

# 北京理工大学

## 本科生毕业设计（论文）外文翻译

外文原文题目： Geometric Transformer for Fast and

Robust Point Cloud Registration

中文翻译题目： Geometric Transformer: 快速且稳健的点云配准

**基于深度学习和聚类优化的多实例点云配准**

**Optimized Multi-Instance Point Cloud Registration based on  
Deep Learning and Cluster Methods**

学 院： 自动化学院

专 业： 自动化

班 级： 06011902

学生姓名： 杨润一

学 号： 1120191211

指导教师： 由育阳

## Geometric Transformer: 快速且稳健的点云配准

### 摘要

我们研究了如何提取精确的点云配准对应关系的问题。最近的无关键点方法绕过了在低重叠场景中检测可重复关键点的难题，这在配准中显示出巨大的潜力。它们在下采样的超点上寻找对应关系，然后将这些对应关系传播到密集点。这些超点基于其相邻区域是否重叠进行匹配。这种稀疏和松散的匹配需要捕获点云的几何结构的上下文特征。我们提出了几何变换器来学习用于稳健超点匹配的几何特征。它编码了成对距离和三元角度，使其在低重叠情况下稳健，且对刚性变换具有不变性。这种简单的设计令人惊讶地获得了高的匹配精度，以至于在估计对齐变换时不需要 RANSAC，从而加速了 100 倍。我们的方法在具有挑战性的 3DLoMatch 基准测试中将内点比率提高了 17~30 个百分点，将配准召回率提高了超过 7 个百分点。

关键词：点云配准；描述子；特征描述；深度学习；毕业设计（外文翻译）

## 目 录

摘要 .....	I
第1章 引言 .....	1
第2章 相关工作 .....	3
2.1 基于对应关系的方法。 .....	3
2.2 直接配准方法。 .....	3
2.3 深度稳健估计器。 .....	3
第3章 方法 .....	4
3.1 超点采样和特征提取 .....	4
3.2 超点匹配模块 .....	5
3.2.1 Geometric Transformer. ....	5
3.2.2 几何自注意力. ....	6
3.2.3 基于特征的交叉注意力. ....	7
3.2.4 超点匹配. ....	8
3.2.5 点匹配模块 .....	8
3.2.6 无需 RANSAC 的局部到全局配准 .....	9
3.2.7 损失函数 .....	10
第4章 实验分析 .....	12
4.1 室内基准测试：3DMatch 和 3DLoMatch .....	12
4.1.1 数据集 .....	12
4.1.2 指标 .....	12
4.1.3 对应性结果 .....	12
4.1.4 配准结果 .....	12
4.1.5 消融实验 .....	14
4.1.6 定性结果 .....	16
4.2 室外基准测试：KITTI odometry .....	16
4.2.1 数据集 .....	16
4.2.2 评估指标 .....	16
4.2.3 配准结果。 .....	17
第5章 总结 .....	18
参考文献 .....	19

## 第 1 章 引言

我们研究了用于点云配准的精确对应关系的提取问题。近期的进步主要由基于学习的，基于对应关系的方法主导<sup>[1-6]</sup>。这些方法训练神经网络从两个输入的点云中提取点对应关系，然后基于此使用稳健的估计器，例如 RANSAC，计算对齐变换。大多数基于对应关系的方法依赖于关键点的检测<sup>[3-5,7]</sup>。然而，在两个点云中，特别是当它们的重叠区域很小时，检测可重复的关键点是具有挑战性的。这通常导致在推测的对应关系中的内点比率较低。

受到近期图像匹配进展的启发<sup>[8-10]</sup>，无关键点方法<sup>[6]</sup>将输入的点云下采样成超点，并通过检查它们的局部邻域（区块）是否重叠进行匹配。这种超点（区块）匹配然后传播到单独的点，产生密集的点对应关系。因此，密集点对应关系的准确性高度依赖于超点匹配的准确性。

超点匹配是稀疏和松散的。其优点是它将严格的点匹配降低到松散的区块重叠，从而放宽了可重复性的要求。同时，区块重叠比基于距离的点匹配更可靠和更有信息量，这对于学习对应关系是一个重要的约束。另一方面，超点匹配需要捕获更多的全局上下文。

为此，Transformer<sup>[11]</sup>已被采用<sup>[6,12]</sup>以在点云配准中编码上下文信息。然而，普通的 Transformer 忽视了点云的几何结构，这使得学习的特征在几何上不够区分，并导致大量的异常匹配。尽管我们可以注入位置嵌入<sup>[13-14]</sup>，但基于坐标的编码是变换不变的，这在配准给定任意姿态的点云时是有问题的。

我们主张，为配准任务学习的点 Transformer 应该使用点云的几何结构，以便提取变换不变的几何特征。我们提出了针对 3D 点云的 Geometric Transformer（简称 GeoTransformer），它仅编码点对的距离和点三元组的角度。

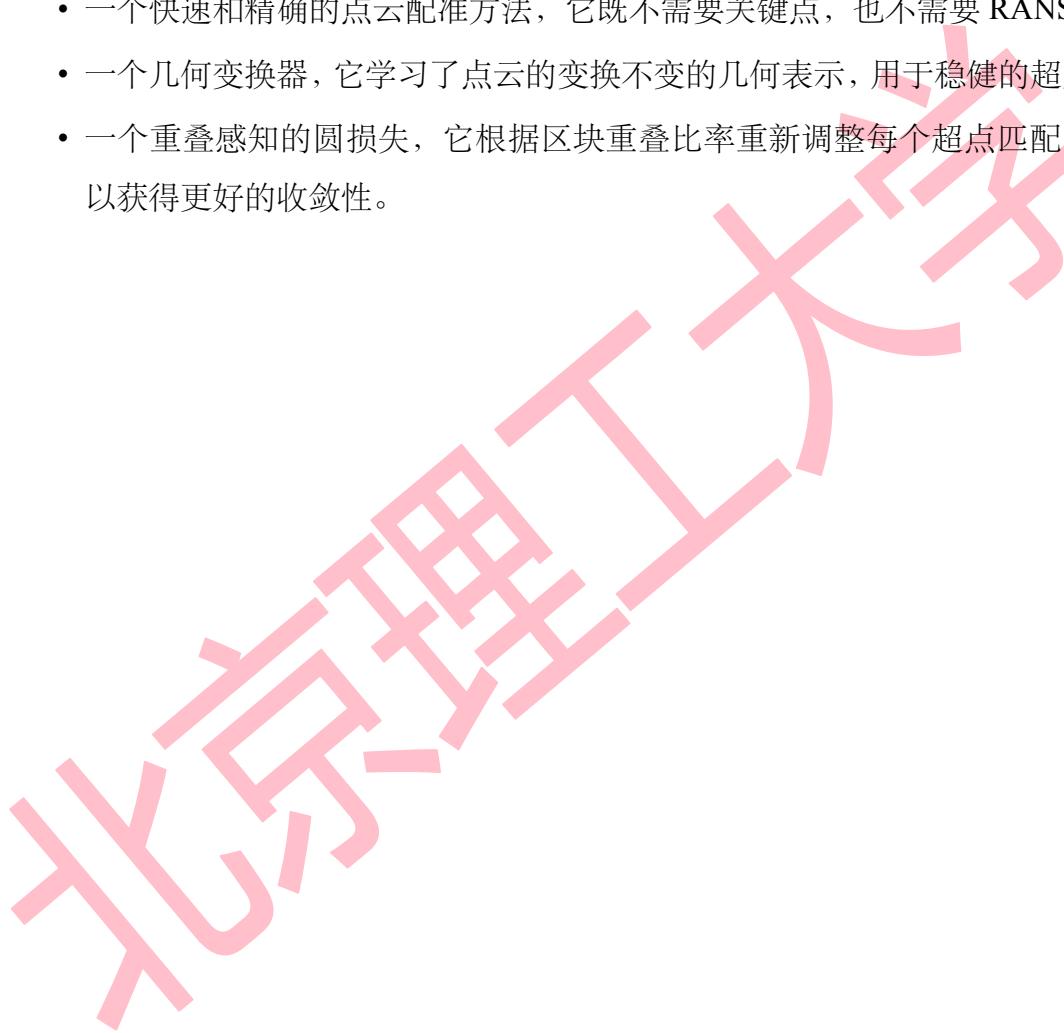
对于一个超点，我们通过在所有其他超点之间的距离和角度的基础上几何地“定位”它来学习一个非局部表示。自注意力机制被用来权衡那些锚定超点的重要性。由于距离和角度对刚性变换是不变的，GeoTransformer 有效地学习点云的几何结构，即使在低重叠的情况下也能进行高度稳健的超点匹配。

受益于高质量的超点匹配，我们的方法利用最优传输层获得高内点比率的密集点对应关系，以及高度稳健和精确的配准，而无需依赖 RANSAC。因此，我们方法的

配准部分运行非常快，例如，对于具有 5K 对应关系的两个点云，只需要 0.01 秒，比 RANSAC 快 100 倍。在室内和室外的基准测试上的大量实验都证明了 GeoTransformer 的效果。我们的方法在 3DLoMatch 基准测试上将内点比率提高了 17 到 30 个百分点，将配准召回率提高了超过 7 个百分点。

