Hindawi
Journal of Sensors
Volume 2023, Article ID 3427758, 12 pages
https://doi.org/10.1155/2023/3427758

Research Article

A Partial-to-Partial Point Cloud Registration Method Based on Geometric Attention Network

Yi Chen, Yong Wang, Jinlong Li, Yu Zhang, and Xiaorong Gao

School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 610031, China

Correspondence should be addressed to Yong Wang; wangyonga@swjtu.edu.cn

Received 28 October 2022; Revised 26 April 2023; Accepted 25 September 2023; Published 27 October 2023

Academic Editor: Giovanni Diraco

Partial point cloud registration is an important step in generating a full 3D model. Many deep learning-based methods show good performance for the registration of complete point clouds but cannot deal with the registration of partial point clouds effectively. Recent methods that seek correspondences over downsampled superpoints show great potential in partial point cloud registration. Therefore, this paper proposes a partial-to-partial point cloud registration network based on geometric attention (GAP-Net), which mainly includes a backbone network optimized by a spatial attention module and an overlapping attention module guided by geometric information. The former aggregates the feature information of superpoints, and the latter focuses on superpoint matching in overlapping regions. The experimental results show that the method achieves better registration performance on ModelNet and ModelLoNet with lower overlap. The rotation error is reduced by 14.49% and 17.12%, respectively, which is robust to the overlap rate.

1. Introduction

As a key technology in computer vision and robotics, point cloud registration is a fundamental guarantee for accomplishing various downstream tasks, such as 3D reconstruction and simultaneous localization and mapping. Due to the rapid development of LiDAR, sensor technology, and stereo cameras, point clouds have become a very important data format and are widely used in many fields, such as autonomous driving, robotics, medical care, and cultural relics protection [1, 2]. However, the point cloud data of the actual 3D model need to be collected from different perspectives, which are incomplete. When registering, they only have partial correspondences. Therefore, the registration of real-world point clouds is still a challenge.

Since the point cloud registration was proposed, scholars at home and abroad have contributed a lot of research results. Iterative closest point [3] is the most classic algorithm. It searches for the closest point between the two point clouds to find the point-pair matching relationship and uses the Euclidean distance between the matching point pairs as the objective function to iterate until the accuracy meets the requirements or to iterate to convergence. Subsequently, handcrafted descriptors such as point pair feature (PPF) [4], signature of histogram of orientation [5], and rotational projection statistics [6] are designed to find point cloud local invariance features to aid in transform estimation for point cloud registration. However, the point cloud registration based on traditional methods cannot be generalized to a large amount of multiclass data.

In recent years, point cloud registration methods based on deep learning have attracted the attention of many scholars. The early deep learning-based point cloud registration network PointNetLK [7] utilizes the PointNet framework to extract the global features of the point cloud, and the Lucas and Kanade (LK) algorithm minimizes the distance between the point cloud features. Deep closest point (DCP) [8] adopts a graph-based approach to learn pointwise features of structures, establishing soft correspondences between point clouds by using rigid-invariant features extracted by an attention mechanism network. PointNetLK and DCP perform well when the input is a full point cloud but cannot handle part-to-part registration scenarios. Subsequently, RPM-Net [9] predicts the correspondence of partial point clouds through the Sinkhorn layer, which can process point clouds with partial visibility. PRNet [10] proposes a partial point cloud registration network for feature-based L2-norm keypoint detection to find

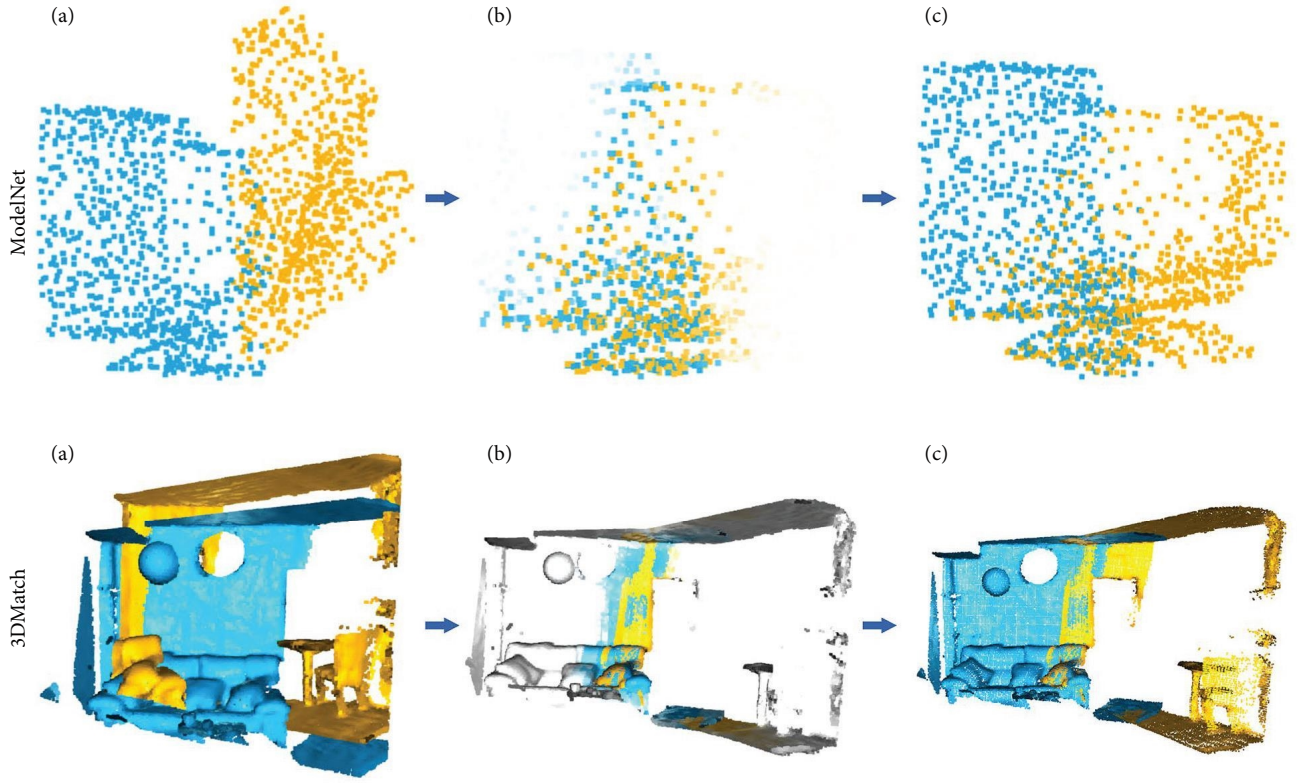

Figure 1: (a) Input pair. (b) Overlapping regions point pair matching. (c) Output pair. For two partially overlapping point clouds, more attention is needed to the overlapping regions. When the correct corresponding point pairs in the overlapping regions are obtained, the exact correspondence can be generated.

common points in input point clouds. At the same time, the method based on feature learning focuses on extracting useful information such as the geometry of point clouds to form discriminative features, which has become the focus of recent scholars. PPFNet [11] introduces PPF features for feature encoding. Fully convolutional geometric features (FCGF) [12] utilizes a 3D fully convolutional network to expand the receptive field and extract geometric features. D3Feat [13] utilizes KPConv [14] to build a fully convolutional encoder–decoder architecture for joint dense detection and description. Based on D3Feat, PREDATOR [15] introduces a module for extracting key points in overlapping regions to establish correspondences. The above algorithms show that the method of finding correspondences on the downsampled superpoints has great potential for partial point cloud registration, and these algorithms have been able to achieve good performance in partial point cloud registration. However, it is still a challenging task to extract the common key points of two partially overlapping point clouds, as shown in Figure 1. Therefore, the method proposed in this paper to combine geometric features with Transformer provides a novel idea for the method of finding correspondences on the downsampled superpoints.

The accuracy of point cloud registration networks based on key point extraction is highly dependent on the accuracy of superpoint matching, so the extracted superpoints need to capture more global context features. Based on this, this paper proposes an improved partial point cloud registration network (GAP-Net) for accurate point cloud registration. Inspired by Transformer [16], our method employs Transformer to encode contextual information in registration before skipping connection blocks of the KPConv backbone network. In the overlap geometric attention module (OGA), the Transformer layer is guided to further aggregate the geometric features of the point cloud by using the coordinate and normal information and exchange information between the two point clouds. The main contributions of this paper can be summarized as follows:

(1) A novel framework for partial-to-partial point cloud registration is proposed, which uses a spatial self-attention mechanism to optimize the KPConv backbone to capture the extracted features of each point cloud.

