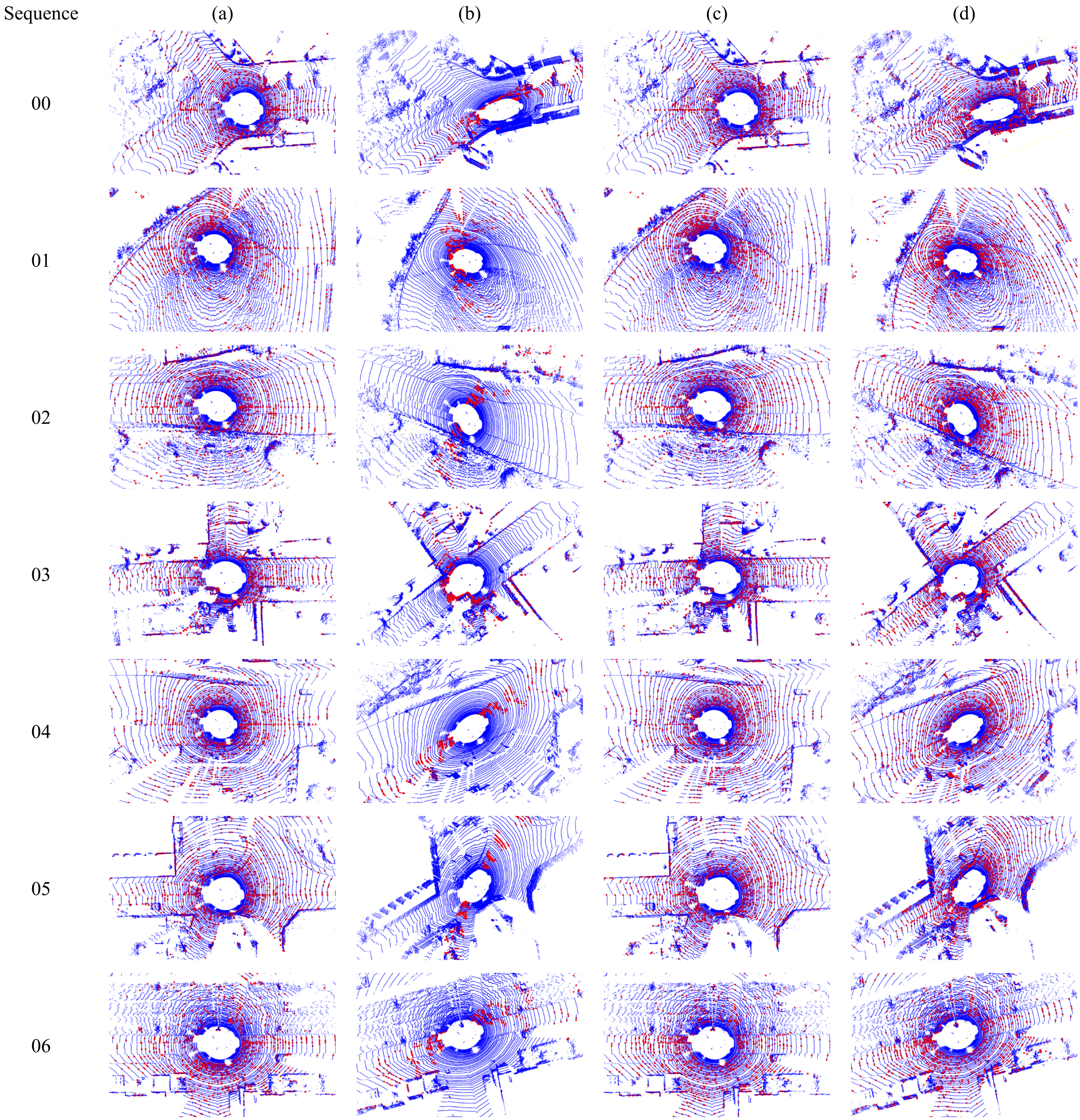
IV. Experimental Results and Analysis

In this section, we used the KITTI [19] and PandaSet [20] large-scale outdoor datasets as the benchmark to evaluate our improved keypoints extracting method and the proposed convolutional end-to-end unsupervised registration network. Firstly, we evaluated the effectiveness and robustness of our improved keypoints extracting method separately in scenarios where the point cloud frame had and didn't have an inclination with respect to the sensor's XOY plane. Secondly, we compared the performance of the proposed registration network with existing state-of-the-art methods in terms of registration accuracy, efficiency, and recall rate.

A. Evaluation of Improved Keypoints Extracting Method

Due to the more prominent keypoints in the KITTI dataset's LiDAR point cloud, we chose to evaluate the improved keypoints extracting method based on the KITTI dataset. The KITTI odometry dataset is the one of the most widely used public dataset in the field of autonomous driving. It consists of various modalities, including calibrated and synchronized images, Velodyne HDL-64E LiDAR scans, high-precision GPS information, and IMU acceleration data. The dataset contains 22 sequences covering urban streets, highways, and other scenarios. Among these sequences, sequences 00-10, totaling 11 sequences, come with ground truth pose, which can be used for training, validation, and testing of registration networks.

To verify the effectiveness and robustness of the improved keypoints extracting method, we randomly selected one point cloud frame from 00-10 KITTI sequences. We applied a random rotation transformation to the selected frames, including pitch, roll, and yaw angles within the range of -45 to 45 degrees. Then, we compared the results of the improved keypoints extracting method with the results obtained using the LOAM's method of extracting surf points and corner points, visualizing the keypoints for comparison.