我们的主要贡献包括：

- 一个快速和精确的点云配准方法，它既不需要关键点，也不需要 RANSAC。
- 一个几何变换器，它学习了点云的变换不变的几何表示，用于稳健的超点匹配。
- 一个重叠感知的圆损失，它根据区块重叠比率重新调整每个超点匹配的损失，以获得更好的收敛性。



## 第 2 章 相关工作

### 2.1 基于对应关系的方法。

我们的工作遵循基于对应关系的方法的思路<sup>[1-3,15]</sup>。它们首先提取两个点云之间的对应关系，然后使用稳健的姿态估计器，例如 RANSAC，恢复变换。由于稳健的估计器，它们在室内和室外场景配准中取得了最新的性能。这些方法可以根据它们提取对应关系的方式进一步分类为两类。第一类旨在检测更可重复的关键点<sup>[4-5]</sup>并为关键点学习更强大的描述符<sup>[3,7,16]</sup>。而第二类<sup>[6]</sup>通过考虑所有可能的匹配而不需要检测关键点来检索对应关系。我们的方法遵循无需检测的方法，并通过利用几何信息来提高对应关系的准确性。

### 2.2 直接配准方法。

最近，直接配准方法已经出现。他们以端到端的方式使用神经网络估计变换。这些方法可以进一步分为两类。第一类<sup>[12,17-19]</sup>遵循 ICP<sup>[20]</sup>的思路，该方法迭代地建立软对应关系，并使用可微分的加权 SVD 计算变换。第二类<sup>[21-23]</sup>首先为每个点云提取一个全局特征向量，并使用全局特征向量回归变换。虽然直接配准方法在单个合成形状上取得了有希望的结果，但在大规模场景中它们可能会失败，如<sup>[5]</sup>所述。

### 2.3 深度稳健估计器。

由于传统的稳健估计器，如 RANSAC 在高离群值比率的情况下会出现收敛速度慢和不稳定的问题，因此提出了深度稳健估计器<sup>[24-26]</sup>作为替代方案。它们通常包含一个分类网络来拒绝离群值和一个估计网络来计算变换。与传统的稳健估计器相比，它们在准确性和速度上都有所改进。然而，它们需要训练一个特定的网络。相比之下，我们的方法通过一个无参数的局部到全局的配准方案实现了快速和精确的配准。

### 第 3 章 方法

给定两个点云  $\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}$  和  $\mathcal{Q} = \{\mathbf{q}_i \in \mathbb{R}^3 \mid i = 1, \dots, M\}$ , 我们的目标是估计一个刚性变换  $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$  来对齐这两个点云, 其中  $\mathbf{R} \in SO(3)$  是 3D 旋转,  $\mathbf{t} \in \mathbb{R}^3$  是 3D 平移。变换可以通过以下方式求解:

$$\min_{\mathbf{R}, \mathbf{t}} \sum_{(\mathbf{p}_{x_i}^*, \mathbf{q}_{y_i}^*) \in \mathcal{C}^*} \|\mathbf{R} \cdot \mathbf{p}_{x_i}^* + \mathbf{t} - \mathbf{q}_{y_i}^*\|_2^2. \quad (3-1)$$

在这里,  $\mathcal{C}^*$  是  $\mathcal{P}$  和  $\mathcal{Q}$  之间的真实对应关系集合。由于在实际中我们无法知道  $\mathcal{C}^*$ , 所以我们需要首先在两个点云之间建立点对应关系, 然后估计对齐变换。

我们的方法采用分层对应关系模式, 以从粗到细的方式找到对应关系。我们采用 KPConv-FPN 同时对输入点云进行下采样并提取点特征 (3.1)。第一级和最后一级 (最粗糙的) 下采样点对应于需要匹配的密集点和超点。使用超点匹配模块提取超点对应关系, 其相邻的局部区域与彼此重叠 (3.2)。基于此, 点匹配模块进一步将超点对应关系细化到密集点 (3.2.5)。最后, 从密集对应关系中恢复对齐变换, 无需依赖 RANSAC (3.2.6)。流程图示在3-1中。

#### 3.1 超点采样和特征提取

我们利用 KPConv-FPN 主干<sup>[27-28]</sup>为点云提取多级特征。点特征学习的一个副产品是点下采样。我们在下采样点上进行工作, 因为点云配准实际上可以通过一组更粗糙的点的对应关系来确定。原始点云通常过于密集, 以至于点对点的对应关系是多余的, 有时甚至过于集中而无法使用。

对于最粗糙的分辨率的点, 由  $\hat{\mathcal{P}}$  和  $\hat{\mathcal{Q}}$  表示, 被视为需要匹配的超点。相关的学习特征被表示为  $\hat{\mathbf{F}}^{\mathcal{P}} \in \mathbb{R}^{|\hat{\mathcal{P}}| \times \hat{d}}$  和  $\hat{\mathbf{F}}^{\mathcal{Q}} \in \mathbb{R}^{|\hat{\mathcal{Q}}| \times \hat{d}}$ 。在原始分辨率的  $1/2$  处计算密集点对应关系, 即, 由  $\tilde{\mathcal{P}}$  和  $\tilde{\mathcal{Q}}$  表示的第一级下采样点。他们的学习特征由  $\tilde{\mathbf{F}}^{\mathcal{P}} \in \mathbb{R}^{|\tilde{\mathcal{P}}| \times \tilde{d}}$  和  $\tilde{\mathbf{F}}^{\mathcal{Q}} \in \mathbb{R}^{|\tilde{\mathcal{Q}}| \times \tilde{d}}$  表示。

对于每个超点, 我们使用点到节点分组策略<sup>[6,29]</sup>在其周围构建一个局部贴片。特别地,  $\tilde{\mathcal{P}}$  中的每个点及其来自  $\tilde{\mathbf{F}}^{\mathcal{P}}$  的特征都分配给几何空间中最近的超点:

$$\mathcal{G}_i^{\mathcal{P}} = \{\tilde{\mathbf{p}} \in \tilde{\mathcal{P}} \mid i = \arg \min_j (\|\tilde{\mathbf{p}} - \hat{\mathbf{p}}_j\|_2), \hat{\mathbf{p}}_j \in \hat{\mathcal{P}}\}. \quad (3-2)$$

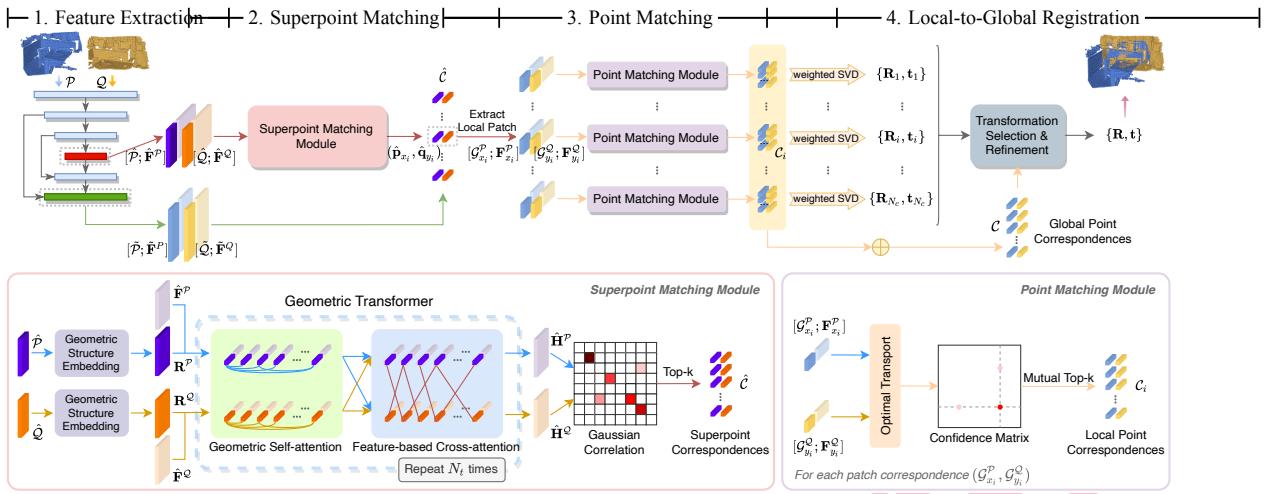


图 3-1 The backbone downsamples the input point clouds and learns features in multiple resolution levels. The Superpoint Matching Module extracts high-quality superpoint correspondences between  $\hat{P}$  and  $\hat{Q}$  using the Geometric Transformer which iteratively encodes intra-point-cloud geometric structures and inter-point-cloud geometric consistency. The superpoint correspondences are then propagated to dense points  $\tilde{P}$  and  $\tilde{Q}$  by the Point Matching Module. Finally, the transformation is computed with a local-to-global registration method.