(2) This paper adopts a random expansion strategy for the extracted superpoints to prevent the problem of K-nearest neighbor (KNN) layer breakage due to too few superpoints in the KNN algorithm. At the same time, it can expand the receptive field to facilitate the extraction of geometric features.

(3) A new geometric self-attention (GSA) module is proposed that uses coordinate and normal feature information to guide the attention mechanism to integrate more global contexts with the learned geometric features. It allows for information exchange between the two point clouds, and the subsequent steps can be focused on overlapping regions for robust superpoint matching.

2. Related Works

2.1. Traditional Registration Methods. Traditional point cloud registration methods generally include two stages: coarse registration and fine registration. Coarse registration provides a good initial position for fine registration, avoids falling into a local optimal solution during fine registration, and improves the accuracy of fine registration. The fine matching criterion is based on the coarse registration, which minimizes the differences between point clouds, such as spatial position differences, so as to obtain a more accurate rotation and translation matrix. For the commonly used algorithm sample consensus initial alignment [17] in the coarse registration stage, the FPFH feature is used to search for point correspondences, which makes the algorithm insensitive to the initial position of the point cloud. 4PCS [18] randomly selects four coplanar points in the target point cloud as the basic point pair for feature matching and uses the largest common pointset strategy to find the optimal matching point pair in the source point cloud. Super4PCS [19] adopts an intelligent indexing strategy, which reduces the computational complexity of 4PCS. Among the precise registration methods, the most classic ICP algorithm can obtain high-precision registration results and is widely used. However, ICP is very sensitive to initial values and outliers, and it is easy to fall into the local optimal solution. So, a series of variant algorithms, such as GO-ICP [20], are derived. In addition, there are some methods that use probability for registration. The normal distributions transform [21] algorithm determines the optimal transformation relationship between the point clouds to be registered based on optimization theory by discretizing the transformation space and combining the objective function to measure the registration error. The coherent point drift [22] algorithm transforms point cloud registration into a probability density estimation problem and uses a Gaussian mixture model and an EM algorithm to complete the registration. However, traditional registration methods have less research on overlapping regions, and they reduce the influence of outliers by dividing corresponding points into inliers and outliers after feature matching. Algorithms to find the correct inliers are RANdom SAmple Consensus (RANSAC) [23], 3DHV [24], etc. But their effect is limited when the proportion of overlapping regions is reduced. Therefore, extracting the points of overlapping regions accurately in the partial point clouds can ensure the registration performance of algorithms such as RANSAC and 3DHV. This paper chooses RANSAC for registration because it is easy to implement.

2.2. Learning-Based Registration Methods. Currently, learning-based methods are popular for registration tasks. DGCNN [25] extracts the feature information of the point cloud through EdgeConv [26]. The EdgeConv proposed by it can extract the local aggregation information of the point cloud under the premise of ensuring that the permutation is unchanged. SiamesePointNet [27] extracts pointwise descriptors directly for registration by introducing the Siamese Point Network, which contains a global shape constraint module and a feature transformation operator. However, some of the initially studied networks assumed that all points in the two point clouds were completely overlapping. Therefore, they are mostly unable to complete the task of partial point cloud registration. OPRNet [28] utilizes the Sinkhorn algorithm for partial registration. OMNet [29] learns masks in a coarse-to-fine manner to reject nonoverlapping regions, which converts the partial-to-partial registration to the registration of the same shapes. ROPNet [30] proposes a context-guided module to extract global features to predict point overlap scores, which are then registered using representative overlapping points with discriminative features. SCANet [31] effectively utilizes global information at different levels by introducing a spatial self-attention aggregation module in the feature extraction part and a channel cross-attention regression module in the pose estimation part for information interaction between the global features of the two point clouds to complete partial point cloud registration. SANet [32] proposes a subtract attention module to aggregate the pointwise features and then obtain the local correspondence between each point to complete the partial point cloud registration. MaskNet++ [33] utilizes spatial self-attention and channel cross-attention mechanisms to extract pointwise features and exchange information, respectively. STORM [34] employs EdgeConv and Transformer [16] to map the input points to a feature space, then performs overlap prediction to identify common points, and Transformer to refine the features, finally completing registration. G3DOA [35] proposes an overlap attention that extracts cocontextual information between the feature encodings of two point clouds to construct a feature descriptor suitable for partial point cloud registration. Inspired by PREDATOR [15], CoFiNet [36], which extracts hierarchical correspondences from coarse to fine, and GeoTransformer [37], which learns geometric features by using the designed attention module, both achieve robust matching on downsampled superpoints.

In summary, previous work has validated the potential of methods for matching on downsampled superpoints with partial registration. The local features and information interaction through the attention mechanism can improve registration performance even further. Based on these, in order to better realize the registration task for partial point clouds, this paper proposes an optimized partial-to-partial point cloud registration framework.

3. Method

3.1. Problem Statement. Given two point clouds $\boldsymbol{P} = \{p_i \in \mathbb{R}^3 | i = 1, 2\ldots, N\}$ and $\boldsymbol{Q} = \{q_j \in \mathbb{R}^3 | j = 1, 2\ldots, M\}$, the purpose of point cloud registration is to estimate a rigid transformation $\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0 & 1 \end{bmatrix} \in \mathrm{SE}(3)$ to align the two point clouds, where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ is a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^{3\times1}$ is a translation matrix. Rigid transformations can be implemented in the following ways:

$$\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{\left(p'_{x_i}, q'_{y_j}\right) \in \mathbf{GT}'} \left\| \boldsymbol{R} \cdot p'_{x_i} + \boldsymbol{t} - q'_{y_j} \right\|_2^2, \tag{1}$$

where $\mathbf{GT}'$ is the set of ground truth corresponding point pairs between the $\boldsymbol{P}$ and $\boldsymbol{Q}$ point clouds.

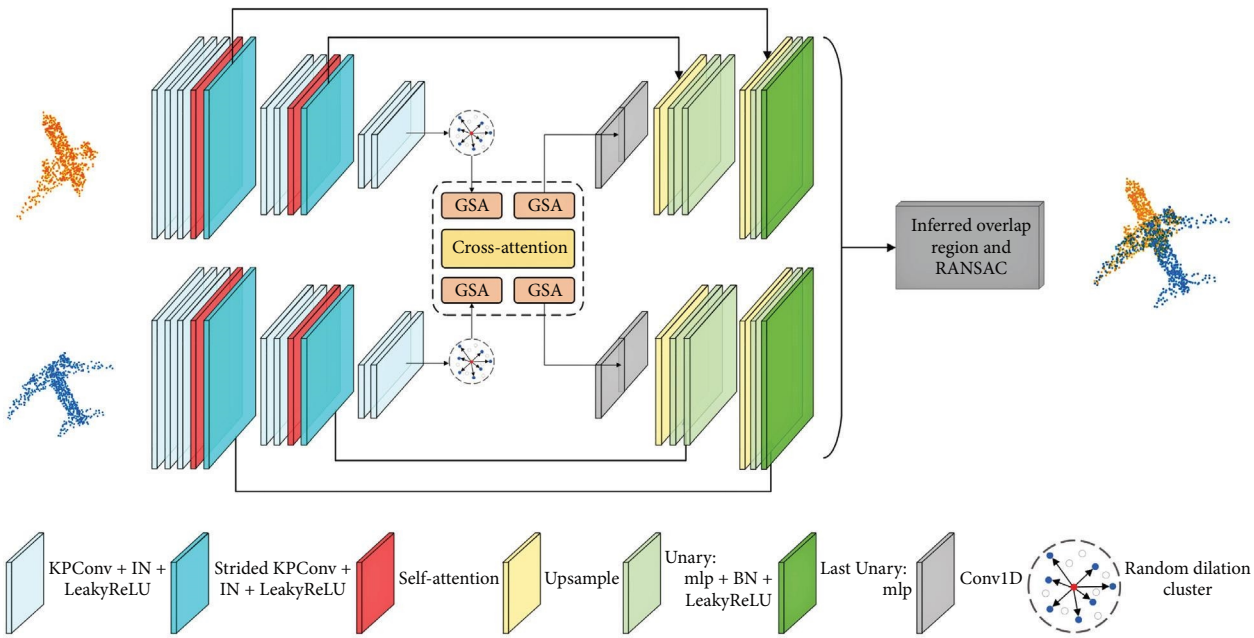

Figure 2: The network architecture of GAP-Net proposed in this paper. The upper line is the overall structure of the entire network, and the lower line is the composition of each module. To extract superpoints and aggregate their features, the KPConv-SSA backbone network is used to downsample the input point clouds. The overlapping attention module, guided by geometric information, is applied to superpoints to encode the information of the point clouds and infer the overlapping regions. Finally, RANSAC is used for registration.