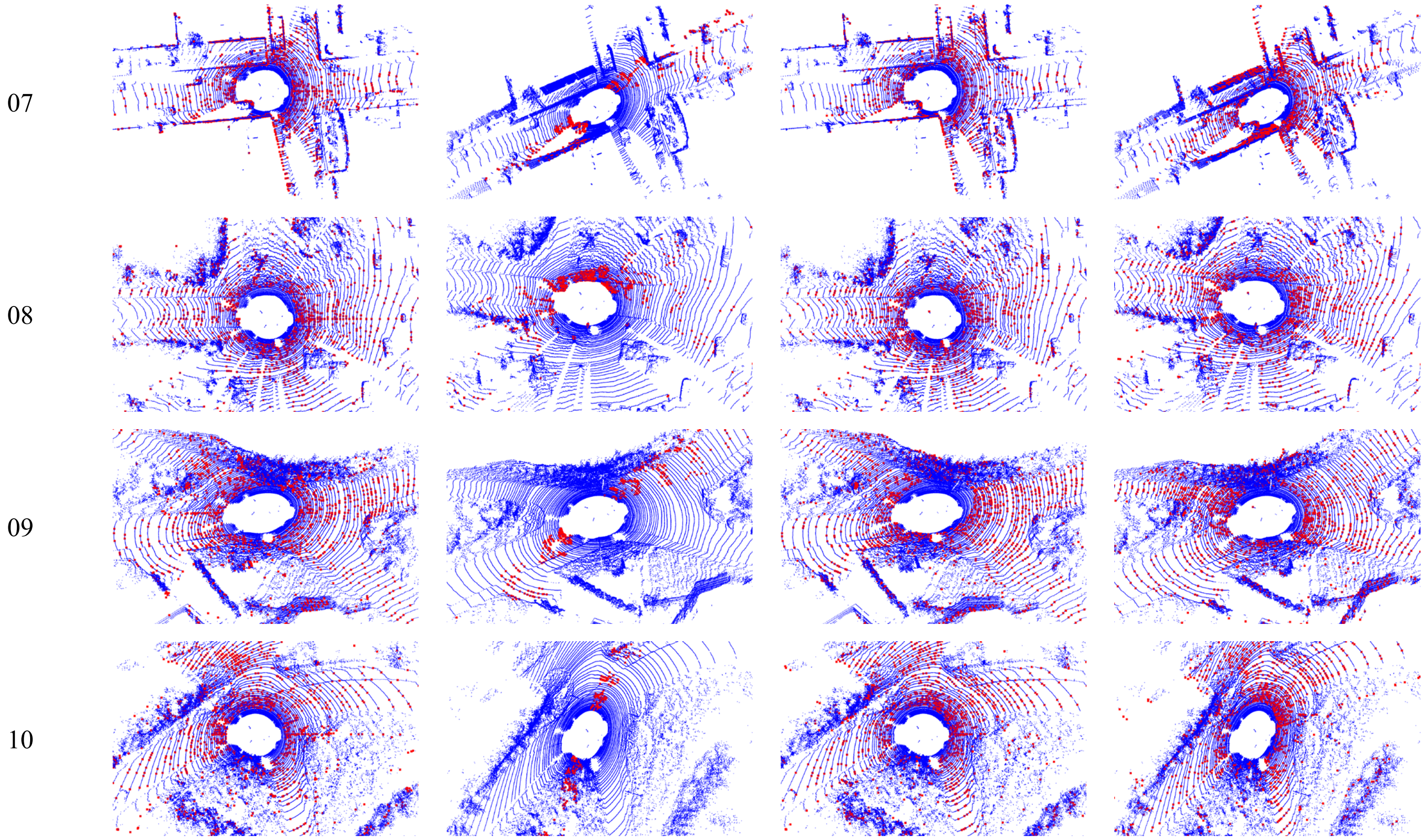


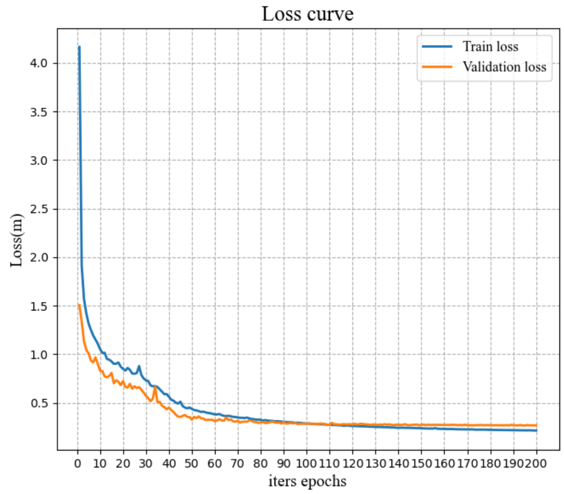
Fig. 4 (a) Visualization results of keypoints extracting on initial point cloud frames using the LOAM method. (b) Visualization results of keypoints extracting on point cloud frames that have been applied large-angle simulated transformations using the LOAM method. (c) Visualization results of keypoints extracting on initial point cloud frames using the improved method. (d) Visualization results of keypoints extracting on point cloud frames that have been applied large-angle simulated transformations using the improved method

As illustrated in Fig. 3, when there is no significant angular distortion between the point cloud frame and the sensor's XOY plane, the results (a) (c) of the improved keypoints extracting method and the traditional LOAM keypoints extracting method are basically consistent, indicating the effectiveness of the improved method. However, when facing large angular distortion between the laser point cloud frame and the sensor's coordinate system, the keypoints extracting results (b) using LOAM method are inaccurate and sparse compared to (a), while the improved method can still extract globally valid keypoints shown as (d), and this indicates its good robustness.

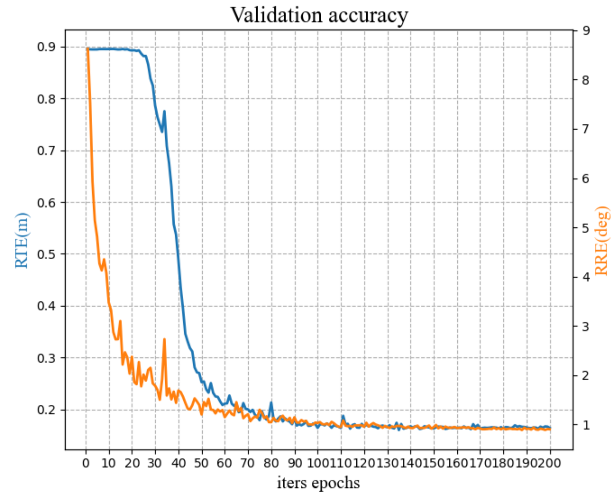
B. Registration Network Experiment Settings