这本质上导致了由超点生成的输入点云的 Voronoi 分解。与  $\mathcal{G}_i^P$  中的点相关的特征矩阵表示为  $\mathbf{F}_i^P \subset \tilde{\mathbf{F}}^P$ 。具有空贴片的超点将被移除。对于  $\mathcal{Q}$ , 计算并以类似方式表示贴片  $\{\mathcal{G}_i^Q\}$  和特征矩阵  $\{\mathbf{F}_i^Q\}$ 。

### 3.2 超点匹配模块

#### 3.2.1 Geometric Transformer.

全局上下文在许多计算机视觉任务中已被证明至关重要<sup>[6,10,30]</sup>。因此, Transformer 被用来利用全局上下文信息进行点云配准。然而, 现有的方法<sup>[5-6,12]</sup>通常只将高级点云特征提供给变换器, 并不明确地编码几何结构。这使得学习到的特征在几何上的区分度较低, 导致严重的匹配模糊和大量的异常匹配, 特别是在重叠度低的情况下。一种直接的解决方案是明确注入 3D 点坐标的位姿嵌入<sup>[13-14]</sup>。然而, 由此产生的基于坐标的变换器自然是变换变化的, 而配准需要变换不变性, 因为输入点云可以处于任意姿态。

为此, 我们提出了 *Geometric Transformer*, 它不仅编码高级点特征, 而且明确捕捉点云内部的几何结构和点云间的几何一致性。GeoTransformer 由一个几何自注意模块组成, 用于学习点云内部特征, 以及一个基于特征的交叉注意模块, 用于建模点云间的一致性。这两个模块交错进行  $N_t$  次, 以提取混合特征  $\hat{\mathbf{H}}^P$  和  $\hat{\mathbf{H}}^Q$ , 用于可靠的超点匹配 (见3-1 (左下角))。

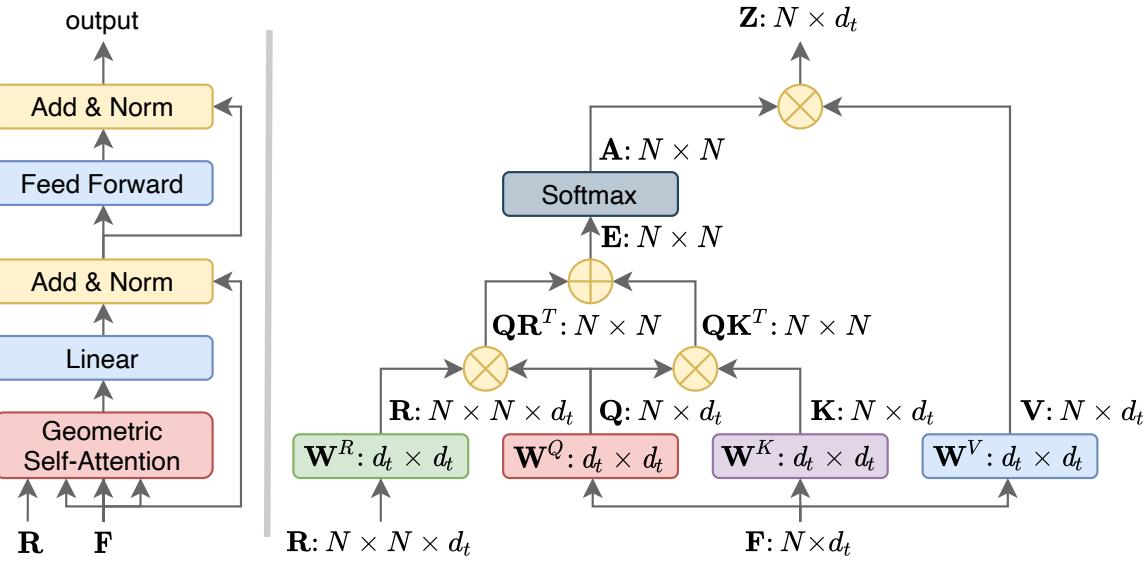


图 3-2 Left: The structure of geometric self-attention module. Right: The computation graph of geometric self-attention.

### 3.2.2 几何自注意力.

我们设计了一个几何自注意力机制，以学习每个点云中超点在特征和几何空间中的全局关联性。以下，我们描述了对  $\hat{\mathcal{P}}$  的计算，对  $\hat{\mathcal{Q}}$  的计算也是相同的。给定输入特征矩阵  $\mathbf{X} \in \mathbb{R}^{|\hat{\mathcal{P}}| \times d_t}$ ，输出特征矩阵  $\mathbf{Z} \in \mathbb{R}^{|\hat{\mathcal{P}}| \times d_t}$  是所有投影输入特征的加权和：

$$\mathbf{z}_i = \sum_{j=1}^{|\hat{\mathcal{P}}|} a_{i,j} (\mathbf{x}_j \mathbf{W}^V), \quad (3-3)$$

其中，权重系数  $a_{i,j}$  是对注意力得分  $e_{i,j}$  进行行方向的 softmax 计算得到的， $e_{i,j}$  计算如下：

$$e_{i,j} = \frac{(\mathbf{x}_i \mathbf{W}^Q)(\mathbf{x}_j \mathbf{W}^K + \mathbf{r}_{i,j} \mathbf{W}^R)^T}{\sqrt{d_t}}. \quad (3-4)$$

这里， $\mathbf{r}_{i,j} \in \mathbb{R}^{d_t}$  是一个将在下文描述的几何结构嵌入。 $\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V, \mathbf{W}^R \in \mathbb{R}^{d_t \times d_t}$  分别是查询、键、值和几何结构嵌入的投影矩阵。3-2显示了几何自注意力的结构和计算。

我们设计了一种新颖的几何结构嵌入，用来编码超点的变换不变的几何结构。核心思想是利用与超点计算的距离和角度，这些是在同一场景的不同点云中保持一致的。给定两个超点  $\hat{\mathbf{p}}_i, \hat{\mathbf{p}}_j \in \hat{\mathcal{P}}$ ，它们的几何结构嵌入由一个成对距离嵌入和一个三元角嵌入组成，这些将在下面详细描述。

(1) 成对距离嵌入. 给定  $\hat{\mathbf{p}}_i$  和  $\hat{\mathbf{p}}_j$  之间的距离  $\rho_{i,j} = \|\hat{\mathbf{p}}_i - \hat{\mathbf{p}}_j\|_2$ ，它们之间的距离嵌入  $\mathbf{r}_{i,j}^D$  通过对  $\rho_{i,j}/\sigma_d$  应用正弦函数<sup>[11]</sup>来计算。这里， $\sigma_d$  是一个超参数，用来调整对

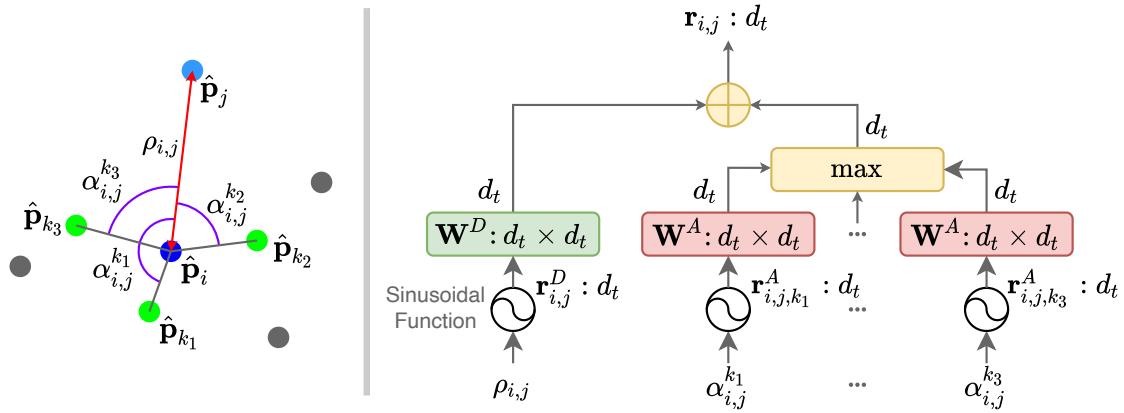


图 3-3 An illustration of the distance-and-angle-based geometric structure encoding and its computation.  
距离变化的敏感性。

(2) 三元角嵌入. 我们用超点的三元组计算角嵌入。首先，我们选择  $\hat{\mathbf{p}}_i$  的  $k$  个最近邻点  $\mathcal{K}_i$ 。对于每个  $\hat{\mathbf{p}}_x \in \mathcal{K}_i$ ，我们计算角度  $\alpha_{i,j}^x = \angle(\Delta_{x,i}, \Delta_{j,i})$ ，其中  $\Delta_{i,j} := \hat{\mathbf{p}}_i - \hat{\mathbf{p}}_j$ 。然后，三元角嵌入  $\mathbf{r}_{i,j,x}^A$  通过对  $\alpha_{i,j}^x / \sigma_a$  应用正弦函数来计算，其中  $\sigma_a$  控制对角度变化的敏感性。

最后，通过聚合成对距离嵌入和三元角嵌入来计算几何结构嵌入  $\mathbf{r}_{i,j}$ :

$$\mathbf{r}_{i,j} = \mathbf{r}_{i,j}^D \mathbf{W}^D + \max_x \{ \mathbf{r}_{i,j,x}^A \mathbf{W}^A \}, \quad (3-5)$$

其中， $\mathbf{W}^D, \mathbf{W}^A \in \mathbb{R}^{d_t \times d_t}$  分别是两种嵌入的投影矩阵。我们在这里使用最大池化来提高对由于自遮挡而导致的超点的不同最近邻的鲁棒性。3-3 显示了几何结构嵌入的计算。

