基于混合注意力机制自适应权重的点云配准算法研究

武　斌[1]，臧金卓[1]，赵　洁[1*]，吉　愈[2]

（1. 天津城建大学计算机信息与工程学院，天津 300380；2. 天津科技大学经济与管理学院，天津 300202）

摘要：三维点云配准作为一项关键技术，在计算机视觉、自动驾驶和机器人导航等领域有着广泛的应用场景。为了解决三维点云配准网络中存在的空间结构信息提取、融合不充分等问题，提出一种混合注意力自适应权重的点云配准算法。利用混合注意力机制进行特征提取，将关键点的局部细节与全局上下文信息相结合以强化特征的表征能力。采用自适应权重网络，根据数据内在特性学习并分配不同特征的重要性权重实现特征融合。在 ModelNet40 数据集和 KITTI odometry 数据集上分别进行实验，实验结果表明在 ModelNet40 数据集上配准准确率达到 96.7%。在 KITTI odometry 数据集上相对旋转误差和相对平移误差分别降低到 6.2cm 和 0.32°。本文算法使用的混合注意力机制自适应权重，有效提升点云配准精度。

关键词：点云配准；计算机视觉；注意力机制；自适应权重；特征提取；特征融合

中图分类号：TP301.6　　**文献标志码：**A

1　引言

点云配准是三维视觉中的一项基本工作，在三维建模、SLAM、图像处理和模式识别等多个领域发挥着重要作用。点云配准是将不同角度的点云通过旋转、平移等刚性变换，统一到同一参考坐标系中，形成完整点云的过程。传统的方法是根据几何特征点对的相似度来匹配关键点[1]。然而，这些方法往往依赖于良好的初始位置，噪声鲁棒性较差。

传统方法中经典的点云配准方法是迭代最近点算法[2]（Iterative Closest Point，ICP），通过迭代的方式不断寻找两片点云之间的最佳匹配点对，并基于这些点对计算出最优的旋转矩阵和平移向量使得经过变换后的点云能够尽可能地与参考点云重合，但其性能在很大程度上受到点云初始位置及外界噪声干扰的影响。为减少上述因素对配准精度的影响，ICP 的基础上，研究者们提出了多种改进算法，例如 Go-ICP、DICP 等。Go-ICP 算法，该方法不仅能够在不考虑待配准点云与目标点云初始位置的情

基金项目：天津市企业科技特派员项目（19JCTPJC47200）；天津市重点研发计划科技支撑重点项目（19YFZCGX00130）；天津市研究生科研创新项目（2021YJSS351）。

* 通信作者：赵洁

况下完成配准并且由于在分支定界方案中加入了局部 ICP 因此也具有很高的配准精度和速度[3]。为增加对噪声的鲁棒性，点特征直方图（Persistent Feature Histograms，PFH）[4] 和快速点特征直方图（Fast Persistent Feature Histograms，FPFH）[5] 利用手工制作的特征描述子建立对应关系，但描述子设计十分复杂。

相较于传统的点云配准算法，深度学习的点云配准算法在精确度、执行效率及鲁棒性上均展现出显著提升。例如，PointNetLK 将二维图像经典配准 LK 算法引入点云配准网络，提高了配准精度，但其迭代的计算方式效率不高；深度最近点（Deep Closest Point，DCP）通过 DGCNN 将点云嵌入到高维空间，然后利用 Transformer 网络合并点云的上下文信息与全局特征，利用 SVD 预测相对运动参数[6]。DCP 对噪声取得了较好的鲁棒性，但在低重叠的情况下表现不佳；PRNet 在 DCP 的基础上进行改进，提出使用关键点检测模块识别关键匹配点，在重叠点中选取关键点，消除了非重叠点的干扰[7]。FireNet 在网络中提出特征交互机制，实现了两个点云的点对信息交互和全局信息交互[8]。PCRNet 使用 PointNet 提取点云特征并利用完全连接层预测刚性变换，对噪声具有强鲁棒性，计算速度有所提高；但 PCRNet 只关注点云的全局特征，没有考虑点云的局部相关性，因此配准精度不足[9]。