In this section, to evaluate the registration performance of our network when dealing with large-scale LIDAR point clouds, we utilized 11 KITTI odometry sequences and 103 PandaSet sequences with ground truth. PandaSet datasets is an open-source dataset for L5 level autonomous driving, which includes 103 sequences, with each sequence representing data from different scenarios. Each sequence in PandaSet datasets contains 80 consecutive frames of LIDAR point clouds, images, and semantic segmentation annotation data. As for KITTI datasets, sequences 0-7 were used for training, sequence 8 for validation, and sequences 9-10 for testing, while as for PandaSet datasets, sequences 0-72 were used for training, sequence 73-87 for validation, and sequences 88-102 for testing. The ground truth of both two datasets were only used to evaluate the accuracy of the registration results on the final test set and were not used as label data for network training. We selected consecutive frames from LIDAR scans as samples (source and target). For each frame, after extracting surf point clouds and corner point clouds, we sampled them with zero padding to 1280 and 640 points respectively. As for the non-planar point clouds extracted using plane segmentation algorithms, we sampled them with random sampling to 8192 points. For each sample, we applied rigid transformations to the preprocessed target frame (for each target frame, we pre-align it to the coordinate system of the source frame using a global registration algorithm, ensuring that all samples to be data-augmented have a consistent scale.), including yaw, pitch, and roll rotation angles in the range of 0-45°, and translation along the three axes in the range of -1.0 m-1.0 m to achieve data augmentation. The training of the entire network was completed on four NVIDIA 1080ti GPUs, we used the Adam optimizer with an initial learning rate of 0.0001 and set the decay rate to 0.99 for exponential decay. The network was trained for a total of 200 epochs, which is shown in Fig. 4.

We compared the performance of our network with both classical registration methods and state-of-the-art learning-based registration methods. The classical methods include ICP (point to point and point to plane) [21], [23], RANSAC [30], FGR (Fast Global Registration) [41] and Go-ICP [24]. We implemented the first three methods using the Open3D library in Python, and we utilized the open-source library to implement Go-ICP. As for learning-based methods, we chose four representative one to compare: DGR (Deep Global Registration) [42], USIP [36], HRegNet [39], PCAM [43], SpinNet [44], DIP [45], GeDi [46] and BTreeNet [47]. We performed voxel downsampling with a voxel size of 0.1m on all point clouds used for testing, and randomly sampled them to 8192 points. All the aforementioned comparative methods and our network were implemented on the same platform and hardware environment.



(a)



(b)

Fig. 5 (a) The training and validation's loss curve. (b) The validation accuracy decrease during training

C. Registration Experiments on KITTI Odometry Dataset

1) Quantitative evaluation

To evaluate the registration performance of our network, we selected three evaluation metrics: registration accuracy, efficiency, and recall rate. The accuracy evaluation includes the relative translation error (RTE) and the relative rotation error (RRE). RTE and RRE can be calculated by Eq. 6 and Eq. 7, where t' and R' are the estimated results, and t and R are the translation and rotation matrices corresponding to the ground truth.

$$\Gamma_{RTE} = \|t - t'\|_2 \qquad (6)$$

$$\Gamma_{RRE} = arccos(Tr(R'R^T - 1)/2) \qquad (7)$$

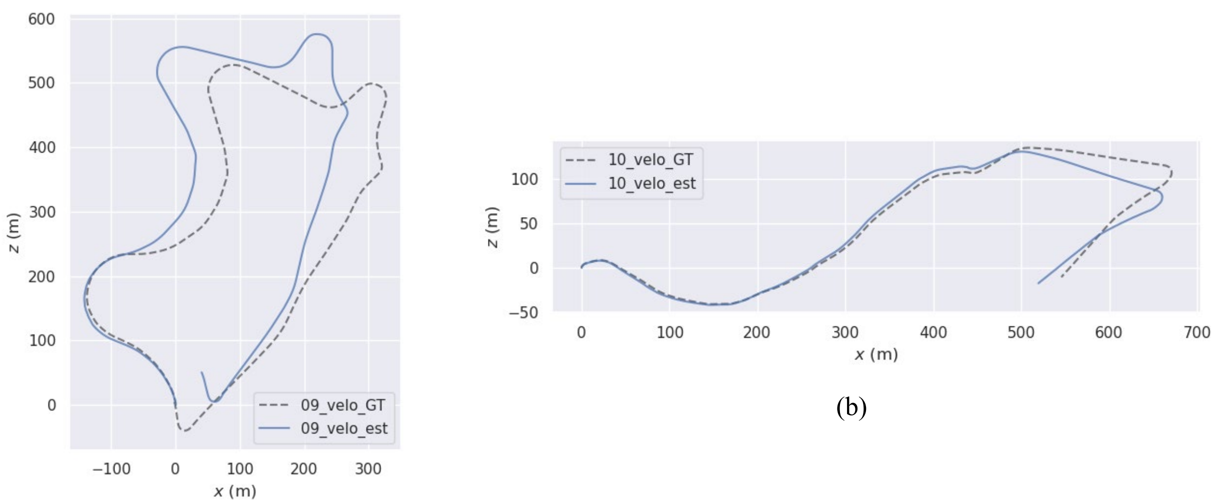
The evaluation of registration efficiency can be achieved by comparing the time taken for registration of individual consecutive frames. The registration recall rate represents the success rate of the registration process. When both RTE and RRE are within a certain threshold ε_{RTE} and ε_{RRE}, the registration is considered effective. A higher recall rate indicates a stronger robustness of the registration method. In this evaluation, we set $\varepsilon_{RTE} = 1m$ and $\varepsilon_{RRE} = 5deg$ respectively. In Tab. 1, we list the experimental results of the comparative methods and our network, including the three metrics mentioned above, to help us conduct a more detailed analysis of the comparative results.