### 3.2.3 基于特征的交叉注意力.

交叉注意力是点云配准任务的典型模块<sup>[5-6,12]</sup>，用于在两个输入点云之间进行特征交换。给定  $\hat{\mathcal{P}}$ ,  $\hat{\mathcal{Q}}$  的自注意力特征矩阵  $\mathbf{X}^P$ ,  $\mathbf{X}^Q$ ,  $\hat{\mathcal{P}}$  的交叉注意力特征矩阵  $\mathbf{Z}^P$  使用  $\hat{\mathcal{Q}}$  的特征计算:

$$\mathbf{z}_i^P = \sum_{j=1}^{|\hat{\mathcal{Q}}|} a_{i,j} (\mathbf{x}_j^Q \mathbf{W}^V). \quad (3-6)$$

同样， $a_{i,j}$  是对交叉注意力得分  $e_{i,j}$  进行行方向的 softmax 计算，而  $e_{i,j}$  是  $\mathbf{X}^P$  和  $\mathbf{X}^Q$  之间的特征相关性计算:

$$e_{i,j} = \frac{(\mathbf{x}_i^P \mathbf{W}^Q)(\mathbf{x}_j^Q \mathbf{W}^K)^T}{\sqrt{d_t}}. \quad (3-7)$$

$\mathcal{Q}$  的交叉注意力特征以相同的方式计算。虽然几何自注意力模块为每个单独的点云编码了变换不变的几何结构，基于特征的交叉注意力模块可以在两个点云之间建模几何一致性。所得到的混合特征既不变于变换，又对推理对应关系具有鲁棒性。

### 3.2.4 超点匹配.

为了找到超点的对应关系，我们提出了一种基于全局特征相关性的匹配方案。我们首先将  $\hat{\mathbf{H}}^P$  和  $\hat{\mathbf{H}}^Q$  归一化到单位超球面，并计算一个高斯相关矩阵  $\mathbf{S} \in \mathbb{R}^{|\hat{P}| \times |\hat{Q}|}$ ，其中  $s_{i,j} = \exp(-\|\hat{\mathbf{h}}_i^P - \hat{\mathbf{h}}_j^Q\|_2^2)$ 。在实践中，点云的一些区域在几何上较不具有区分性，并且在另一个点云中有许多相似的区域。除了我们强大的混合特征外，我们还对  $\mathbf{S}$  进行双向归一化操作<sup>[8,10]</sup>，进一步抑制模糊的匹配，得到  $\bar{\mathbf{S}}$ ：

$$\bar{s}_{i,j} = \frac{s_{i,j}}{\sum_{k=1}^{|\hat{Q}|} s_{i,k}} \cdot \frac{s_{i,j}}{\sum_{k=1}^{|\hat{P}|} s_{k,j}}. \quad (3-8)$$

我们发现这种抑制可以有效地消除错误的匹配。最后，我们选择  $\bar{\mathbf{S}}$  中最大的  $N_c$  个项作为超点对应关系：

$$\hat{\mathcal{C}} = \{(\hat{\mathbf{p}}_{x_i}, \hat{\mathbf{q}}_{y_i}) \mid (x_i, y_i) \in \text{topk}_{x,y}(\bar{s}_{x,y})\}. \quad (3-9)$$

由于 GeoTransformer 的强大几何结构编码能力，我们的方法能够在低重叠情况和少量点对应关系下，最显著的是，以一种无需 RANSAC 的方式实现准确的配准。

### 3.2.5 点匹配模块

得到超点对应关系后，我们使用一个简单而有效的点匹配模块来提取点对应关系。在点级别，我们只使用由骨干网络学习的局部点特征。其原理是，一旦通过超点匹配解决了全局的模糊性，点级别的匹配主要由两个匹配点的邻域决定。这种设计选择提高了鲁棒性。

对于每个超点对应关系  $\hat{\mathcal{C}}_i = (\hat{\mathbf{p}}_{x_i}, \hat{\mathbf{q}}_{y_i})$ ，我们使用一个最优传输层<sup>[31]</sup> 来提取  $\mathcal{G}_{x_i}^P$  和  $\mathcal{G}_{y_i}^Q$  之间的局部稠密点对应关系。具体来说，我们首先计算一个成本矩阵  $\mathbf{C}_i \in \mathbb{R}^{n_i \times m_i}$ ：

$$\mathbf{C}_i = \mathbf{F}_{x_i}^P (\mathbf{F}_{y_i}^Q)^T / \sqrt{d}, \quad (3-10)$$

其中  $n_i = |\mathcal{G}_{x_i}^P|$ ,  $m_i = |\mathcal{G}_{y_i}^Q|$ 。然后，我们将成本矩阵  $\mathbf{C}_i$  通过追加一个新行和一个新

列扩增为  $\bar{\mathbf{C}}_i$ , 新行和新列的值由一个可学习的垃圾箱参数  $\alpha$  填充, 如文献<sup>[31]</sup> 中所述。然后我们利用 Sinkhorn 算法<sup>[32]</sup> 在  $\bar{\mathbf{C}}_i$  上计算一个软分配矩阵  $\bar{\mathbf{Z}}_i$ , 然后通过丢弃最后一行和最后一列恢复为  $\mathbf{Z}_i$ 。我们使用  $\mathbf{Z}_i$  作为候选匹配的置信度矩阵, 并通过互相选择前  $k$  个进行点对应关系的提取, 其中, 如果一个点匹配位于它所在的行和列的  $k$  个最大项中, 则选择该点匹配:

$$\mathcal{C}_i = \{(\mathcal{G}_{x_i}^P(x_j), \mathcal{G}_{y_i}^Q(y_j)) \mid (x_j, y_j) \in \text{mutual\_topk}_{x,y}(z_{x,y}^i)\}. \quad (3-11)$$

然后将每个超点匹配计算出的点对应关系收集到一起, 形成最终的全局稠密点对应关系:  $\mathcal{C} = \bigcup_{i=1}^{N_c} \mathcal{C}_i$ 。

### 3.2.6 无需 RANSAC 的局部到全局配准

先前的方法通常依赖于稳健的姿态估计器来估计变换, 因为假定的对应关系常常被离群值主导。大多数稳健估计器, 如 RANSAC, 收敛速度较慢。鉴于 GeoTransformer 的高内点比例, 我们能够实现稳健的配准, 而无需依赖稳健估计器, 这也大大降低了计算成本。

我们设计了一个局部到全局配准 (LGR) 方案。作为一种假设验证方法, LGR 包括一个局部阶段的变换候选生成和一个全局阶段的变换选择。在局部阶段, 我们使用其局部点对应关系为每个超点匹配求解一个变换  $\mathbf{T}_i = \{\mathbf{R}_i, \mathbf{t}_i\}$ :

$$\mathbf{R}_i, \mathbf{t}_i = \min_{\mathbf{R}, \mathbf{t}} \sum_{(\tilde{\mathbf{p}}_{x_j}, \tilde{\mathbf{q}}_{y_j}) \in \mathcal{C}_i} w_j^i \|\mathbf{R} \cdot \tilde{\mathbf{p}}_{x_j} + \mathbf{t} - \tilde{\mathbf{q}}_{y_j}\|_2^2. \quad (3-12)$$

这可以使用加权 SVD<sup>[20]</sup>以闭式解决。 $\mathbf{Z}_i$  中每个对应关系的相应置信度得分用作权重  $w_j^i$ 。由于对应关系的高质量, 这个阶段获得的变换已经非常准确。在全局阶段, 我们选择在整个全局点对应关系中接受最多内点匹配的变换:

$$\mathbf{R}, \mathbf{t} = \max_{\mathbf{R}_i, \mathbf{t}_i} \sum_{(\tilde{\mathbf{p}}_{x_j}, \tilde{\mathbf{q}}_{y_j}) \in \mathcal{C}} [\|\mathbf{R}_i \cdot \tilde{\mathbf{p}}_{x_j} + \mathbf{t}_i - \tilde{\mathbf{q}}_{y_j}\|_2^2 < \tau_a], \quad (3-13)$$

其中  $[\cdot]$  是 Iverson 括号。 $\tau_a$  是接受半径。然后, 我们通过求解3-12, 迭代地使用存活的内点匹配重新估计变换  $N_r$  次。如4.1所示, 我们的方法在 RANSAC 的配准精度上实现了可比较的结果, 但将计算时间减少了 100 倍以上。此外, 与深度稳健估计器<sup>[24-26]</sup>不同, 我们的方法是无参数的, 不需要网络训练。

### 3.2.7 损失函数

损失函数  $\mathcal{L} = \mathcal{L}_{oc} + \mathcal{L}_p$  由用于超点匹配的重叠感知圆形损失  $\mathcal{L}_{oc}$  和用于点匹配的点匹配损失  $\mathcal{L}_p$  组成。

**重叠感知圆形损失。**现有的方法<sup>[6,10]</sup>通常将超点匹配形式化为多标签分类问题，并采用带双重 softmax<sup>[10]</sup>或最优传输<sup>[6,31]</sup>的交叉熵损失。每个超点被分配（分类）给一个或多个其他超点，其中基于 patch 重叠计算的 ground truth，很可能一个 patch 会与多个 patch 重叠。通过分析交叉熵损失的梯度，我们发现在多标签分类中，具有高置信度得分的正类被正梯度抑制。这阻碍了模型从中提取可靠的超点对应关系。