为了弥补深度学习的点云配准方法在特征提取和融合方面的不足，从而提高配准的精度，本文提出混合注意力机制自适应权重的点云配准算法。本文的主要贡献如下。

1）提出自注意力机制与交叉注意力机制结合的混合注意力机制，更好捕获上下文信息，使点云的局部相关性和全局信息相结合，突出点云的几何特征与空间信息，提高模型的表达能力和泛化能力。

2）引入自适应权重网络，包括权重学习模块和融合模块。权重学习模块负责根据数据的特性和任务需求学习权重，而融合模块则利用学习到的权重对不同的特征或数据点进行融合，进一步完成点云配准。

2 相关工作

神经网络与深度学习在点云配准中广泛应用，通过训练网络模型来自动学习点云的特征表示和配准参数，提高配准的准确性和效率，在模型训练过程中，注意力机制、自适应权重等起到了关键作用，允许模型关注与配准任务相关的点云区域，有效地描述点云的局部和全局结构，使配准结果更加精准。

注意力机制能够让模型有选择地关注到输入数据中的重要信息，从而提高模型的表示能力和信息利用率。Choy[10] 等提出引入注意力机制来优化特征匹配过程，能够关注到点云中的重要区域或特征；但在处理具有噪声或缺失数据的点云时，该方法的鲁棒性可能受到一定影响。Yew[11] 等提出鲁棒性匹配网络（RPM-Net），能够处理具有噪声、外点等挑战的点云数据，但该方法主要关注点云之间的刚性变换估计，对于非刚性变换的处理比较有限。Zhao[12] 等提出 Point Transformer 进行点云处理，该方法通过注意力机制优化了点云特征之间的匹配，显著提高了配准的准确性。但由于 Transformer 架构中的自注意力机制需要计算所有点之间的交互，因此该方法的计算复杂度较高，可能不适合处理大规模的点云数据。施成浩[13] 等提出一种新颖的注意力机制 3D-RoFormer，能够以轻量、快速的方式编码相对位置信息，具备强大的上下文和几何特征描述能力，在解决点云配准和地点识别问题上的表现良好，并具有强大的泛化能力，但其在解决全局定位问题上还有欠缺。刘旭珩[14] 等提出的 AMGNet 模型，融入了 Transformer 架构以深化点云特征的上下文信息融合策略，增强了模型的性能表现，并且该模型具有优良的兼容性，能够适配并优化多种点云特征提取网络，但其抗干扰能力和在真实数据集上的表现仍需通过进一步的研究与优化。

自适应权重允许网络在融合过程中根据数据的特性和问题的复杂度，动态地调整不同网络或数据之间的权重。这种技术旨在提高融合结果的准确性和可靠性，以应对复杂多变的环境与数据集。Liu[15] 等提出自适应空间特征融合（ASFF）策略，ASFF 方法通过学习如何在不同层次上对特征进行空间过滤，从而解决了单镜头检测器中特征金字塔的不一致性问题，但在对特定数据集的依赖、计算资源需求及对复杂场景的适应性等方面存在一些限制。Yu Guo[16] 等提出一种鲁棒的点云配准方法，该方法可以同时学习具有自适应稀疏邻居的权向量并执行加权 ICP。自动学习权向量来局部抑制和全局消除噪声和异常值的影响。但权重调整策略的参数可能需要针对具体应用进行仔细调整。Mungoli[17] 等使用线性和非线性融合函数来组合自适应融合层中的特征，将其整合到现有

的深度学习架构中，实现自适应地在深度学习模型中融合特征。李杰[18]等引入自适应权重调节技术优化卡尔曼滤波算法，解决跨模态传感器低复杂度高效融合跟踪问题，最终实现多传感器融合的目标检测与跟踪。Zhang[19]等提出基于文本模态指导的层次自适应融合方法。该方法能够充分考虑模态之间的特征交互，同时平衡局部与全局的模态相关特征和独特特征，有效提高情感分类性能，但由于需要处理来自多个模态的数据，并进行复杂的融合操作，导致计算复杂度较高。