Tab. 1 Registration performance on KITTI dataset

Methods	RTE (m)		RRE (deg)		Time (s)	Recall
	Mean	Std	Mean	Std		
ICP (p2point) [21]	0.258	0.374	0.169	0.388	0.491	92.3%
ICP (p2plane) [23]	0.253	0.361	0.133	0.371	0.527	94.2%
Go-ICP [24]	0.916	1.159	2.298	2.416	40.049	52.8%
FGR [41]	0.121	0.459	0.219	**0.239**	4.241	98.8%
RANSAC [30]	0.168	0.886	0.759	5.735	2.163	96.5%
DGR [42]	0.358	0.316	0.415	0.448	1.389	96.3%
USIP+RANSAC [30], [36]	0.152	0.186	0.586	0.405	2.526	97.3%
HRegNet [39]	0.609	0.252	1.034	0.573	0.112	86.2%
PCAM-soft+ICP [21], [43]	0.120	0.362	0.790	0.452	0.206	98.5%
SpinNet [44]	0.125	0.182	0.760	0.680	0.317	86.7%
DIP [45]	0.098	0.085	0.560	0.610	0.292	91.2%
GeDi [46]	0.084	0.075	0.420	0.460	0.342	97.3%
BTreeNet [47]	0.139	0.572	2.697	1.908	0.812	85.5%
Ours	0.176	0.112	0.828	0.736	**0.014**	97.8%
Ours+ICP	**0.034**	**0.052**	**0.108**	0.885	0.328	**99.4%**

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3417415

As the results shown in Tab. 1 and Fig. 5, classical methods still have a certain level of reliability when dealing with large-scale outdoor LiDAR point cloud registration problems like KITTI. Whether it is point-to-point or point-to-surface ICP algorithms, their registration accuracy metrics, such as RTE and RRE, are around 25cm and 0.15deg respectively, with registration recall rates exceeding 90%. FGR and RANSAC are the two best-performing traditional methods, as they utilize the idea of global registration, to avoid registration results getting trapped in local optima. Our method achieves a similar level of accuracy as FGR and RANSAC in three precision evaluation metrics but far surpasses them in terms of registration efficiency, with a single registration taking only 13.5ms, much less than FGR's 4.241s and RANSAC's 2.163s. Furthermore, our method can provide stable initial values (the standard deviations metric Std for both RTE and RRE metrics are very small), by using the registration results from our method as initial values for ICP, we can further improve the registration accuracy, resulting in an RTE of 3.37cm and an RRE of 0.108deg, which are the best results in comparison, the efficiency also shows improvement compared to the original ICP. On the other hand, Go-ICP, a global registration algorithm based on branch and bound, fails to achieve satisfactory results in large-scale point cloud registration in the comparative experiments, with all its metrics being at a low level. Learning-based methods have received extensive attention in recent years, as shown in Tab. 1, DGR performs modestly in terms of registration accuracy but exhibits good recall rates. USIP+RANSAC, a learning-based strategy combining point cloud keypoints extraction and global registration, outperforms RANSAC in the experiments. HRegNet shows good performance in computing the transformation matrix between source point clouds and keyframes according to its paper but performs averagely in point cloud registration of consecutive frames in comparative experiments. GeDi is currently demonstrating the best performance among learning-based methods. However, our method combined with ICP achieves higher registration accuracy at a similar level of computational efficiency of GeDi. In addition, BTreeNet is the only unsupervised learning method among the comparative approaches specifically designed for large-scale 3D point cloud registration. This network features a novel forward propagation that separates the learning of rotation and translation features, avoiding interference between rotation and translation estimates within a single matrix. It utilizes Chamfer distance and Earth Mover's Distance as loss functions for unsupervised learning. BTreeNet demonstrates good generalization, manifested by small RTE, but it performs poorly in terms of RRE, indicating significant angular errors when applying this method to large-scale point cloud registration. Fig. 5 demonstrates the application of our registration method in LiDAR odometry, where continuous registration results between point cloud frames are continuously outputted to achieve odometry estimation. Nevertheless, in practical applications, additional optimization and loop detection methods are still needed to address the issue of odometry error accumulation and drift.



(a)

(b)

Fig. 6 The motion estimation using our network plus ICP on sequence 09 (a) and 10 (b) of the KITTI dataset

To evaluate the performance of our network in registering consecutive point cloud frames with large rotation variation, we applied a Euclidean transformation to the original target frames in the test samples. This transformation included random variations of pitch, yaw, and roll angles between 0-45°, as well as translations along the three axes between -1.0m to 1.0m. This was done to simulate scenarios where there are significant rotation variation between consecutive point cloud frames. The comparison results are shown in Tab. 2.

Tab. 2 Registration performance with large rotation variation in consecutive frames on KITTI dataset

Methods	RTE (m)		RRE (deg)		Time (s)	Recall
	Mean	Std	Mean	Std		
ICP (p2point) [21]	0.885	0.662	18.264	15.062	0.756	27.3%
ICP (p2plane) [23]	1.056	0.814	21.216	16.492	2.933	23.3%
FGR [41]	0.172	0.469	1.229	1.266	4.984	95.2%
RANSAC [30]	0.289	0.731	2.471	10.714	3.371	84.5%
DGR [42]	0.926	1.105	5.567	4.425	1.486	53.5%
USIP+RANSAC [30], [36]	0.269	0.862	2.146	5.642	3.548	85.7%
HRegNet [39]	1.107	1.224	8.926	6.735	0.156	35.2%
PCAM-soft+ICP [21], [43]	0.164	0.516	0.844	**0.558**	0.327	98.1%
SpinNet [44]	0.372	0.462	1.540	1.680	0.369	74.3%
DIP [45]	0.353	0.372	0.983	0.829	0.324	75.6%
GeDi [46]	0.324	0.432	0.906	0.846	0.371	80.2%
BTreeNet [47]	0.245	0.705	3.132	2.544	0.932	79.7%
Ours	0.323	0.111	1.359	0.732	**0.014**	93.6%
Ours+ICP	**0.038**	**0.063**	**0.152**	1.036	0.283	**99.2%**

According to the comparison results, when there is a large rotation variation between consecutive frames, the ICP can hardly compute effective registration results, with a registration recall rate of less than 30%. The performance of the FGR and RANSAC is somewhat satisfactory, but they also require increased computation time. As for the learning-based methods, the significant rotation variation between the source and target frames prevents them from obtaining reliable feature correspondences during the registration process, resulting in registration failures. Only RANSAC+USIP, PCAM-soft+ICP and the unsupervised method BTreeNet can achieve correct registration output in certain scenarios. Although our method shows a slight decrease in accuracy compared to the original scenes, both the RTE and RRE exhibit very small standard deviations, indicating excellent overall registration stability. Furthermore, by using the registration results obtained from our method as initial values input to ICP, significantly better registration results can be achieved compared to other methods. The good initial values also enable ICP to converge more quickly, enhancing the robustness of the registration.