为了解决这个问题，我们选择以度量学习的方式提取超点描述符。一个直接的解决方案是采用类似于<sup>[4-5]</sup>的圆形损失<sup>[33]</sup>。然而，圆形损失忽视了正样本之间的差异，并对它们进行了等权重的处理。因此，它在匹配重叠相对较低的 patch 时遇到困难。出于这个原因，我们设计了一个重叠感知的圆形损失，以使模型关注那些重叠较高的匹配。我们选择在  $Q$  中至少有一个正 patch 的  $P$  中的 patch，以形成一组锚定 patch， $A$ 。如果一对 patch 至少有 10% 的重叠，那么它们是正的，如果它们没有重叠，那么它们是负的。所有其他对都被忽略。对于每个锚定 patch  $G_i^P \in A$ ，我们将其在  $Q$  中的正 patch 集合表示为  $\varepsilon_p^i$ ，其负 patch 集合表示为  $\varepsilon_n^i$ 。然后在  $P$  上定义重叠感知的圆形损失为：

$$\mathcal{L}_{oc}^P = \frac{1}{|A|} \sum_{G_i^P \in A} \log \left[ 1 + \sum_{G_j^Q \in \varepsilon_p^i} e^{\lambda_i^j \beta_p^{i,j} (d_i^j - \Delta_p)} \cdot \sum_{G_k^Q \in \varepsilon_n^i} e^{\beta_n^{i,k} (\Delta_n - d_i^k)} \right], \quad (3-14)$$

其中  $d_i^j = \|\hat{\mathbf{h}}_i^P - \hat{\mathbf{h}}_j^Q\|_2$  是特征空间中的距离， $\lambda_i^j = (o_i^j)^{\frac{1}{2}}$ ， $o_i^j$  代表  $G_i^P$  和  $G_j^Q$  之间的重叠比率。每个样本的正权重和负权重分别为  $\beta_p^{i,j} = \gamma(d_i^j - \Delta_p)$  和  $\beta_n^{i,k} = \gamma(\Delta_n - d_i^k)$  计算。边缘超参数设置为  $\Delta_p = 0.1$  和  $\Delta_n = 1.4$ 。重叠感知的圆形损失根据重叠比率重新对  $\varepsilon_p^i$  上的损失值进行权重计算，以便给予重叠更高的 patch 对更多的的重要性。同样的，对于  $Q$  上的损失  $\mathcal{L}_{oc}^Q$  也是如此。总的损失是  $\mathcal{L}_{oc} = (\mathcal{L}_{oc}^P + \mathcal{L}_{oc}^Q)/2$ 。

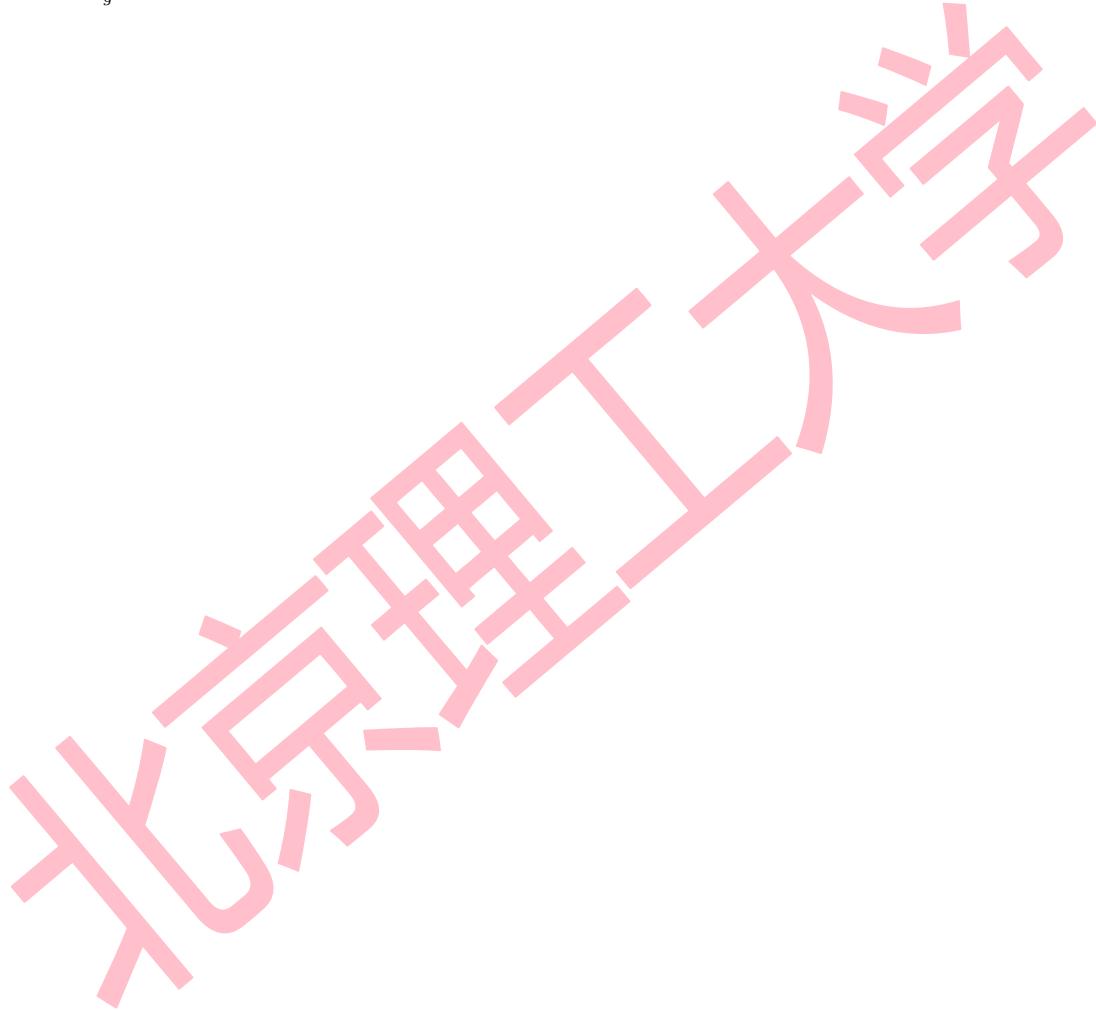
**点匹配损失。**地面真实点对应关系相对稀疏，因为它们只在下采样的点云中可用。我们简单地在每个超点对应关系的分配矩阵  $\bar{\mathbf{Z}}_i$  上使用负对数似然损失<sup>[31]</sup>。在训练过程中，我们随机采样  $N_g$  个地面真实的超点对应关系  $\{\hat{\mathcal{C}}_i^*\}$ ，而不是使用预测的。对于每个  $\hat{\mathcal{C}}_i^*$ ，我们提取一组地面真实的点对应关系  $\mathcal{M}_i$ ，匹配半径为  $\tau$ 。两个 patch

中未匹配的点集合分别记为  $\mathcal{I}_i$  和  $\mathcal{J}_i$ 。计算  $\hat{\mathcal{C}}_i^*$  的个体点匹配损失为：

$$\mathcal{L}_{p,i} = - \sum_{(x,y) \in \mathcal{M}_i} \log \bar{z}_{x,y}^i - \sum_{x \in \mathcal{I}_i} \log \bar{z}_{x,m_i+1}^i - \sum_{y \in \mathcal{J}_i} \log \bar{z}_{n_i+1,y}^i, \quad (3-15)$$

通过对所有采样的超点匹配进行个体损失的平均，计算最终的损失：

$$\mathcal{L}_p = \frac{1}{N_g} \sum_{i=1}^{N_g} \mathcal{L}_{p,i}$$



## 第 4 章 实验分析

### 4.1 室内基准测试：3DMatch 和 3DLoMatch

#### 4.1.1 数据集

3DMatch<sup>[34]</sup>包含 62 个场景，其中 46 个用于训练，8 个用于验证，8 个用于测试。我们使用了由<sup>[5]</sup>预处理的训练数据，并在 3DMatch 和 3DLoMatch<sup>[5]</sup>协议上进行评估。3DMatch 中的点云对有超过 30% 的重叠，而 3DLoMatch 中的点云对重叠度低，范围在 10% 至 30% 之间。

#### 4.1.2 指标

根据<sup>[4-5]</sup>，我们使用三个指标评估性能：(1) 内点比例 (Inlier Ratio, IR)，在地面真实变换下，残差低于特定阈值（例如，0.1m）的假设对应关系的比例；(2) 特征匹配召回率 (Feature Matching Recall, FMR)，内点比例高于特定阈值（例如，5%）的点云对的比例；(3) 配准召回率 (Registration Recall, RR)，变换误差小于特定阈值（例如，RMSE < 0.2m）的点云对的比例。