3 基于混合注意力机制自适应权重的点云配准网络

3.1 整体网络结构

本文构建的网络模型由特征提取块、特征融合块和姿态回归块组成。整体结构如图 1 所示。首先使用 $P_T=\{P_{T_j} \mid j=1, 2, 3, \cdots, n \subset R^3\}$ 表示输入目标点云，其中每个点 P_{T_j} 包含坐标（x, y, z）和其他相关信息，如强度和颜色等。然后初始化旋转矩阵 $\boldsymbol{R}_0 \in \boldsymbol{SO}$（3）和平移向量 $\boldsymbol{t}_0 \in R^3$，并将其应用于目标点云 P_T，生成源点云 $P_S=\{P_{S_i} \mid i=1, 2, \cdots, n \subset R^3\}$，这个过程可以用下面的公式表示

$$P_S = R_0 P_T + t_0 \tag{1}$$

这样就得到了经过预处理后待配准的点云数据。特征提取块由 5 个多层感知器（MLP）组成，类似于 PointNet 架构，大小分别为 64、64、64、128 和 1 024。通过共享权值的多层感知器与混合注意力机制输出高维特征并进行特征学习，输出融合了关键点局部相关性和全局信息的 1 024 维高维特征。最后，通过最大池化层，将大小为 $n \times 1\ 024$ 的特征整合为 $1 \times 1\ 024$。特征融合块由权重学习模块和融合模块两个核心组件组成，权重学习模块负责动态地调整两个输入点云特征的重要性，融合模块利用权重将两个点云的特征融合成 $1 \times 2\ 048$ 的特征。姿态回归块通过两个连续的全连接层对旋转矩阵 $\boldsymbol{R}$ 和平移向量 $\boldsymbol{t}$ 进行回归。每个全连接层含有五个隐藏层，分别为 1 024，1 024，512，512，256。输出层的大小分别为 4 和 3，表示四元数的四个分量与平移向量的三个分量，四元数通过数学转换得到旋转矩阵 $\boldsymbol{R}$。完成姿态回归后，将得到的旋转矩阵 $\boldsymbol{R}$ 和平移向量 $\boldsymbol{t}$ 应用于源点云，实现其在三维空间中的精确对齐，即完成点云配准。

3.2 混合注意力机制

本文提出将自注意力和交叉注意力结合在一起的混合注意力机制，以解决点云特征的提取问题。可以同时利用两种机制的优点，既考虑到点云内部的结构信息，又能很好的关注到点云之间的空间相关性，模型结构图如图 2 所示。

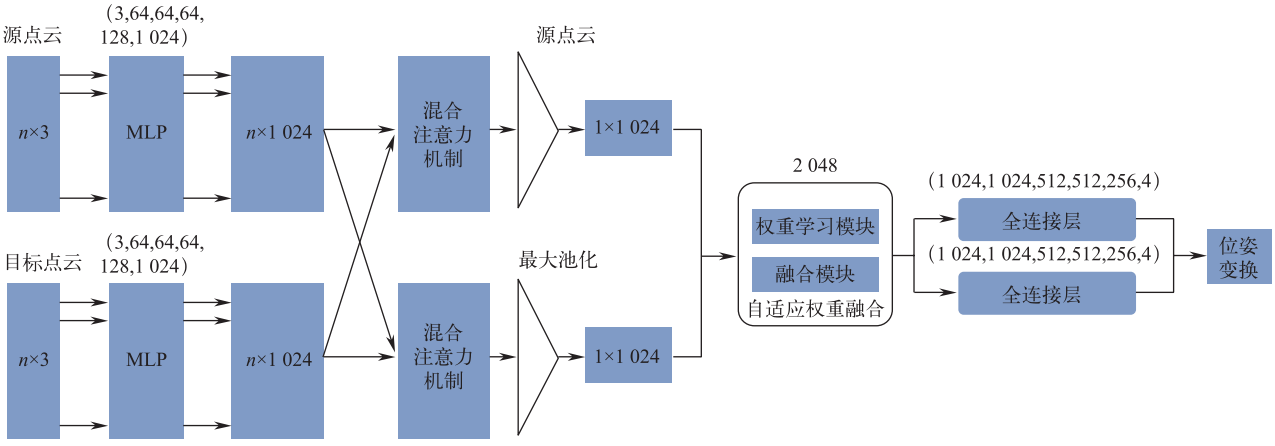


图 1 网络模型的整体结构

首先使用自注意力机制聚合源点云 P_S 上的所有特征，并在每层中逐步进行更新。设 $a_{\varepsilon \to i}$ 是所有关键点 $\{j:(i, j \in) \varepsilon\}$ 的聚合结果，$\varepsilon \in \{\varepsilon_{self}, \varepsilon_{cross}\}$，则元素 i 的消息传递更新为