2) Qualitative evaluation

For qualitative analysis of the registration, we chose to perform comparisons using visualizations. Fig. 6 presents two qualitative examples of our proposed registration method. The qualitative results demonstrate that our method can generate accurate point cloud correspondences even when facing large rotation variation between adjacent frame point clouds.

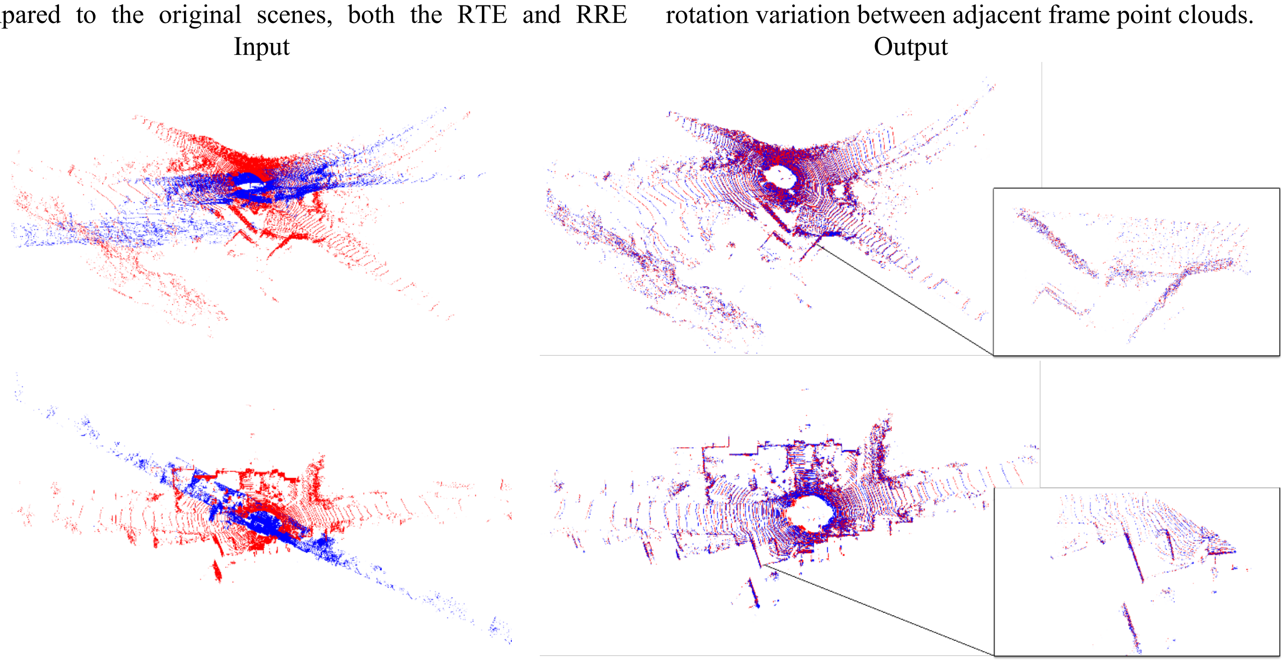


Fig. 7 Qualitative visualization evaluation of the proposed registration method based on KITTI dataset. The left column consists of the source and target point clouds to be registered, while the right column shows the results after precise registration using the proposed method, including local point cloud alignment information.

D. Registration Experiments on PandaSet Dataset

1) Quantitative evaluation

In order to further evaluate the performance of our registration method in different large-scale outdoor scene point clouds, we used the PandaSet dataset. The PandaSet dataset consists of sequential data from 103 different scenes, with each sequence containing point cloud data collected by two types of LiDAR sensors. Each frame of point cloud data is associated with its corresponding pose. We selected point cloud data collected by the 360° rotating LiDAR sensor, Pandar64. We used sequences 0-72 for training, sequences 73-87 for validation, and sequences 88-102 for testing. Similar to the handling of the KITTI dataset, for each sample, we performed zero-padding sampling on the extracted surf point clouds and corner point clouds, resulting in 1280 and 640 points, respectively. As for the non-planar point clouds extracted using plane segmentation algorithms, we sampled them randomly to obtain 8192 points. We compared our network with ICP, FGR, RANSAC, DGR, USIP, HRegNet, PCAM, SpinNet, DIP, GeDi and BTreeNet in registering consecutive point cloud frames with large rotation variation. The comparison results are shown in Tab. 3.

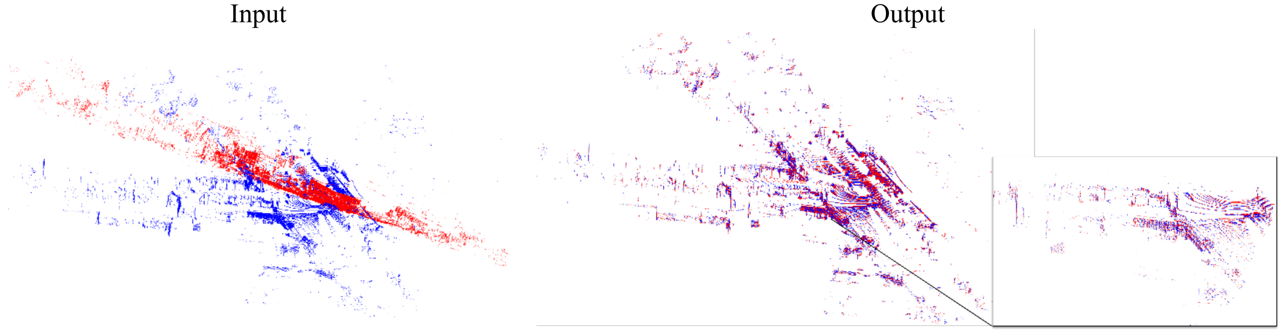
Tab. 3 Registration performance with large rotation variation in consecutive frames on PandaSet dataset