However, in reality, point clouds are often collected from different perspectives, and they are incomplete. In order to form a complete point cloud of an object or scene, it is necessary to perform part-to-part point cloud registration on the two point clouds. Obviously, at this time, it registers two partial point clouds based on the information about the overlapping regions. According to first establishing the point correspondence between the two point clouds and then estimating the path of the transformation matrix, this paper mainly focuses on the former and establishes the point correspondence in the overlapping regions. To this end, this paper proposes GAP-Net, which takes two point clouds as input, outputs point correspondences, and then uses RANSAC [23] to estimate rigid transformations.

3.2. Network Architecture. GAP-Net is an encoder–decoder network, as shown in Figure 2. The encoder adopts the KPConv-SSA backbone network proposed in this paper to simultaneously downsample the input point clouds and extract multilevel features. The basic convolution block is composed of a ResNet-like KPConv/strided KPConv layer, an instance norm layer, and a LeakyReLU layer. At the same time, a spatial self-attention block is added before the strided KPConv block for pointwise feature encoding, which can utilize pointwise and global information at different levels. The spatial self-attention block is shown in Figure 3. The spatial self-attention mechanism in this paper consists of three operations: query (Q), key (K), and value (V). Specifically, given a source feature map $\boldsymbol{F_X}$, the self-attention map $\boldsymbol{A_X}$ is obtained via the softmax function by multiplying the query (Q) in row i, the key (K) in column j, and the value (V)

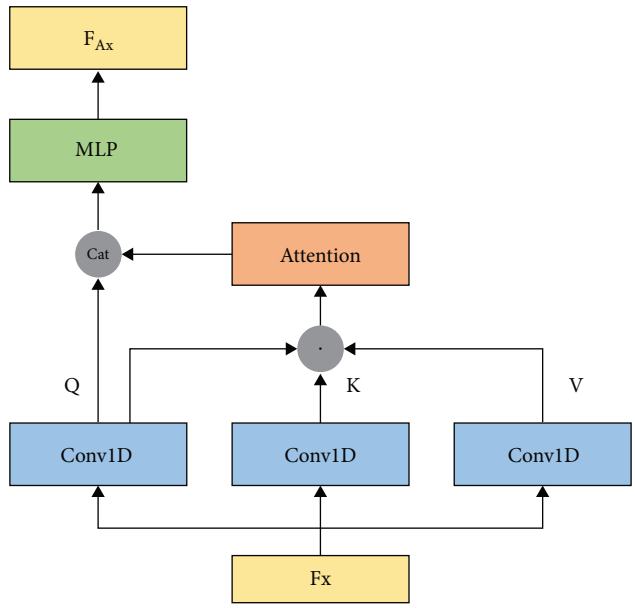

Figure 3: The spatial self-attention mechanism in this paper.

in column j. Second, the attention-based feature map $\boldsymbol{F}_{\boldsymbol{A}_X}$ is obtained by concatenating the query (Q) and the attention map, respectively. Finally, update feature $\boldsymbol{F_X}$ is shown in Equation 2. It is worth noting that, for simplicity, the operations of query and key share weights. The decoder consists of upsampling blocks and linear blocks. The upsampling block uses the nearest search for feature interpolation, and the

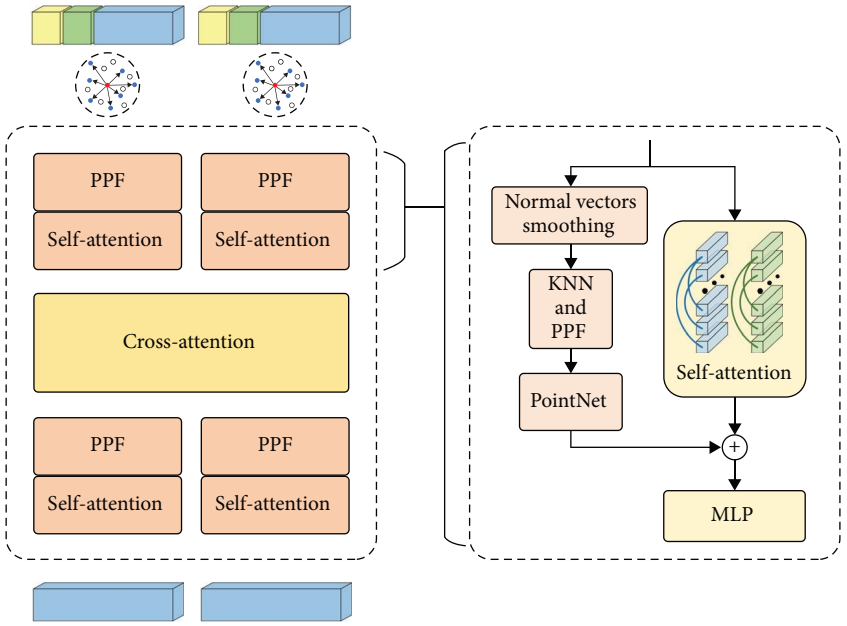

FIGURE 4: (a) Overlap geometric attention module (OGA). The OGA module takes the overlap and corresponding normals and latent features as input and outputs geometric fusion features. (b) Geometric self-attention (GSA). The normal vector is used to enhance geometric encoding.

linear block consists of a linear (MLP) layer, an instance norm layer, and a LeakyReLU layer.

$$\boldsymbol{F}_X = \boldsymbol{F}_X + \boldsymbol{F}_{A_X}. \tag{2}$$

GAP-Net takes the source point cloud $\boldsymbol{P}$, the target point cloud $\boldsymbol{Q}$, and their feature descriptors $\boldsymbol{F}_{P'}, \boldsymbol{F}_{Q'}$ that are corresponding size matrices initialized to 1 as input. First, the encoder performs downsampling and extracts features to obtain superpoints $\boldsymbol{P}', \boldsymbol{Q}'$ and corresponding features $\boldsymbol{F}'_{P'}$, $\boldsymbol{F}'_{Q'}$. Information interaction is then performed in the OGA, guided by geometric information, and the features $\boldsymbol{F}'_{P'}, \boldsymbol{F}'_{Q}$ and scores $\boldsymbol{S}_{P'}, \boldsymbol{S}_{Q'}$ corresponding to the superpoints are output. Then, getting the feature and score of each point through the decoder. Finally, under the guidance of the score, enough key points are extracted to complete the registration task with RANSAC.

3.3. Overlap Geometric Attention Module. Exploiting attention mechanisms to capture global contextual information has played an important role in many computer vision tasks. At present, there are some methods that use attention to extract features using global context information for point cloud registration. However, these methods usually only exploit the high-level point cloud features provided by attention and neglect to use the geometric information of the point cloud to encode with attention. Therfore, this paper proposes OGA, an overlapping attention module guided by point cloud geometric information, to capture the geometric structure of point clouds and encode superpoint features. The OGA module is a bridge between encoders and decoders, and it mainly consists of a geometric information-guided self-attention module and a feature-based cross-attention module, as shown in Figure 4(a).

3.3.1. Random Dilation Cluster. Inspired by RSKDD [38], before using the attention mechanism to encode the point cloud features, this paper randomly expands the superpoints input to the OGA module to deal with the problem that the number of superpoints extracted from the sparse point clouds during registration is not enough to support the subsequent KNN algorithm operation, then causing feature layer breaks and the registration to fail. For the superpoints extracted by the encoder, a KNN search needs to be performed for each point in geometric coding. At this time, in order to solve the problem that the number of superpoints is too small, this paper adopts the random dilation cluster strategy to generate clusters, as shown in Figure 5. Assume that KNN are selected for a single cluster with an expansion rate of α. This paper first searches the $\alpha \times k$ nearest neighbors of the center point and then randomly samples K points from them. Although this strategy is simple, it can effectively avoid the feature layer breakage of superpoints extracted from sparse point clouds when performing geometric encoding.