$$^{(l+1)}x_i^{P_S} = {}^{(l)}x_i^{P_S} + MLP([{}^{(l)}x_i^{S} \| a_{\varepsilon \to i}]) \tag{2}$$

式中，$^{(l)}x_i^{S}$ 表示源点云 P_s 中第 l 层中节点 i 的中间表示，$[\cdot \| \cdot]$ 表示连接函数，目标点云 P_T 的更新步骤与其相同。

然后使用交叉注意力机制作用与两点云，实现点云之间的信息交互，$a_{\varepsilon \to i}$ 可表示为平均加权值，如下：

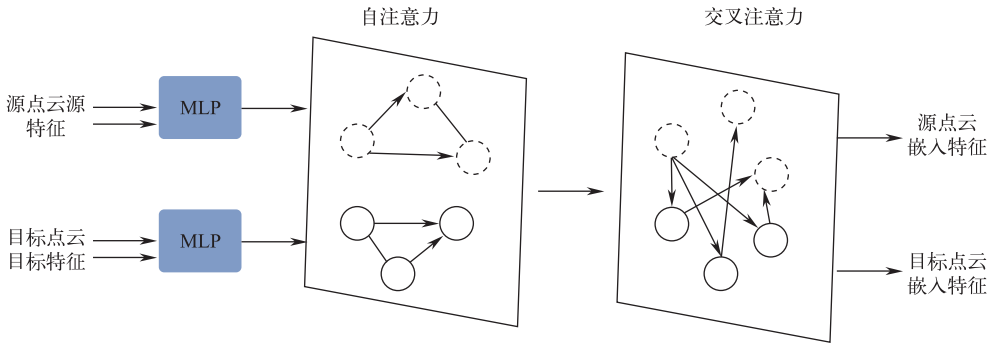


图 2　混合注意力模型结构图

$$a_{\varepsilon\to i}=\sum_{j:(i,j)\in\varepsilon}\alpha_{ij}\upsilon_j \tag{3}$$

注意力权重 $\alpha_{ij}=Soft\max_j\left(q_i^{\mathrm{T}}k_j\right)$ 表示查询的相似度。关键字 k_j、查询 q_i 和值 v_j 表示对于图神经网络深层特征的线性投影，$(Q,H)\in\{A,B\}^2$，故

$$q_i=W_1^{(l)}x_i^Q+b_1 \tag{4}$$

$$\begin{bmatrix}k_j\\v_j\end{bmatrix}=\begin{bmatrix}W_2\\W_3\end{bmatrix}^{(l)}x_i^H+\begin{bmatrix}b_2\\b_3\end{bmatrix} \tag{5}$$

最后使用多头注意力进行特征提取，全面的捕获点云的不同方面和细节，输出源点云 P_s 的交互特征：

$$head_i=soft\max\left(\frac{q_i\cdot k_i^{\mathrm{T}}}{\sqrt{D/4}}\right) \tag{6}$$

$$F_S^{\text{inter}}=L_i^{\text{NET}}+MLP(cat(head_1,head_2,head_3,head_4)) \tag{7}$$

式中，L_i^{NET} 表示特征提取后的特征，$cat\ (head_1,\ head_2,\ head_3,\ head_4)$ 表示一个全连接网络，目标点云 P_{T} 的计算与其同理。

3.3　自适应权重网络

在点云配准和融合的过程中，特征向量的提取和自适应权重的应用是至关重要的步骤。通过提取点云的特征并构建特征向量，我们能够捕获点云的局部和全局信息。这些特征向量随后被用作自适应权重网络的输入，该网络可以根据特征向量的相似度、重要性或质量等因素，动态地调整每个点云的权重，如图 3 所示。