Methods	RTE (m)		RRE (deg)		Time (s)	Recall
	Mean	Std	Mean	Std		
ICP (p2point) [21]	1.198	0.671	21.942	14.683	0.308	18.9%
FGR [41]	0.131	0.461	0.239	**0.172**	11.369	98.7%
RANSAC [30]	0.110	0.352	0.279	0.181	5.854	98.2%
DGR [42]	0.879	1.032	4.978	4.568	1.512	58.2%
USIP+RANSAC [30], [36]	0.108	0.113	0.293	0.269	2.564	98.5%
HRegNet [39]	1.032	1.178	6.793	5.325	0.132	37.6%
PCAM-soft+ICP [21], [43]	0.138	0.423	0.678	0.546	0.337	98.4%
SpinNet [44]	0.316	0.417	1.214	1.543	0.354	75.1%
DIP [45]	0.295	0.358	0.842	0.674	0.311	76.4%
GeDi [46]	0.273	0.383	0.789	0.743	0.357	82.3%
BTreeNet [47]	0.227	0.675	2.762	2.112	0.894	80.5%
Ours	0.276	0.209	0.581	0.565	**0.025**	97.6%
Ours+ICP	**0.034**	**0.179**	**0.117**	0.463	0.217	**99.8%**

By analyzing the results in Tab. 3, it is evident that when facing significant rotational variations between consecutive frames in the PandaSet dataset, relying solely on ICP results in almost registration failure. FGR and RANSAC demonstrate excellent registration accuracy and achieve a recall rate of over 98%. However, both methods incur further increases in the average registration time per frame, which may not meet the real-time requirements of practical applications. In the case of learning-based methods, the combinations of RANSAC+USIP and PCAM-soft+ICP both perform acceptably, while HRegNet continues to perform poorly. BTreeNet and GeDi also exhibit relatively stable performance. Our method, when compared to the computational results on the KITTI dataset, still exhibits outstanding performance on the PandaSet dataset. It not only provides more stable outputs in terms of metrics like RTE and RRE but also achieves a registration recall rate of 97.6%, making the proposed approach combined with ICP deliver the best performance.

2) Qualitative evaluation

Similar to the KITTI dataset, we conducted a qualitative analysis of the registration results based on the PandaSet dataset through visualization. Fig. 7 presents two qualitative examples of our proposed registration method. From the qualitative global and local visual results, it can be observed that our method is capable of generating accurate point cloud correspondences even when faced with large rotation variation between adjacent point cloud frames in the PandaSet dataset.



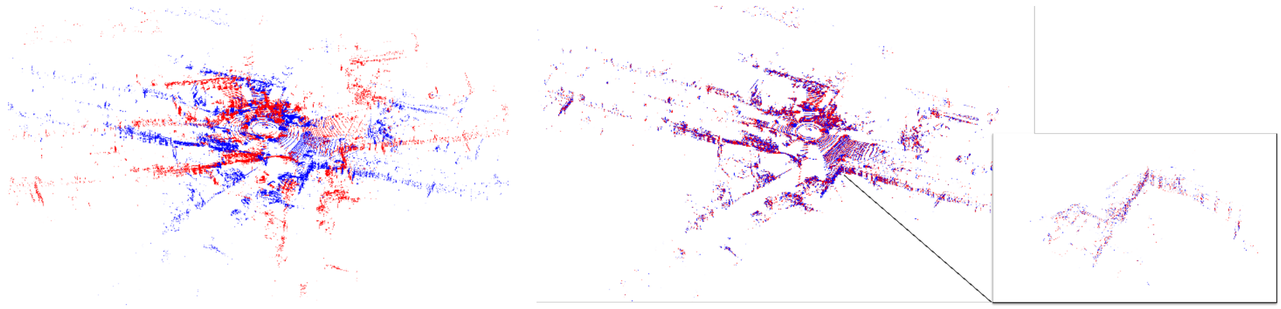


Fig. 8 Qualitative visualization evaluation of the proposed registration method based on PandaSet dataset. The left column consists of the source and target point clouds to be registered, while the right column shows the results after precise registration using the proposed method, including local point cloud alignment information.

V. Conclusion

In this paper, a convolutional end-to-end unsupervised registration network is proposed for large-scale outdoor LiDAR point cloud registration, and it utilizes an improved keypoints extracting method based on rotation compensation and a dynamic overlap ratio weighted chamfer distance loss for efficient and robust registration of large-scale sparse point cloud. The improved point cloud keypoints extracting method utilizes our proposed rotation compensation technique to reliably extract features from point cloud frames with a significant tilt angle relative to the sensor coordinate system. Within the proposed end-to-end unsupervised registration network, we extract global features from the keypoint point clouds and learn the information about the overlapping regions of the point clouds through a spatial attention weight encoder. We achieve reliable convergence by employing an improved dynamic overlap ratio weighted chamfer distance loss. In the registration tests based on the KITTI and PandaSet dataset, it is concluded that our method demonstrates an enhanced performance in terms of accuracy and computational efficiency, when it is applied to either original consecutive frames or the case of simulating large angular variations in real-world scenarios between consecutive frames by randomly transforming the target frame. Furthermore, by combining our method with the precise registration method ICP, we achieve optimal accuracy and robustness for the registration of outdoor large-scale point clouds.

Although our method performs excellently in registering large-scale outdoor LiDAR point clouds, it still has some limitations that require further research for resolution. Firstly, when applied independently, our method often does not yield ideal results for registering low-overlap large-scale point clouds. It may require the use of methods such as ICP or nonlinear optimization to further enhance accuracy. Secondly, since our method draws inspiration from 3D keypoints extraction in laser SLAM, it can be effectively applied to point cloud registration problems collected by multi-line LiDARs. However, when faced with the challenge of registering model point clouds in small indoor scenes, our method cannot be directly applied. In such cases, it may be beneficial to replace the existing keypoints extracting strategy with other excellent 3D interest point extracting methods like USIP, allowing for the unified treatment of indoor and outdoor point cloud registration problems and improving the overall generalizability of our registration approach.