3.3.2. Geometric Self-Attention. As shown in Figure 4(b), the geometry-guided encoding module (GSA) takes superpoints and corresponding latent features as input and outputs geometrically enhanced features. Inspired by RPM-Net [9], the geometric feature $\boldsymbol{G}_{p_i'}$ of the superpoint $p_i' \in \boldsymbol{P}'$ is constructed with PPF [4], which can be formulated given as follows:

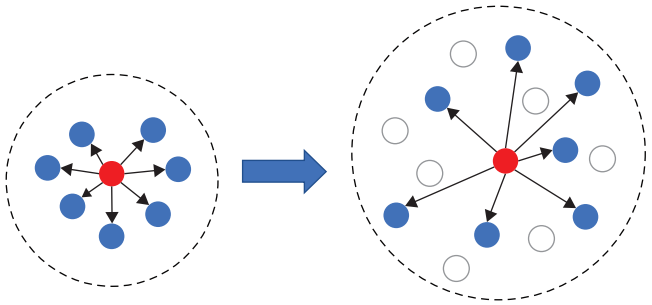

Figure 5: Random dilation cluster strategy. The red point is the center point, and the blue points are the selected neighbor points. The left part is the standard KNN-based cluster, and the right part is the random dilation cluster. It is obvious that it is a significant enlargement of the receptive field.

$$
\begin{aligned}
&\boldsymbol{N}_{p_i'} = \frac{1}{\left|\boldsymbol{J}_i^N\right|} \sum_{p_j \in J_i^N} \boldsymbol{N}_{p_j},\\
&\mathrm{PPF}\left(p_i', p_j'\right) = \left(\angle\left(p_j' - p_i', \boldsymbol{N}_{p_i'}\right), \angle\left(p_j' - p_i', \boldsymbol{N}_{p_j'}\right),\right.\\
&\qquad\qquad \left.\angle\left(\boldsymbol{N}_{p_i'}, \boldsymbol{N}_{p_j'}\right), \left\|p_i' - p_j'\right\|_2\right),\\
&\boldsymbol{G}_{p_j'} = \mathrm{f}_1\left(p_i', p_j' - p_i', \mathrm{PPF}\left(p_i', p_j'\right)\right),\\
&\boldsymbol{G}_{p_i'} = \max\left\{\boldsymbol{G}_{p_j'} \middle| p_j' \in \boldsymbol{J}_i^G\right\}.
\end{aligned}
\tag{3}
$$

where $\boldsymbol{N}_{p'}$ is the normal vector of a point in the superpoint set $\boldsymbol{P}'$ obtained via the encoder, it is calculated by averaging the normal $\boldsymbol{N}_p$ calculated by Open3D over its surrounding points in $\boldsymbol{P}$. $\boldsymbol{J}_i^N = \left\{p_j \middle| \left\|p_j - p_i'\right\| < r^N\right\}, p_j \in \boldsymbol{P}$ and r^N is the radius of p_i''s neighborhood. $\angle(\cdot,\cdot) \in [0, \pi]$ represents the angle between two vectors. f_1 is implemented by PointNet. $\boldsymbol{J}_i^G = \left\{p_j' \middle| \left\|p_j' - p_i'\right\| < r^G\right\}, p_j' \in \boldsymbol{P}'$ and r^G is the radius of p_i''s neighborhood. $\max(\cdot)$ represents channelwise maximum pooling. Inspired by PREDATOR and CoFiNet, a self-attention mechanism is introduced in the GSA module to further aggregate and enhance their contextual relations and obtain the semantic features $\boldsymbol{F}_{p'}'$ output by the encoder as $\boldsymbol{E}_{p'}$. Then, this paper fuses geometric features and semantic features to generate GSA features:

$$
\boldsymbol{F}_{p'}^{gsa} = \mathrm{MLP}\left(\boldsymbol{G}_{p'}, \boldsymbol{E}_{p'}\right). \tag{4}
$$

For computational efficiency, this paper adopts the same architecture as the cross-attention module but acquires features from the same point cloud to implement the self-attention mechanism, i.e., from $\boldsymbol{F}_{p'}'$ to $\boldsymbol{E}_{p'}$.

3.3.3. Information Interaction. The information interaction module in this paper consists of a cross-attention mechanism for information interaction and another GSA module for explicitly updating the local context. The cross-attention module adopts multihead attention, as shown in Figure 6. For the fusion feature $\left(\boldsymbol{F}_{p'}^{gsa}, \boldsymbol{F}_{q'}^{gsa}\right)$ obtained from the previous GSA block, the information of the potential overlap regions is obtained by mixing the feature information of the two point clouds through cross-attention and updating and enhancing the contextual information with the GSA block to complete the information interaction. The features of information interaction are calculated given as follows:

$$
\begin{aligned}
&\mathrm{head}_{\mathrm{ij}-1} = softmax\left(\frac{\mathrm{Q_i K_j^T}}{\sqrt{d_k}}\right) \cdot \mathrm{V_j},\\
&\boldsymbol{F}_{p'}^{ca} = \boldsymbol{F}_{p'}^{gsa} + \mathrm{MLP}\left(\mathrm{cat}\left[\mathrm{head}_{\mathrm{ij}-1}, \ldots, \mathrm{head}_{\mathrm{ij}-\mathrm{k}}\right]\right),\\
&\boldsymbol{F}_{p'}^{in} = \mathrm{MLP}\left(\boldsymbol{G}_{p'}^{ca}, \boldsymbol{E}_{p'}^{ca}\right),
\end{aligned}
\tag{5}
$$

where $\mathrm{Q_i} = \boldsymbol{F}_{p'}^{gsa} \cdot \mathrm{W_i^Q}, \mathrm{K_j} = \boldsymbol{F}_{q'}^{gsa} \cdot \mathrm{W_j^K}, \mathrm{V_j} = \boldsymbol{F}_{q'}^{gsa} \cdot \mathrm{W_j^V}$, and d_k is the dimension of the parameter $\mathrm{K_j}$. $\mathrm{W_i^Q}, \mathrm{W_j^K}$, and $\mathrm{W_j^V}$ are learnable weight matrices. Therefore, updating the information of a superpoint p_i' requires combining the query of that point with the keys and values of all superpoints $q_j' \in \boldsymbol{Q}'$. k is the number of heads, and $\left(\boldsymbol{G}_{p'}^{ca}, \boldsymbol{E}_{p'}^{ca}\right)$ refers to the subsection "Geometric Self-Attention."

3.4. Loss Function

3.4.1. Feature Loss. Circle losses for feature descriptors $\boldsymbol{F}_{\boldsymbol{P}}'$ and $\boldsymbol{F}_{\boldsymbol{Q}}'$ are computed from the randomly sampled correspondences (p_c and q_c) from $\boldsymbol{P}$ and $\boldsymbol{Q}$:

$$
\begin{aligned}
L_C^P = \frac{1}{N_c} \sum_{i=1}^{N_c} \log\Bigg[1 + \sum_{j \in \rho} \exp\left(\alpha_\rho^j \left(D_j^i - \Delta_\rho\right)\right)\\
\cdot \sum_{k \in \eta} \exp\left(\alpha_\eta^k \left(\Delta_\eta - D_k^i\right)\right)\Bigg],
\end{aligned}
\tag{6}
$$

where N_c is the number of the sampled correspondences, $D_j^i = \left\| \boldsymbol{F}_{p_i} - \boldsymbol{F}_{q_j} \right\|_2$ denotes distance in feature space, and Δ_ρ, Δ_η are positive and negative margins, respectively. The weights $\alpha_\rho^j = \beta\left(D_j^i - \Delta_\rho\right)$ and $\alpha_\eta^k = \beta\left(\Delta_\eta - D_k^i\right)$ are determined individually for each positive and negative points and β is a scale factor. Then, the loss L_c^Q is defined in the same way and the total circle loss for feature descriptors is $L_C(F) = \frac{1}{2}\left(L_C^P + L_C^Q\right)$.