#### 4.1.3 对应性结果

首先，我们将我们的方法与近期的最先进技术：PerfectMatch<sup>[2]</sup>，FCGF<sup>[3]</sup>，D3Feat<sup>[4]</sup>，SpinNet<sup>[7]</sup>，Predator<sup>[5]</sup>，YOHO<sup>[16]</sup>和 CoFiNet<sup>[6]</sup>进行比较，比较结果见4-1(顶部和中部)。按照<sup>[4-5]</sup>的方式，我们报告了不同数量对应性的结果。对于特征匹配召回率，我们的方法在 3DLoMatch 上至少提高了 5 百分点 (pp)，显示了其在低重叠情况下的有效性。对于内点比例，改进更为显著。它在 3DMatch 和 3DLoMatch 上均超过了基准线，分别提高了 7~33 pp 和 17~31 pp。对应性越少，提升越大。这表明我们的方法提取的对应性更为可靠。

#### 4.1.4 配准结果

为了评估配准性能，我们首先比较了通过 RANSAC 获得的配准召回率，比较结果见4-1(底部)。按照<sup>[4-5]</sup>的方式，我们运行了 50K 次 RANSAC 迭代来估计转换。GeoTransformer 在 3DMatch 和 3DLoMatch 上都获得了新的最先进的结果。它在 3DMatch 上超过了之前的最好结果 1.2 pp，在 3DLoMatch 上超过了 7.5 pp，显示了

# Samples	3DMatch					3DLoMatch				
	5000	2500	1000	500	250	5000	2500	1000	500	250
<i>Feature Matching Recall (%) ↑</i>										
PerfectMatch <sup>[2]</sup>	95.0	94.3	92.9	90.1	82.9	63.6	61.7	53.6	45.2	34.2
FCGF <sup>[3]</sup>	97.4	97.3	97.0	96.7	96.6	76.6	75.4	74.2	71.7	67.3
D3Feat <sup>[4]</sup>	95.6	95.4	94.5	94.1	93.1	67.3	66.7	67.0	66.7	66.5
SpinNet <sup>[7]</sup>	97.6	97.2	96.8	95.5	94.3	75.3	74.9	72.5	70.0	63.6
Predator <sup>[5]</sup>	96.6	96.6	96.5	96.3	96.5	78.6	77.4	76.3	75.7	75.3
YOHO <sup>[16]</sup>	<b>98.2</b>	97.6	97.5	97.7	96.0	79.4	78.1	76.3	73.8	69.1
CoFiNet <sup>[6]</sup>	<u>98.1</u>	<b>98.3</b>	<b>98.1</b>	<b>98.2</b>	<b>98.3</b>	<u>83.1</u>	<u>83.5</u>	<u>83.3</u>	<u>83.1</u>	<u>82.6</u>
GeoTransformer ( <i>ours</i> )	97.9	<u>97.9</u>	<u>97.9</u>	<u>97.9</u>	<u>97.6</u>	<b>88.3</b>	<b>88.6</b>	<b>88.8</b>	<b>88.6</b>	<b>88.3</b>
<i>Inlier Ratio (%) ↑</i>										
PerfectMatch <sup>[2]</sup>	36.0	32.5	26.4	21.5	16.4	11.4	10.1	8.0	6.4	4.8
FCGF <sup>[3]</sup>	56.8	54.1	48.7	42.5	34.1	21.4	20.0	17.2	14.8	11.6
D3Feat <sup>[4]</sup>	39.0	38.8	40.4	41.5	41.8	13.2	13.1	14.0	14.6	15.0
SpinNet <sup>[7]</sup>	47.5	44.7	39.4	33.9	27.6	20.5	19.0	16.3	13.8	11.1
Predator <sup>[5]</sup>	58.0	58.4	<u>57.1</u>	<u>54.1</u>	49.3	<u>26.7</u>	<u>28.1</u>	<u>28.3</u>	<u>27.5</u>	25.8
YOHO <sup>[16]</sup>	64.4	60.7	55.7	46.4	41.2	25.9	23.3	22.6	18.2	15.0
CoFiNet <sup>[6]</sup>	49.8	51.2	51.9	52.2	<u>52.2</u>	24.4	25.9	26.7	26.8	<u>26.9</u>
GeoTransformer ( <i>ours</i> )	<b>71.9</b>	<b>75.2</b>	<b>76.0</b>	<b>82.2</b>	<b>85.1</b>	<u>43.5</u>	<u>45.3</u>	<u>46.2</u>	<u>52.9</u>	<u>57.7</u>
<i>Registration Recall (%) ↑</i>										
PerfectMatch <sup>[2]</sup>	78.4	76.2	71.4	67.6	50.8	33.0	29.0	23.3	17.0	11.0
FCGF <sup>[3]</sup>	85.1	84.7	83.3	81.6	71.4	40.1	41.7	38.2	35.4	26.8
D3Feat <sup>[4]</sup>	81.6	84.5	83.4	82.4	77.9	37.2	42.7	46.9	43.8	39.1
SpinNet <sup>[7]</sup>	88.6	86.6	85.5	83.5	70.2	59.8	54.9	48.3	39.8	26.8
Predator <sup>[5]</sup>	89.0	89.9	<u>90.6</u>	88.5	86.6	59.8	61.2	62.4	60.8	58.1
YOHO <sup>[16]</sup>	90.8	<u>90.3</u>	89.1	<u>88.6</u>	84.5	65.2	65.5	63.2	56.5	48.0
CoFiNet <sup>[6]</sup>	89.3	88.9	88.4	87.4	<u>87.0</u>	<u>67.5</u>	<u>66.2</u>	<u>64.2</u>	<u>63.1</u>	<u>61.0</u>
GeoTransformer ( <i>ours</i> )	<b>92.0</b>	<b>91.8</b>	<b>91.8</b>	<b>91.4</b>	<b>91.2</b>	<b>75.0</b>	<b>74.8</b>	<b>74.2</b>	<b>74.1</b>	<b>73.5</b>

表 4-1 在 3DMatch 和 3DLoMatch 上的结果。

其在高重叠和低重叠场景下的效能。更重要的是，我们的方法在不同数量的样本下都很稳定，因此它不需要像以前的方法那样<sup>[3,6-7,16]</sup>采样大量的对应性来提高性能。

我们接着在 4-2 中比较了不使用 RANSAC 的配准结果。我们开始时通过对应性上的加权 SVD 来求解对齐变换。基线要么无法达到合理的结果，要么性能下降严重。相比之下，GeoTransformer（使用加权 SVD）在 3DMatch 上达到了 86.5% 的配准召回率，在 3DLoMatch 上达到了 59.9%，接近 Predator 使用 RANSAC 的结果。在没有 RANSAC 的异常值过滤下，高内点比率对于成功配准是必需的。然而，高内点比率并不一定会导致高配准召回率，因为对应性可能会聚集在一起，如<sup>[5]</sup>所述。尽管如此，我们的方法通过提取可靠且分布良好的超点对应性，在没有 RANSAC 的情况下仍然表现良好。

Model	Estimator	#Samples	RR(%)		Time(s)		
			3DM	3DLM	Model	Pose	Total
FCGF <sup>[3]</sup>	RANSAC-50k	5000	85.1	40.1	0.052	3.326	3.378
D3Feat <sup>[4]</sup>	RANSAC-50k	5000	81.6	37.2	0.024	3.088	3.112
SpinNet <sup>[7]</sup>	RANSAC-50k	5000	88.6	59.8	60.248	0.388	60.636
Predator <sup>[5]</sup>	RANSAC-50k	5000	89.0	59.8	0.032	5.120	5.152
CoFiNet <sup>[6]</sup>	RANSAC-50k	5000	89.3	67.5	0.115	1.807	1.922
GeoTransformer ( <i>ours</i> )	RANSAC-50k	5000	<b>92.0</b>	<b>75.0</b>	0.075	1.558	1.633
FCGF <sup>[3]</sup>	weighted SVD	250	42.1	3.9	0.052	0.008	0.056
D3Feat <sup>[4]</sup>	weighted SVD	250	37.4	2.8	0.024	0.008	0.032
SpinNet <sup>[7]</sup>	weighted SVD	250	34.0	2.5	60.248	0.006	60.254
Predator <sup>[5]</sup>	weighted SVD	250	50.0	6.4	0.032	0.009	0.041
CoFiNet <sup>[6]</sup>	weighted SVD	250	64.6	21.6	0.115	0.003	0.118
GeoTransformer ( <i>ours</i> )	weighted SVD	250	<b>86.5</b>	<b>59.9</b>	0.075	0.003	0.078
CoFiNet <sup>[6]</sup>	LGR	all	87.6	64.8	0.115	0.028	0.143
GeoTransformer ( <i>ours</i> )	LGR	all	<b>91.5</b>	<b>74.0</b>	0.075	0.013	0.088

表 4-2 Registration results w/o RANSAC on 3DMatch (3DM) and 3DLoMatch (3DLM). The *model time* is the time for feature extraction, while the *pose time* is for transformation estimation.