References

[1] R. C. Shinde, S. S. Durbha, and A. V. Potnis, “LidarCSNet: A Deep Convolutional Compressive Sensing Reconstruction Framework for 3D Airborne Lidar Point Cloud,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 180, pp. 313–334, Oct. 2021, doi: 10.1016/j.isprsjprs.2021.08.019.

[2] I. Belkin, A. Abramenko, and D. Yudin, “Real-Time Lidar-based Localization of Mobile Ground Robot,” *Procedia Computer Science*, vol. 186, pp. 440–448, 2021, doi: 10.1016/j.procs.2021.04.164.

[3] F. Huang, D. Shen, W. Wen, J. Zhang, and L.-T. Hsu, “A Coarse-to-Fine LiDAR-Based SLAM with Dynamic Object Removal in Dense Urban Areas,” presented at the Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021), Sep. 2021, pp. 3162–3172. doi: 10.33012/2021.18083.

[4] J. Liu, Y. Liang, D. Xu, · Gong, and J. Hyyppä, “A ubiquitous positioning solution of integrating GNSS with LiDAR odometry and 3D map for autonomous driving in urban environments,” May 2023.

[5] S. Ham, J. Im, M. Kim, and K. Cho, “Construction and Verification of a High-Precision Base Map for an Autonomous Vehicle Monitoring System,” *IJGI*, vol. 8, no. 11, p. 501, Nov. 2019, doi: 10.3390/ijgi8110501.

[6] R. Imai, K. Nakamura, Y. Tsukada, N. Aso, and J. Yamamoto, “Automatic Generation of Road Feature Identification Models from Point Cloud Data Using HD Maps,” in *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, Nov. 2022, pp. 1–6. doi: 10.1109/SCISISIS55246.2022.10001908.

[7] J. Li, X. Huang, and J. Zhan, “High-Precision Motion Detection and Tracking Based on Point Cloud Registration and Radius Search,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 6, pp. 6322–6335, Jun. 2023, doi: 10.1109/TITS.2023.3250209.

[8] D. Xu, J. Liu, Y. Liang, X. Lv, and J. Hyyppä, “A LiDAR-based single-shot global localization solution using a cross-section shape context descriptor,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 189, pp. 272–288, Jul. 2022, doi: 10.1016/j.isprsjprs.2022.05.005.

[9] D. Xu, J. Liu, J. Hyyppä, Y. Liang, and W. Tao, “A heterogeneous 3D map-based place recognition solution using virtual LiDAR and a polar grid height coding image descriptor,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 183, pp. 1–18, Jan. 2022, doi: 10.1016/j.isprsjprs.2021.10.020.

[10] H. Deng, T. Birdal, and S. Ilic, “PPFNet: Global Context Aware Local Features for Robust 3D Point Matching,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 195–205. doi: 10.1109/CVPR.2018.00028.

[11] H. Ma, D.-Yu. Yin, J.-B. Liu, and R.-Z. Chen, “3D convolutional auto-encoder based multi-scale feature extraction for point cloud registration,” *Optics & Laser Technology*, vol. 149, p. 107860, May 2022, doi: 10.1016/j.optlastec.2022.107860.

[12] Y. Wang and J. Solomon, “Deep Closest Point: Learning Representations for Point Cloud Registration,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea

(South): IEEE, Oct. 2019, pp. 3522–3531. doi: 10.1109/ICCV.2019.00362.

[13] Y. Wang and J. M. Solomon, "PRNet: Self-Supervised Learning for Partial-to-Partial Registration," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2019. Accessed: Jul. 07, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/hash/ebad33b3c9fa1d10327bb55f9e79e2f3-Abstract.html

[14] D. Yin *et al.*, "Interest point detection from multi-beam light detection and ranging point cloud using unsupervised convolutional neural network," *IET Image Processing*, vol. 15, no. 2, pp. 369–377, 2021, doi: 10.1049/ipr2.12027.

[15] D. Zai *et al.*, "Pairwise registration of TLS point clouds using covariance descriptors and a non-cooperative game," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 134, pp. 15–29, Dec. 2017, doi: 10.1016/j.isprsjprs.2017.10.001.

[16] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D registration," in *2009 IEEE International Conference on Robotics and Automation*, May 2009, pp. 3212–3217. doi: 10.1109/ROBOT.2009.5152473.

[17] Z. Dong, B. Yang, Y. Liu, F. Liang, B. li, and Y. Zang, "A novel binary shape context for 3D local surface description," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 431–452, Aug. 2017, doi: 10.1016/j.isprsjprs.2017.06.012.

[18] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," in *Robotics: Science and Systems X*, Robotics: Science and Systems Foundation, Jul. 2014. doi: 10.15607/RSS.2014.X.007.

[19] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013, doi: 10.1177/0278364913491297.

[20] P. Xiao *et al.*, "PandaSet: Advanced Sensor Suite Dataset for Autonomous Driving," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Sep. 2021, pp. 3095–3101. doi: 10.1109/ITSC48978.2021.9565009.

[21] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, SPIE, Apr. 1992, pp. 586–606. doi: 10.1117/12.57955.

[22] H. Liu, Y. Zhang, L. Lei, H. Xie, Y. Li, and S. Sun, "Hierarchical Optimization of 3D Point Cloud Registration," *Sensors*, vol. 20, no. 23, p. 6999, Dec. 2020, doi: 10.3390/s20236999.

[23] Segal, Aleksandr and Haehnel, Dirk and Thrun, Sebastian, "Generalized-icp," *Robotics: science and systems*, p. 435, 2009.

[24] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016, doi: 10.1109/TPAMI.2015.2513405.

[25] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, Oct. 2003, pp. 2743–2748 vol.3. doi: 10.1109/IROS.2003.1249285.

[26] T. Nakamura and Y. Tashita, "Congruence Transformation Invariant Feature Descriptor for Robust 2D Scan Matching," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 2013, pp. 1648–1653. doi: 10.1109/SMC.2013.284.

[27] T. Nakamura and S. Wakita, "Robust global scan matching method using congruence transformation invariant feature descriptors and a geometric constraint between keypoints," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2014, pp. 3103–3108. doi: 10.1109/SMC.2014.6974404.