3.4.2. Overlap and Saliency Loss. To supervise key points in the overlap regions, we follow PREDATOR [15] and use the overlap loss and matchability loss. Binary cross-entropy loss is used for overlap loss L_o and saliency loss L_s, i.e.,

$$
\begin{aligned}
L_O^P &= \frac{1}{|P|} \sum_{i=1}^{|P|} \overline{O}_{p_i} \log\left(O_{p_i}\right) + \left(1 - \overline{O}_{p_i}\right) \log\left(1 - O_{p_i}\right),\\
L_S^P &= \frac{1}{|P|} \sum_{i=1}^{|P|} \overline{S}_{p_i} \log\left(S_{p_i}\right) + \left(1 - \overline{S}_{p_i}\right) \log\left(1 - S_{p_i}\right),
\end{aligned}
\tag{7}
$$

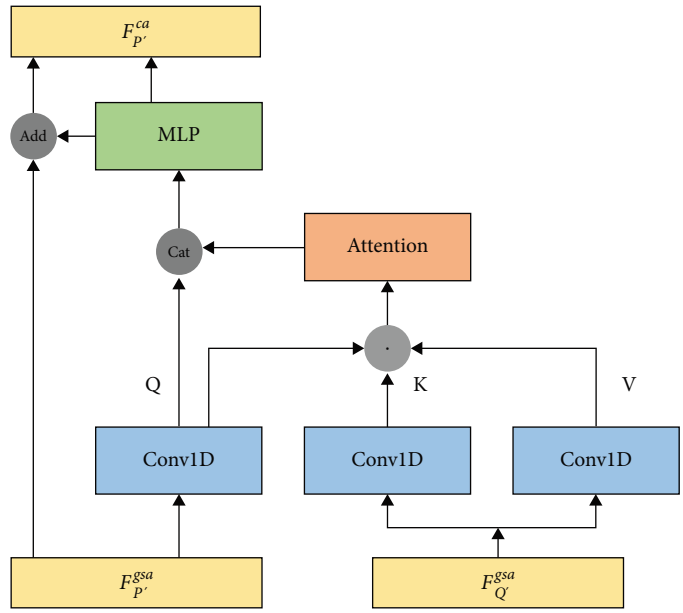

Figure 6: The cross-attention mechanism in this paper.

where $\overline{O}_{p_i}$ and $\overline{S}_{p_i} \in \{0, 1\}$ are the ground truth labels of point p_i. Then, L_O^Q and L_S^Q are defined in the same way. The total overlap loss and the total saliency loss are $L_O = 1/2(L_O^P + L_O^Q)$, $L_S = 1/2(L_S^P + L_S^Q)$, respectively.

3.4.3. Combined Loss. The complete loss function of GAP-Net is given as follows:

$$L = L_C(F) \cdot \omega_i^c + L_O \cdot \omega_i^o + L_S \cdot \omega_i^s, \tag{8}$$

where ω_i^c, ω_i^o, and ω_i^s are weighting factors for sample balance.

4. Experiments

This paper compares GAP-Net with registration methods on synthetic, object-centric ModelNet40 and ModelLoNet (Section 4.1) and tests it using Stanford 3D scanning (Section 4.2). It is proved that the method in this paper can be used for partial registration of point clouds. Furthermore, this paper compares GAP-Net with registration methods on indoor scene point clouds, 3DMatch and 3DLoMatch (Section 4.3), proving that our method is not limited to simple geometric objects but can also be used for large-scale scene point cloud registration.

4.1. ModelNet40 and ModelLoNet

4.1.1. Dataset. ModelNet40 is a widely used point cloud registration dataset consisting of 9,843 CAD models of 40 different object categories for training and 2,468 models for testing. This paper uses 5,112 models for training, 1,202 models for validation, and 1,266 models for testing according to RPM-Net [9]. For a given point cloud, first copy the point cloud and randomly generate a rotation within (0°, 45°) and a translation within (−0.5, 0.5). Then, in order to generate partially overlapping point clouds, we randomly crop along one direction, retaining about 70% of the points. A further 50% downsampling was performed to retain 717 points. In addition to generating a ModelNet with an average pairwise overlap of 73.5%, this paper also generates a ModelLoNet with a lower (53.6%) average overlap according to PREDATOR [15] by retaining about 50% of the points when cropping and then randomly sampling the 717 points that remain in the end. The network was trained by the SGD optimizer, and the network parameters were updated on Intel(R) Xeon(R) CPU E3-1230 V2 3.3 GHz and NVIDIA GeForce GTX 1080 Ti GPU.

4.1.2. Metrics. This paper evaluates the registration based on the relative rotation error (RRE) and relative translation error (RTE) proposed in RPM-Net and the improved chamfer distance.

$$\begin{aligned}
&\text{Error}(\boldsymbol{R}) = \arccos \frac{\text{tr}\left(\widehat{\boldsymbol{R}}^{-1}\boldsymbol{R}\right) - 1}{2}, \\
&\text{Error}(\boldsymbol{t}) = \left\|\widehat{\boldsymbol{R}}^{-1}\boldsymbol{t} - \widehat{\boldsymbol{t}}\right\|_1, \\
&\tilde{\text{CD}}(\boldsymbol{P}, \boldsymbol{Q}) = \frac{1}{|\boldsymbol{P}|}\sum_{p\in\boldsymbol{P}} \min_{q\in\boldsymbol{Q}_{clean}} \|p - q\|_2^2 + \frac{1}{|\boldsymbol{Q}|}\sum_{q\in\boldsymbol{Q}} \min_{p\in\boldsymbol{P}_{clean}} \|p - q\|_2^2,
\end{aligned} \tag{9}$$

where $\boldsymbol{R}, \boldsymbol{t}$ and $\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{t}}$ represent the prediction and ground truth transformation, respectively, and $\text{tr}(\cdot)$ represents the trace of the matrix.

4.1.3. Comparisons. This paper compares GAP-Net with DCP [8], RPM-Net, and PREDATOR, and the experimental results are shown in Table 1. Obviously, GAP-Net outperforms existing methods on ModelNet. GAP-Net's RRE is reduced by 13.14% when compared to the next-best-performing

Table 1: Evaluation results on ModelNet and ModelLoNet.

Methods	ModelNet			ModelLoNet		
	RRE	RTE	CD	RRE	RTE	CD
DCP-v2 [8]	11.975	0.171	0.01170	16.501	0.300	0.0268
RPM-Net [9]	<u>1.712</u>	<u>0.018</u>	<u>0.00085</u>	7.342	<u>0.124</u>	**0.0050**
PREDATOR [15]	1.739	0.019	0.00089	<u>5.235</u>	0.132	0.0083
This study	**1.487**	**0.015**	**0.00079**	**4.339**	**0.114**	<u>0.0059</u>

Note. The best performance is highlighted in bold, while the next-best performance is underlined.