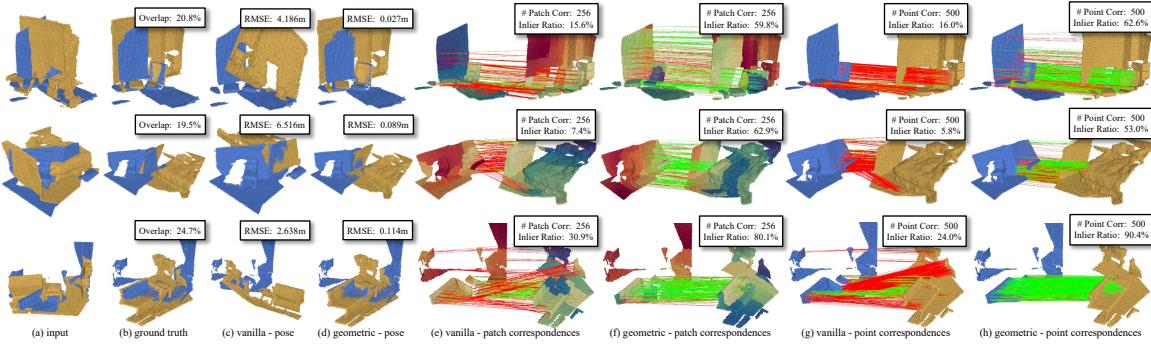


图 4-1 基于自注意力和几何自注意力的模型的配准结果。在列 (e) 和 (f) 中，我们使用 t-SNE 可视化了这些补丁的特征。在第一行，几何自注意力通过基于它们与更显著区域（例如椅子）之间的几何关系，帮助找到结构较弱的墙壁上的内点匹配。在接下来的行中，几何自注意力通过基于它们与床或沙发之间的几何关系，帮助拒绝了类似的平面或角落补丁之间的异常匹配。

当我们使用我们的局部到全局配准(LGR)来计算变换时，我们的方法将 3DMatch 的配准召回率提升到 91.5%，3DLoMatch 提升到 74.0%，大幅超过所有基于 RANSAC 的基线。这些结果也非常接近我们使用 RANSAC 的结果，但 LGR 在姿态时间上比 RANSAC 快了 100 倍以上。这些结果展示了我们的方法在准确性和速度上的优越性。

#### 4.1.5 消融实验

我们进行了广泛的消融实验，以更好地理解我们方法中的各个模块<sup>1</sup>。为了评估超点 (patch) 匹配，我们引入了另一个指标 *Patch Inlier Ratio* (PIR)，它表示具有实

<sup>1</sup>由于篇幅限制，我们在附录 D.3 中呈现了一些消融实验的结果。

Model	3DMatch				3DLoMatch			
	PIR	FMR	IR	RR	PIR	FMR	IR	RR
(a) graph neural network	73.3	97.9	56.5	89.5	39.4	84.9	29.2	69.8
(b) vanilla self-attention	79.6	97.9	60.1	89.0	45.2	85.6	32.6	68.4
(c) self-attention w/ ACE	83.2	<b>98.1</b>	68.5	89.3	48.2	84.3	38.9	69.3
(d) self-attention w/ RCE	80.0	97.9	66.1	88.5	46.1	84.6	37.9	68.7
(e) self-attention w/ PPF	83.5	97.5	68.5	88.6	49.8	83.8	39.9	69.5
(f) self-attention w/ RDE	84.9	98.0	69.1	90.7	50.6	85.8	40.3	72.1
(g) geometric self-attention	<b>86.1</b>	97.7	<b>70.3</b>	<b>91.5</b>	<b>54.9</b>	<b>88.1</b>	<b>43.3</b>	<b>74.0</b>

表 4-3 Ablation experiments of the geometric self-attention

Model	3DMatch				3DLoMatch			
	PIR	FMR	IR	RR	PIR	FMR	IR	RR
(a) cross-entropy loss	80.0	97.7	65.7	90.0	45.9	85.1	37.4	68.4
(b) weighted cross-entropy loss	83.2	<b>98.0</b>	67.4	90.0	49.0	86.2	38.6	70.7
(c) circle loss	85.1	97.8	69.5	90.4	51.5	86.1	41.3	71.5
(d) overlap-aware circle loss	<b>86.1</b>	97.7	<b>70.3</b>	<b>91.5</b>	<b>54.9</b>	<b>88.1</b>	<b>43.3</b>	<b>74.0</b>

表 4-4 Ablation experiments of the overlap-aware circle loss.

际重叠的 patch 匹配的比例。FMR 和 IR 是使用所有稠密点对应性进行报告的，使用 LGR 进行配准。

为了研究几何自注意力的有效性，我们在4-3中比较了七种点云内部特征学习的方法：(a) 图神经网络<sup>[5]</sup>，(b) 无位置嵌入的自注意力<sup>[6]</sup>，(c) 绝对坐标嵌入<sup>[31]</sup>，(d) 相对坐标嵌入<sup>[13]</sup>，(e) 点对特征嵌入<sup>[35-36]</sup>，(f) 对点对距离进行嵌入，(g) 几何结构嵌入。一般而言，注入几何信息可以提升性能。但是基于坐标的嵌入由于其对变换的不变性而受到一定限制。令人惊讶的是，图神经网络在 RR 上表现良好，这要归功于 kNN 图的变换不变性。然而，图神经网络的感受野受限，这对 IR 性能产生了不利影响。虽然 PPF 嵌入在理论上对于变换是不变的，但在实践中很难为超点估计准确的法线，这导致了较差的性能。

我们的方法在所有指标上都远远优于其他方法，尤其是在低重叠情况下，即使仅使用了点对距离嵌入，也表现出很强的鲁棒性。

下一步，我们对重叠感知的圆形损失进行了消融研究，如4-4所示。我们比较了四种用于监督超点匹配的损失函数：(a) 交叉熵损失<sup>[31]</sup>，(b) 加权交叉熵损失<sup>[6]</sup>，(c) 圆形损失<sup>[33]</sup>和(d) 重叠感知的圆形损失。对于前两种模型，我们使用最优传输层来计算匹配矩阵，就像<sup>[6]</sup>中那样。圆形损失要比两种交叉熵损失的变体效果更好，验证了以度量学习方式监督超点匹配的有效性。我们的重叠感知的圆形损失在所有指标上都远远优于普通的圆形损失。

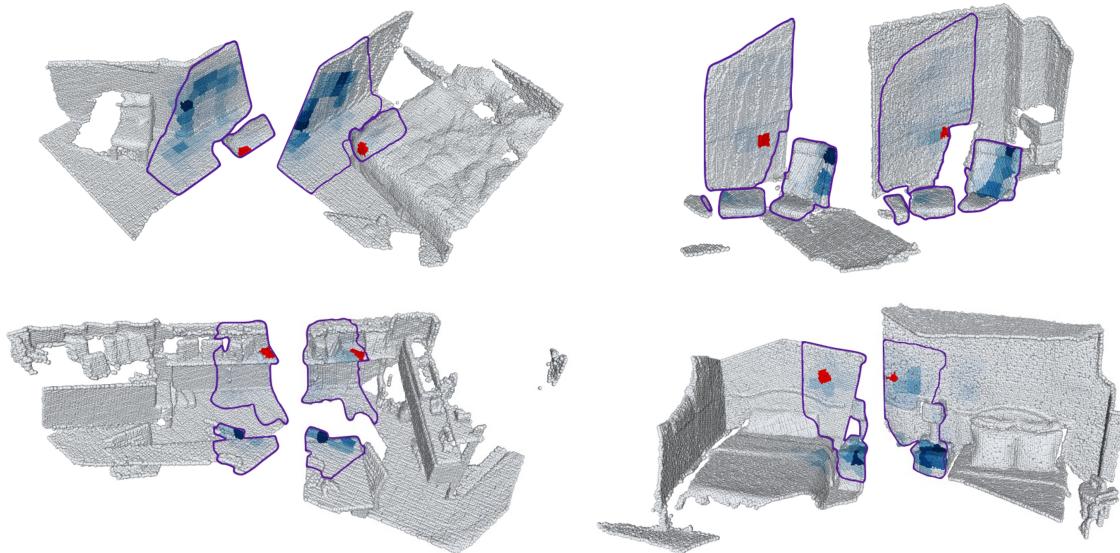


图 4-2 在四组点云对上可视化几何自注意力分数。重叠区域用紫色线条标示。锚点（对应的配对点）以红色高亮显示，而其他点的注意力分数以颜色编码（较深的颜色表示较高的分数）。请注意，即使在不相交的重叠区域，两个匹配锚点的注意力模式仍然是一致的。

#### 4.1.6 定性结果

图 4-1 提供了使用普通自注意力和我们的几何自注意力模型的配准结果图库。几何自注意力有助于从结构较弱的区域中根据它们与更显著区域的几何关系推断出补丁匹配 (1<sup>st</sup> 行)，并且能够拒绝在特征空间中相似但位置不同的异常匹配 (2<sup>nd</sup> 和 3<sup>rd</sup> 行)。

图 4-2 可视化了我们的几何自注意力学习到的注意力分数，显示了锚点补丁匹配之间的显著一致性。这表明我们的方法能够学习到跨点云的几何一致性，这对于准确的对应关系至关重要。

## 4.2 室外基准测试：KITTI odometry

### 4.2.1 数据集

KITTI odometry<sup>[37]</sup> 包含由 LiDAR 扫描的 11 个室外驾驶场景序列。我们遵循<sup>[3-5]</sup>，使用序列 0-5 进行训练，序列 6-7 进行验证，序列 8-10 进行测试。与<sup>[3-5]</sup>一样，我们使用 ICP 对地面真值位姿进行优化，并且仅评估距离至少为 10m 的点云配对。