[28] X. Zezhong, L. Jilin, and X. Zhiyu, "Scan matching based on CLS relationships," in *IEEE International Conference on Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003*, Oct. 2003, pp. 99–104 vol.1. doi: 10.1109/RISSP.2003.1285556.

[29] F. Ramos, D. Fox, and H. Durrant-Whyte, "CRF-Matching: Conditional Random Fields for Feature-Based Scan Matching".

[30] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.

[31] X. Huang, G. Mei, J. Zhang, and R. Abbas, "A comprehensive survey on point cloud registration." arXiv, Mar. 05, 2021. Accessed: Jul. 07, 2023. [Online]. Available: http://arxiv.org/abs/2103.02690

[32] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 77–85. doi: 10.1109/CVPR.2017.16.

[33] A. V. Phan, M. L. Nguyen, Y. L. H. Nguyen, and L. T. Bui, "DGCNN: A convolutional neural network over large-scale labeled graphs," *Neural Networks*, vol. 108, pp. 533–543, Dec. 2018, doi: 10.1016/j.neunet.2018.09.001.

[34] W. Lu, G. Wan, Y. Zhou, X. Fu, P. Yuan, and S. Song, "DeepVCP: An End-to-End Deep Neural Network for Point Cloud Registration," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 12–21. doi: 10.1109/ICCV.2019.00010.

[35] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jul. 07, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/d8bf84be3800d12f74d8b05e9b89836f-Abstract.html

[36] J. Li and G. H. Lee, "USIP: Unsupervised Stable Interest Point Detection From 3D Point Clouds," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 361–370. doi: 10.1109/ICCV.2019.00045.

[37] F. Lu, G. Chen, Y. Liu, Z. Qu, and A. Knoll, "RSKDD-Net: Random Sample-based Keypoint Detector and Descriptor," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 21297–21308. Accessed: Jul. 07, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/f40723ed94042ea9ea36bfb5ad4157b2-Abstract.html

[38] Z. Zhang, G. Chen, X. Wang, and M. Shu, "DDRNet: Fast point cloud registration network for large-scale scenes," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, pp. 184–198, May 2021, doi: 10.1016/j.isprsjprs.2021.03.003.

[39] F. Lu *et al.*, "HRegNet: A Hierarchical Network for Large-scale Outdoor LiDAR Point Cloud Registration," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 15994–16003. doi: 10.1109/ICCV48922.2021.01571.

[40] T. Shan and B. Englot, "LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid: IEEE, Oct. 2018, pp. 4758–4765. doi: 10.1109/IROS.2018.8594299.

[41] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast Global Registration," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 766–782. doi: 10.1007/978-3-319-46475-6_47.

[42] C. Choy, W. Dong, and V. Koltun, "Deep Global Registration," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 2511–2520. doi: 10.1109/CVPR42600.2020.00259.

[43] A.-Q. Cao, G. Puy, A. Boulch, and R. Marlet, "PCAM: Product of Cross-Attention Matrices for Rigid Registration of Point Clouds." arXiv, Oct. 04, 2021. Accessed: Jul. 21, 2023. [Online]. Available: http://arxiv.org/abs/2110.01269

[44] S. Ao, Q. Hu, B. Yang, A. Markham, and Y. Guo, "SpinNet: Learning a General Surface Descriptor for 3D Point Cloud Registration." arXiv, Apr. 09, 2021. Accessed: Jan. 20, 2024. [Online]. Available: http://arxiv.org/abs/2011.12149

[45] F. Poiesi and D. Boscaini, "Distinctive 3D local deep descriptors." arXiv, Dec. 28, 2020. Accessed: Jan. 20, 2024. [Online]. Available: http://arxiv.org/abs/2009.00258

[46] F. Poiesi and D. Boscaini, "Learning general and distinctive 3D local deep descriptors for point cloud registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2022, doi: 10.1109/TPAMI.2022.3175371.

[47] L. Xi, W. Tang, T. Xue, and T. Wan, "Iterative BTreeNet: Unsupervised learning for large and dense 3D point cloud registration," *Neurocomputing*, vol. 506, pp. 336–354, Sep. 2022, doi: 10.1016/j.neucom.2022.07.082.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2024.3417415

Jingbin Liu received the Ph.D. Degree from Wuhan University, Wuhan, China, in 2008. From 2008 to 2016, he was a Specialist Research Scientist with Finnish Geospatial Research Institute (FGI, formerly known as Finnish Geodetic Institute). He is a Professor in positioning and navigation with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His interest areas include indoor and outdoor positioning, smartphone navigation, indoor mobile mapping, and GNSS/INS/SLAM integration technology.

Xuanfan Lv received the B.S. degree in School of Geodesy and Geomatics from Wuhan University, Wuhan, China in 2021. He is currently working toward the Master degree at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) in Wuhan University, Wuhan, China.

His research interests include laser scanning, SLAM and data processing.

Xiaodong Gong received the M.S. degree in surveying and mapping engineering from the School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, China, in 2018 and the Ph.D. degree in Geodetic Engineering from Wuhan University, China, in 2022.

He is a postdoctoral researcher in the Ubiquitous Navigation Lab (UNL) at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China. His research interests include indoor position and navigation, building information model (BIM), and mobile geographic information system (GIS).

Yifan Liang received the B.S. degree in surveying and mapping engineering from China University of Mining And Technology, Xuzhou, China, in 2017. He is currently working toward the doctoral degree at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) in Wuhan University, Wuhan, China.

His research interests include laser scanning, SLAM and data processing.

Juha Hyyppä received the M.Sc., Dr.Ing., and Dr.Sc. degrees, all with honors, from Helsinki University of Technology, Helsinki, Finland, in 1987, 1990, and 1994, respectively.

He is currently a Professor of remote sensing and photogrammetry, Finnish Geospatial Research Institute, Director of the Centre of Excellence in Laser Scanning Research, laserscanning.fi, and a Distinguished Professor, Shinshu University (Japan). His references include 30 years experiences in research team leadership, coordination of more than ten international science projects, and author of more than 200 ISI Web of Science listed papers. He has focus on laser scanning systems, their performance and new applications, especially related to mobile, personal and ubiquitous laser scanning and their point cloud processing especially to forest information extraction.