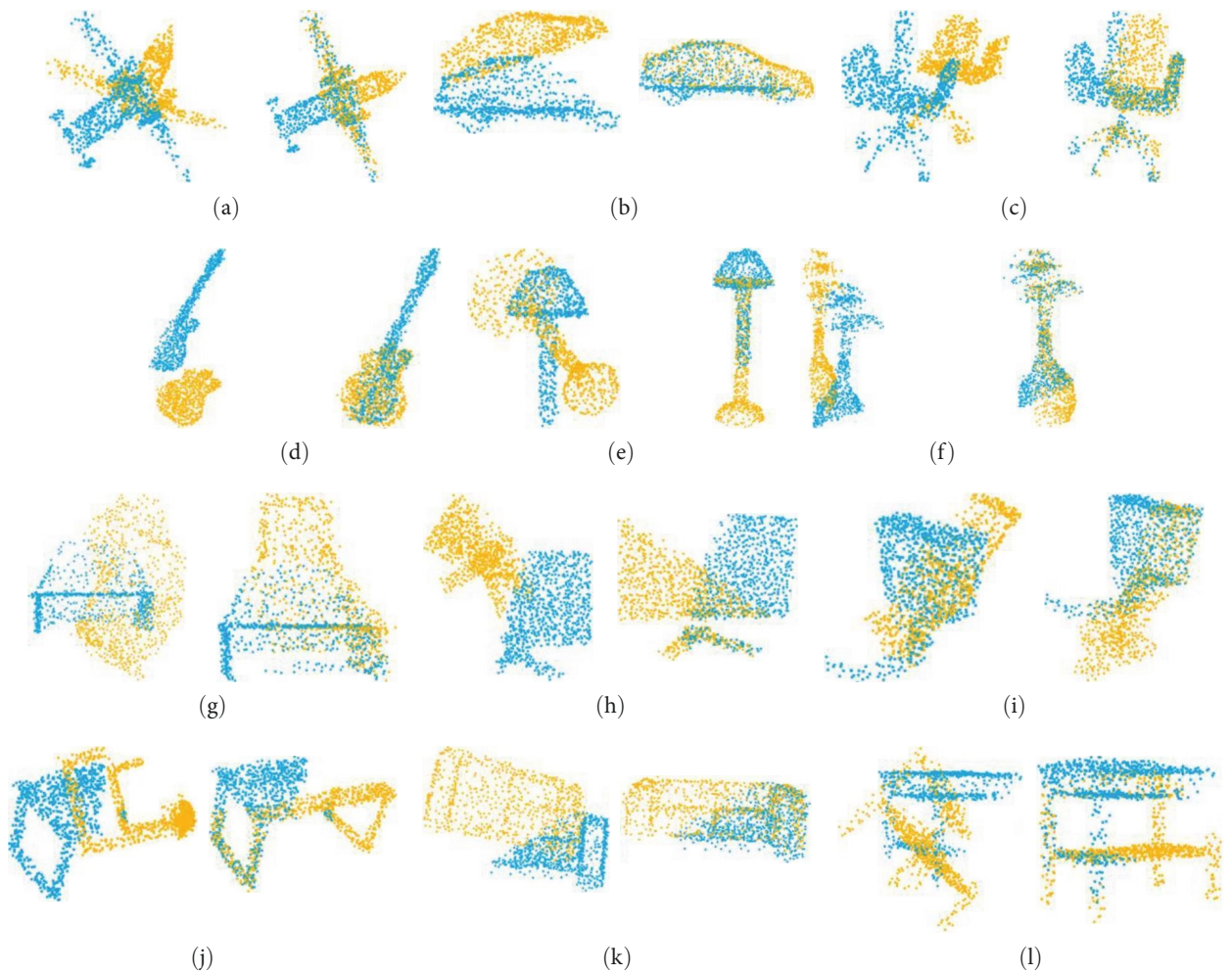

Figure 7: Example results of GPA-Net on partially visible data. The source point cloud is yellow, and the target point cloud is blue. (a)–(f) are cropped to preserve 70% of the original point cloud, and (g)–(l) are cropped to preserve 50% of the original point cloud. (a) airplane, (b) car, (c) chair, (d) guitar, (e) lamp, (f) vase, (g) bottle, (h) laptop, (i) toilet, (j) table, (k) sofa, and (l) bed.

RPM-Net on ModelNet. Furthermore, on the low-overlap ModelLoNet dataset, it not only outperforms RPM-Net, a method specially tuned for ModelNet, in terms of rotation–translation error by a large margin, but also outperforms PREDATOR, a method specially tuned for low-overlap point cloud registration. GAP-Net's RRE is reduced by 17.12% when compared to the next-best-performing PREDATOR on ModelLoNet. This shows that GAP-Net is state-of-the-art in partial registration, especially robust in low-overlap states. Example results of our method on partially visible data are shown in Figure 7.

4.1.4. Relative Overlap Rate. In order to test the registration performance of GAP-Net under different overlap rates, this paper conducts a set of experiments with different cropping rates on the ModelNet40 complete point cloud dataset. The cropping retention rate ranges from 70% to 40% for a total of seven. There are 1,266 test pairs in each group, and the test results are shown in Figure 8. The results show that the registration performance of all three networks is at a high level when the crop retention rate is reduced from 70% to 60%. When the crop retention rate is reduced to 50%, the registration performance of RPM-Net is already significantly

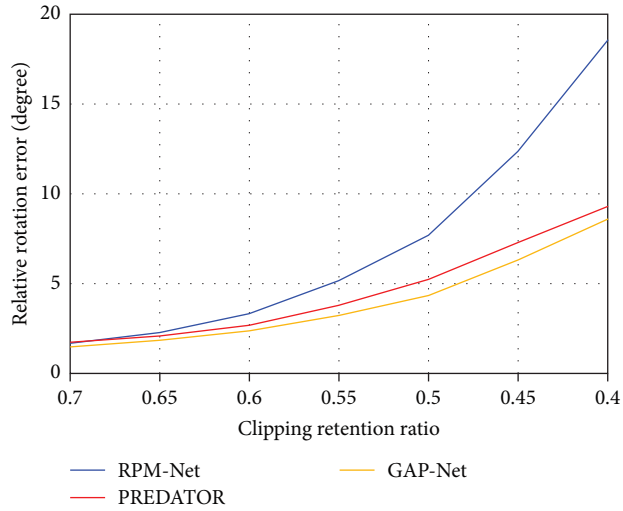

Figure 8: Relative rotation error of registration at different relative overlap ratios.

lower than that of GAP-Net and PREDATOR. When the crop retention rate is reduced from 50% to 40%, the RPM-Net error increases sharply. This is because after 50% and above cropping of cloud pairs, the proportion of overlapping regions decreases sharply, and some even have no overlapping regions. In this case, the extracted features have poor recognition ability, which confuses the model. The relative rotation errors of GAP-Net and PREDATOR can also be controlled within 10; GAP-Net is relatively better. In conclusion, GAP-Net outperforms state-of-the-art RPM-Net and PREDATOR in partial registration of ModelNet40 and is robust to changes in crop retention. Among them, the performance of RPM-Net's partial registration of the point cloud decreases rapidly with the reduction of the clipping retention ratio, that is, the reduction of the relative overlap rate.

4.1.5. Ablations Study. To better understand the importance of the SSA components and the proposed OGA module, this paper conducts module ablation experiments on these two modules on the ModelNet and ModelLoNet datasets. The experimental results are shown in Table 2. GAP-Net is first compared with a baseline model in which the SSA component and the proposed OGA module are completely removed. The error achieves an RRE of 1.91° and 5.405° in the baseline model test. By adding the SSA component, the RRE is reduced by 0.116° and 0.552° on ModelNet and ModelLoNet, and the error is reduced by 6.07% and 10.21%, respectively. This indicates that GAP-Net benefits from the spatial self-attention aggregation (SSA) module, which effectively utilizes the internal and global information of each point cloud at different levels, so the three metrics on both datasets can achieve better performance. Taking this as a new baseline model, three different combinations of GSA and CA components in the OGA module were added, respectively. The combination of GSA and CA achieved the errors of RRE, RTE, and CD on the ModelNet dataset, which were only higher than those of GAP-Net. The gap with other better-performing combined metrics on the ModelLoNet dataset is also small, suggesting that the GSA component used to update the local context before upsampling further improves performance. In addition, compared with the new model, only adding the GSA component in the OGA module reduces the RRE by 0.158° and 0.056° on ModelNet and ModelLoNet, respectively, and adding the CA component reduces the RRE by 0.056° and 0.1° on ModelNet and ModelLoNet, respectively. This suggests that the self-attention mechanism GSA component guided by geometric encoding fuses the extracted features to further enhance their contextual relationship, and the CA component for mixing the feature information of the two point clouds to obtain information of potential overlapping regions are all improved network performance to some extent. Therefore, combining these four parts together, the GAP-Net, can achieve the best overall performance.

4.2. Stanford 3D Scanning

4.2.1. Dataset. This paper uses the Stanford 3D scanning dataset to test the generalization of GAP-Net. Compared to the synthetic ModelNet40 dataset, it is a real-world dataset. For partial registration, the partially overlapping point clouds are generated by randomly cropping about 30% of the points in different directions from two identical point clouds, and then the source and target point clouds are generated by rotation and translation for testing. The model trained on the ModelNet40 dataset is directly used here.