### 4.2.2 评估指标

我们遵循<sup>[5]</sup>以三个指标评估我们的 GeoTransformer：(1) 相对旋转误差 (Relative Rotation Error, RRE)，估计旋转矩阵与地面真值旋转矩阵之间的测地线距离，(2) 相对

Model	RTE(cm)	RRE(°)	RR(%)
3DFeat-Net <sup>[38]</sup>	25.9	<b>0.25</b>	96.0
FCGF <sup>[3]</sup>	9.5	0.30	<b>96.6</b>
D3Feat <sup>[4]</sup>	<u>7.2</u>	0.30	<b>99.8</b>
SpinNet <sup>[7]</sup>	9.9	0.47	99.1
Predator <sup>[5]</sup>	<b>6.8</b>	<u>0.27</u>	<b>99.8</b>
CoFiNet <sup>[6]</sup>	8.2	0.41	<b>99.8</b>
GeoTransformer ( <i>ours</i> , RANSAC-50k)	7.4	<u>0.27</u>	<b>99.8</b>
FMR <sup>[22]</sup>	~66	1.49	90.6
DGR <sup>[25]</sup>	~32	0.37	<b>98.7</b>
HRegNet <sup>[39]</sup>	~12	<u>0.29</u>	<u>99.7</u>
GeoTransformer ( <i>ours</i> , LGR)	<b>6.8</b>	<b>0.24</b>	<b>99.8</b>

表 4-5 在 KITTI odometry 上的实验结果。

平移误差 (Relative Translation Error, RTE)，估计平移向量与地面真值平移向量之间的欧氏距离，以及 (3) 配准成功率 (Registration Recall, RR)，在 RRE<5° 和 RTE<2m 的阈值下，点云配对的比例。

#### 4.2.3 配准结果。

在4-5(顶部) 中，我们将我们的方法与基于 RANSAC 的最先进方法进行比较：3DFeat-Net<sup>[38]</sup>，FCGF<sup>[3]</sup>，D3Feat<sup>[4]</sup>，SpinNet<sup>[7]</sup>，Predator<sup>[5]</sup>和 CoFiNet<sup>[6]</sup>。我们的方法与这些方法相当，展现了在室外场景中的良好通用性。我们进一步与三种无需 RANSAC 的方法进行比较，如4-5(底部) 所示：FMR<sup>[22]</sup>，DGR<sup>[25]</sup>和 HRegNet<sup>[39]</sup>。我们的方法在所有基准方法上都取得了很大的优势。此外，我们的方法在使用 LGR 时超过了所有基于 RANSAC 的方法。

## 第 5 章 总结

我们提出了几何变换器（Geometric Transformer）来学习用于点云配准的鲁棒的粗到精细的对应关系。通过编码超点之间的成对距离和三元角度，我们的方法捕捉了点云之间的几何一致性，并具有转换不变性。凭借可靠的对应关系，我们的方法以一种无需 RANSAC 的方式实现了快速而准确的配准。未来，我们希望将我们的方法扩展到跨模态（例如，2D-3D）配准，并拥有更丰富的应用场景。



## 参考文献

- [1] Deng H, Birdal T, Ilic S. Ppfnet: Global context aware local features for robust 3d point matching[C]//CVPR. 2018: 195-205.
- [2] Gojcic Z, Zhou C, Wegner J D, et al. The perfect match: 3d point cloud matching with smoothed densities[C]//CVPR. 2019: 5545-5554.
- [3] Choy C, Park J, Koltun V. Fully convolutional geometric features[C]//CVPR. 2019: 8958-8966.
- [4] Bai X, Luo Z, Zhou L, et al. D3feat: Joint learning of dense detection and description of 3d local features[C]//CVPR. 2020: 6359-6367.
- [5] Huang S, Gojcic Z, Usyatsov M, et al. PREDATOR: Registration of 3D Point Clouds with Low Overlap[C]//CVPR. 2021: 4267-4276.
- [6] Yu H, Li F, Saleh M, et al. CoFiNet: Reliable Coarse-to-fine Correspondences for Robust Point Cloud Registration[J]. ArXiv preprint arXiv:2110.14076, 2021.
- [7] Ao S, Hu Q, Yang B, et al. SpinNet: Learning a General Surface Descriptor for 3D Point Cloud Registration[C]//CVPR. 2021: 11753-11762.
- [8] Rocco I, Cimpoi M, Arandjelović R, et al. Neighbourhood Consensus Networks[J]. NeurIPS, 2018, 31: 1651-1662.
- [9] Zhou Q, Sattler T, Leal-Taixe L. Patch2pix: Epipolar-guided pixel-level correspondences[C]//CVPR. 2021: 4669-4678.
- [10] Sun J, Shen Z, Wang Y, et al. LoFTR: Detector-free local feature matching with transformers[C]//CVPR. 2021: 8922-8931.
- [11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//NeurIPS. 2017: 5998-6008.
- [12] Wang Y, Solomon J M. Deep closest point: Learning representations for point cloud registration[C]//ICCV. 2019: 3523-3532.
- [13] Zhao H, Jiang L, Jia J, et al. Point transformer[C]//ICCV. 2021: 16259-16268.
- [14] Yang J, Zhang Q, Ni B, et al. Modeling point clouds with self-attention and gumbel subset sampling[C]//CVPR. 2019: 3323-3332.
- [15] Deng H, Birdal T, Ilic S. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors[C]//ECCV. 2018: 602-618.
- [16] Wang H, Liu Y, Dong Z, et al. You Only Hypothesize Once: Point Cloud Registration with Rotation-equivariant Descriptors[J]. ArXiv preprint arXiv:2109.00182, 2021.
- [17] Wang Y, Solomon J. PRNet: self-supervised learning for partial-to-partial registration[C]//NeurIPS. 2019: 8814-8826.
- [18] Yew Z J, Lee G H. Rpm-net: Robust point matching using learned features[C]//CVPR. 2020: 11824-11833.
- [19] Fu K, Liu S, Luo X, et al. Robust Point Cloud Registration Framework Based on Deep Graph Matching[C]//CVPR. 2021: 8893-8902.
- [20] Besl P J, McKay N D. Method for registration of 3-D shapes[C]//Sensor fusion IV: control paradigms and data structures: vol. 1611. 1992: 586-606.
- [21] Aoki Y, Goforth H, Srivatsan R A, et al. Pointnetlk: Robust & efficient point cloud registration using pointnet[C]//CVPR. 2019: 7163-7172.

- [22] Huang X, Mei G, Zhang J. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences[C]//CVPR. 2020: 11366-11374.
- [23] Xu H, Liu S, Wang G, et al. OMNet: Learning Overlapping Mask for Partial-to-Partial Point Cloud Registration[C]//ICCV. 2021: 3132-3141.
- [24] Pais G D, Ramalingam S, Govindu V M, et al. 3dregnet: A deep neural network for 3d point registration[C]//CVPR. 2020: 7193-7203.
- [25] Choy C, Dong W, Koltun V. Deep global registration[C]//CVPR. 2020: 2514-2523.
- [26] Bai X, Luo Z, Zhou L, et al. PointDSC: Robust Point Cloud Registration using Deep Spatial Consistency[C]//CVPR. 2021: 15859-15869.
- [27] Thomas H, Qi C R, Deschaud J E, et al. Kpconv: Flexible and deformable convolution for point clouds[C]//ICCV. 2019: 6411-6420.
- [28] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//CVPR. 2017: 2117-2125.
- [29] Li J, Chen B M, Lee G H. So-net: Self-organizing network for point cloud analysis[C]//CVPR. 2018: 9397-9406.
- [30] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//. 2021.
- [31] Sarlin P E, DeTone D, Malisiewicz T, et al. Superglue: Learning feature matching with graph neural networks[C]//CVPR. 2020: 4938-4947.
- [32] Sinkhorn R, Knopp P. Concerning nonnegative matrices and doubly stochastic matrices[J]. Pacific Journal of Mathematics, 1967, 21(2): 343-348.
- [33] Sun Y, Cheng C, Zhang Y, et al. Circle loss: A unified perspective of pair similarity optimization[C]//CVPR. 2020: 6398-6407.
- [34] Zeng A, Song S, Nießner M, et al. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions[C]//CVPR. 2017: 1802-1811.
- [35] Drost B, Ulrich M, Navab N, et al. Model globally, match locally: Efficient and robust 3D object recognition[C]//CVPR. 2010: 998-1005.
- [36] Raposo C, Barreto J P. Using 2 point+ normal sets for fast registration of point clouds with small overlap[C]//ICRA. 2017: 5652-5658.
- [37] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//CVPR. 2012: 3354-3361.
- [38] Yew Z J, Lee G H. 3dfeat-net: Weakly supervised local 3d features for point cloud registration[C]//ECCV. 2018: 607-623.
- [39] Lu F, Chen G, Liu Y, et al. HRegNet: A Hierarchical Network for Large-scale Outdoor LiDAR Point Cloud Registration[C]//ICCV. 2021: 16014-16023.