4.2.2. Metrics and Experiments. This paper uses RRE and RTE to evaluate the registration effect. The registration results are shown in Figure 9. Obviously, from the registration effect maps, although these object categories did not appear during training, GAP-Net can still perform very well on objects in the Stanford dataset. From both RRE and RTE, the registration errors on objects in the Stanford dataset are within the test error range on Model and ModelLoNet. This shows that our method has good generalization.

4.3. DMatch and 3DLoMatch

4.3.1. Implementation Details. Due to the large scale of the 3DMatch indoor scene point cloud, a group of basic convolution blocks are added at the front end of the network, and corresponding upsampling layers are added to increase the number of network layers to extract features. The experiment was performed on a computer with Intel(R) Core(TM) i9-10980XE CPU @3.00 GHz and NVIDIA GeForce RTX 3090 GPU.

4.3.2. Dataset. 3DMatch contains 62 scenes, of which 46 are for training, eight for validation, and eight for testing. This paper conducts experiments using 3DMatch and 3DLoMatch preprocessed in PREDATOR [15], which contain >30% and 10%–30% partially overlapping scene pairs, respectively. This paper adopts registration recall (RR) as the main metric, since RR corresponds to the actual goal of point cloud registration. RR is the fraction of point cloud pairs for which the root mean square error of the estimated transformation compared to the ground truth is less than 0.2.

Table 2: Ablation of the network architecture.

SSA	OGA			ModelNet			ModelLoNet		
	GSA	CA	GSA	RRE	RTE	CD	RRE	RTE	CD
				1.910	0.0219	0.000983	5.405	0.128	0.00681
√				1.794	0.0205	0.000922	4.853	0.126	0.00671
√	√			1.636	0.0167	0.000854	4.797	0.119	0.00599
√	√	√		1.580	0.0159	0.000807	4.697	0.120	0.00611
√	√		√	1.673	0.0175	0.000853	4.608	0.123	0.00645
√	√	√	√	**1.487**	**0.0148**	**0.000793**	**4.339**	**0.114**	**0.00589**

Note. The best performance is highlighted in bold, while the next-best performance is underlined.

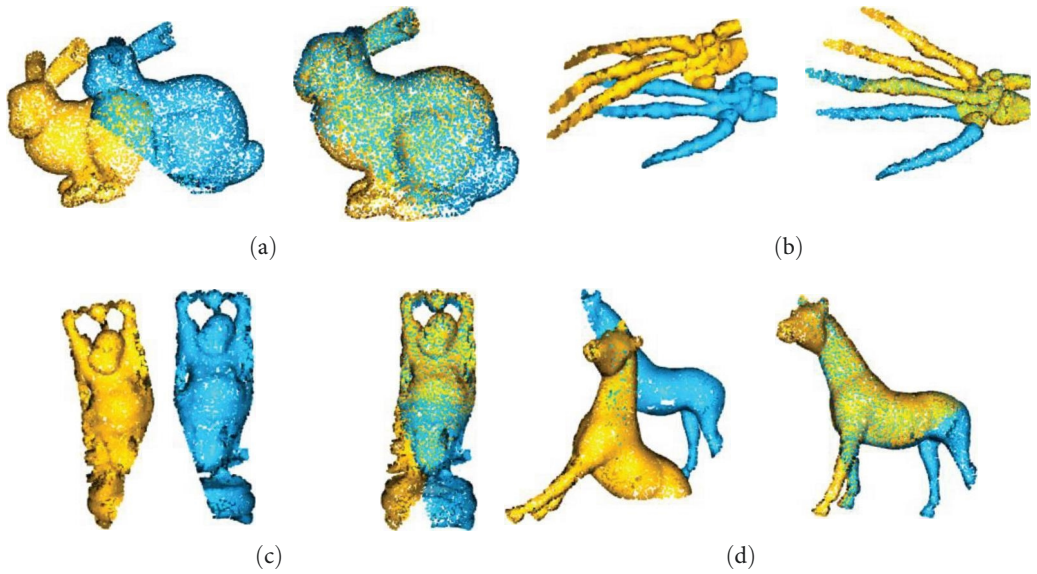

Figure 9: Example results of GAP-Net trained on ModelNet40 and applied to Stanford data. The objects (a)–(d): bunny, hand, happy Buddha, and horse; on the left is the initial position of the two point clouds, on the right is the registration result, and the rotation and translation error of the registration are calculated. (a) RRE = 3.263, RTE = 0.0075; (b) RRE = 0.282, RTE = 0.0183; (c) RRE = 2.232, RTE = 0.0057; (d) RRE = 0.929, RTE = 0.0012.

Table 3: Results on 3DMatch and 3DLoMatch datasets with different sample sizes.

#Samples	3DMatch					3DLoMatch				
	5,000	2,500	1,000	500	250	5,000	2,500	1,000	500	250
Registration recall (%)										
3DSN [36]	78.4	76.2	71.4	67.6	50.8	33.0	29.0	23.3	17.0	11.0
FCGF [12]	85.1	84.7	83.3	81.6	71.4	40.1	41.7	38.2	35.4	26.8
D3Feat [13]	81.6	84.5	83.4	82.4	77.9	37.2	42.7	46.9	43.8	39.1
PREDATOR [15]	**89.0**	**89.9**	**90.6**	88.5	**86.6**	**59.8**	**61.2**	**62.4**	**60.8**	**58.1**
This study	**89.0**	88.8	89.3	**88.7**	86.0	56.1	57.5	57.8	56.3	53.5

Note. The best performance is highlighted in bold, while the next-best performance is underlined. FCGF, fully convolutional geometric features.

4.3.3. Comparisons. This paper compares GAP-Net with other feature-based registration methods: 3DSN [39], FCGF [12], D3Feat [13], and PREDATOR, as shown in Table 3. From the results, our GAP-Net performs only slightly worse than PREDATOR on 3DMatch and 3DLoMatch. It is not significantly different from PREDATOR, with registration recall being only 1.3% lower at worst. It is significantly worse than PREDATOR on 3DLoMatch, but the gap in registration recall is also in the 5% range. GAP-Net still performs better than other feature-based registration methods. An example result of our method on partially visible data is shown in Figure 10, and the registration effect is still ideal in the partial registration of numerous scenes.

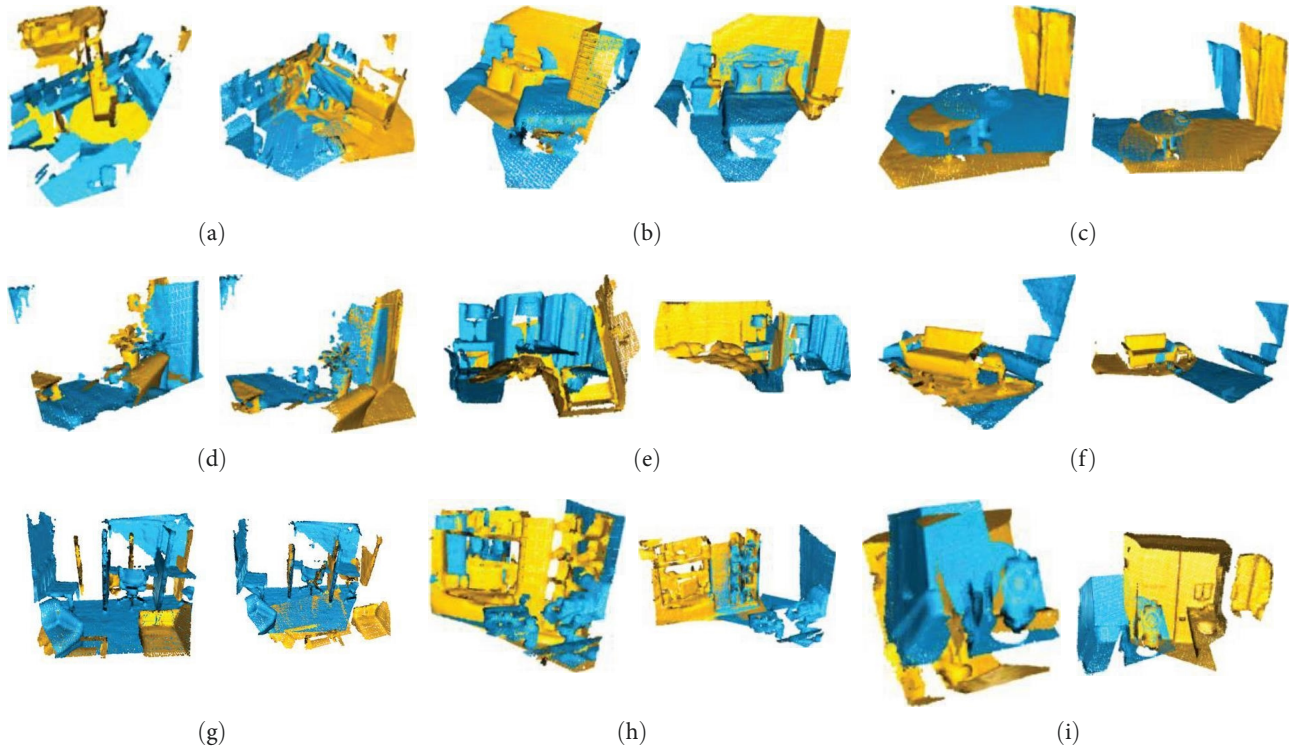

Figure 10: Example results of GAP-Net on partially visible data. The source point cloud is yellow, and the target point cloud is blue. (a)–(d) are partially overlapping scene pairs containing >30%; (e)–(i) are partially overlapping scene pairs containing 10%–30%. (a) Overlap: 0.4532; (b) overlap: 0.3713; (c) overlap: 0.3625; (d) overlap: 0.3301; (e) overlap: 0.2912; (f) overlap: 0.2811; (g) overlap: 0.2314; (h) overlap: 0.2113; (i) overlap: 0.1013.

5. Conclusion

This paper proposes GAP-Net, a partial point cloud registration network. A backbone network optimized using a spatial attention module is proposed to efficiently utilize the internal and global information of each point cloud at different levels. This paper also proposes an overlapping attention module based on geometric information for inferring points in overlapping regions. Experiments on the point cloud data of the ModelNet and ModelLoNet models show that our model has higher registration accuracy compared to state-of-the-art methods. In addition, the experiments on 3DMatch and 3DLoMatch scene point cloud data show that our method is also applicable for large-scale scene partial point cloud registration. In future work on this paper, we will further discuss how to adaptively select geometric information for different types of point cloud data, so that it can have better performance on scene point cloud data.

Data Availability

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We gratefully acknowledge that this work was supported in part by the National Natural Science Foundation of China under grant no. 61960206010 and the Science and Technology Support Program of Sichuan Province under grant no. 2021YJ0080 for providing the project.

References

[1] W. Lu, Y. Zhou, G. Wan, S. Hou, and S. Song, "L3-net: towards learning based LiDAR localization for autonomous driving," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6389–6398, IEEE, Long Beach, CA, USA, 2019.

[2] F. Pomerleau, F. Colas, and R. Siegwart, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends in Robotics*, vol. 4, no. 1, pp. 1–104, 2015.

[3] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, vol. 1611, pp. 586–606, SPIE, 1992.

[4] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: efficient and robust 3D object recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 998–1005, IEEE, San Francisco, CA, USA, 2010.

[5] F. Tombari, S. Salti, and L. Di Stefano, "Unique shape context for 3D data description," in *3DOR '10: Proceedings of the ACM workshop on 3D object retrieval*, pp. 57–62, Association for Computing Machinery, 2010.

[6] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," *International Journal of Computer Vision*, vol. 105, no. 1, pp. 63–86, 2013.

[7] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "Pointnetlk: robust & efficient point cloud registration using pointnet," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7163–7172, IEEE, 2019.

[8] Y. Wang and J. M. Solomon, "Deep closest point: learning representations for point cloud registration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3523–3532, IEEE, 2019.

[9] Z. J. Yew and G. H. Lee, "RPM-net: robust point matching using learned features," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11824–11833, IEEE, 2020.

[10] Y. Wang and J. M. Solomon, "PRNet: self-supervised learning for partial-to-partial registration," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[11] 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*, pp. 195–205, IEEE, Salt Lake City, UT, USA, 2018.

[12] C. Choy, J. Park, and V. Koltun, "Fully convolutional geometric features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8958–8966, IEEE, 2019.

[13] X. Y. Bai, Z. X. Luo, L. Zhou, H. B. Fu, L. Quan, and C. L. Tai, "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*, pp. 6359–6367, IEEE, 2020.

[14] H. Thomas, C. R. Qi, J. E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "KPConv: flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6411–6420, IEEE, 2019.

[15] S. Y. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler, "PREDATOR: registration of 3D point clouds with low overlap," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4267–4276, IEEE, 2021.

[16] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 6000–6010, 2017.

[17] 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*, pp. 3212–3217, IEEE, 2009.

[18] D. Aiger, N. J. Mitra, and D. Cohen-Or, "4-Points congruent sets for robust pairwise surface registration," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 1–10, 2008.

[19] N. Mellado, D. Aiger, and N. J. Mitra, "Super 4PCS fast global pointcloud registration via smart indexing," *Computer Graphics Forum*, vol. 33, no. 5, pp. 205–215, 2014.

[20] 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, 2016.

[21] 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)*, pp. 2743–2748, IEEE, Las Vegas, NV, USA, 2003.

[22] A. Myronenko and X. Song, "Point set registration: coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[23] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[24] F. Tombari and L. Di Stefano, "Object recognition in 3D scenes with occlusions and clutter by hough voting," in *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*, pp. 349–355, IEEE, Singapore, 2010.

[25] 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, 2018.

[26] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics*, vol. 38, no. 5, pp. 1–12, 2019.

[27] J. Zhou, M. J. Wang, W. D. Mao, M. L. Gong, and X. P. Liu, "SiamesePointNet: a siamese point network architecture for learning 3D shape descriptor," *Computer Graphics Forum*, vol. 39, no. 1, pp. 309–321, 2020.

[28] Z. Dang, F. Wang, and M. Salzmann, "Learning 3D-3D correspondences for one-shot partial-to-partial registration," arXiv preprint arXiv: 2006.04523, 2020.

[29] H. Xu, S. C. Liu, G. F. Wang, G. H. Liu, and B. Zeng, "Omnet: learning overlapping mask for partial-to-partial point cloud registration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3132–3141, IEEE, 2021.

[30] L. F. Zhu, D. R. Liu, C. W. Lin et al., "Point cloud registration using representative overlapping points," arXiv preprint arXiv: 2107.02583, 2021.

[31] R. Zhou, X. Li, and W. Jiang, "SCANet: a spatial and channel attention based network for partial-to-partial point cloud registration," *Pattern Recognition Letters*, vol. 151, pp. 120–126, 2021.

[32] L. Li, Y. Xie, L. Cen, and Z. Zeng, "A novel cause analysis approach of grey reasoning Petri net based on matrix operations," *Applied Intelligence*, vol. 52, pp. 1–18, 2022.

[33] R. Zhou, H. Wang, X. Li, Y. Guo, C. Dai, and W. Jiang, "MaskNet++: inlier/outlier identification for two point clouds," *Computers & Graphics*, vol. 103, pp. 90–100, 2022.

[34] Y. Wang, C. Yan, Y. Feng, S. Du, Q. Dai, and Y. Gao, "STORM: structure-based overlap matching for partial point cloud registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 1135–1149, 2023.

[35] H. Zhao, H. Zhuang, C. Wang, and M. Yang, "G3DOA: generalizable 3D descriptor with overlap attention for point cloud registration," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2541–2548, 2022.

[36] H. Yu, F. Li, M. Saleh, B. Busam, and S. Ilic, "Cofinet: reliable coarse-to-fine correspondences for robust pointcloud registration," *Advances in Neural Information Processing Systems*, vol. 34, pp. 23872–23884, 2021.

[37] Z. Qin, H. Yu, C. J. Wang, Y. L. Guo, Y. X. Peng, and K. Xu, "Geometric transformer for fast and robust point cloud registration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11143–11152, IEEE, 2022.

[38] F. Lu, G. Chen, Y. L. Liu, Z. N. Qu, and A. Knoll, "Rskdd-net: random sample-based keypoint detector and descriptor," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21297–21308, 2020.

[39] Z. Gojcic, C. F. Zhou, J. D. Wegner, and A. Wieser, "The perfect match: 3D point cloud matching with smoothed densities," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5545–5554, IEEE, 2019.