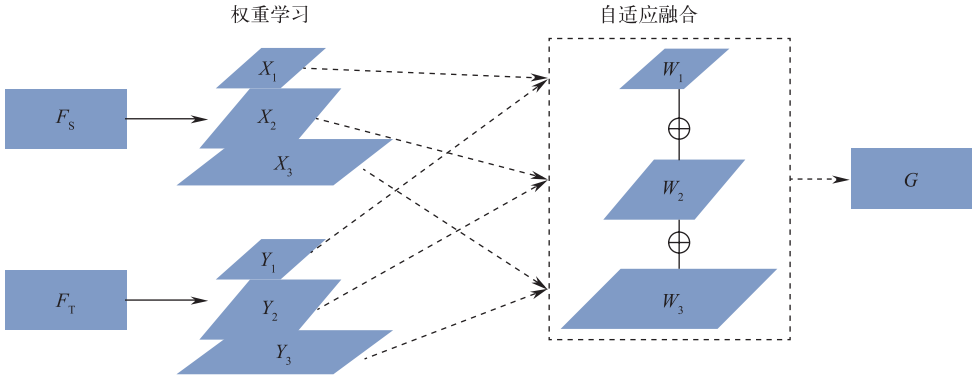


图 3　自适应权重模型结构图

在自适应权重网络中，欧式距离是一个常用的度量，用于计算特征向量之间的相似度。对于提取的点云 F_{S} 和 F_{T}，它们的特征向量分别为 X_i 和 Y_j，其中 i 和 j 分别是 P_{S} 和 P_{T} 中点的索引，它们之间的欧式距离可以通过计算每个维度上对应元素的差的平方和的平方根来得到。它们之间的距离值越小，表示两个特征向量越相似。

$$d(X_i,Y_j)=\sqrt{\sum_{k=1}^{n}(x_{ik}-y_{jk})^2} \tag{8}$$

式中，x_{ik} 是 X_i 中的第 k 个特征值，y_{jk} 是 Y_j 中的第 k 个特征值，n 是特征向量的维度。

一旦权重学习完成，就可以利用这些权重对不同的点云数据进行融合。融合过程通常采用加权平均融合策略，即根据每个点云的权重对相应的点进行加权，然后求和得到融合后的点云数据。自适应权重网络为每个点云数据集中的点分配了权重 ω_i 和 ω_j。则融合后的点云数据 G 可以通过以下表达式计算：

$$G=\frac{\sum_i \omega_i \cdot X_i+\sum_j \omega_j \cdot Y_j}{\sum_i \omega_i+\sum_j \omega_j} \tag{9}$$

式中，$\sum_i \omega_i \cdot X_i$和$\sum_j \omega_j \cdot Y_j$分别表示点云$P_S$和$P_T$中所有点根据其权重进行加权后的和，$\sum_i \omega_i+\sum_j \omega_j$是权重的总和，用于归一化融合后的点云数据。自适应地调整每个点的权重，确保了信息融合时能够准确捕捉关键特征，从而得到更加准确可靠的融合结果。

3.4 配准参数计算

本文采用EMD（Earth Mover's Distance，推土机距离）函数进行计算，在点云配准中，其为最小化模板点云和网络预测后经过位姿变换的源点云P_S^{est}之间的距离，EMD损失函数表达式如下：

$$Loss_{\mathrm{EMD}}\left(P_S^{\mathrm{est}},P_T\right)=\min_{\psi:P_S^{\mathrm{est}}\to P_T}\frac{1}{\left|P_S^{\mathrm{est}}\right|}\sum_{x\in P_S^{\mathrm{est}}}\left\|x-\psi(x)\right\|_2 \tag{10}$$

式中，P_T为目标点云，P_S^{est}为源点云P_S每次更新变换后的点云，该公式求一个双射函数ψ，使得基于该函数的点对距离和最小。EMD的定义是明确的，它提供了一种衡量两个点云集合之间差异的新视角。相比于传统的最近点距离（Nearest Point Distance，NPD）和豪斯多夫距离（Hausdorff Distance，HD），EMD在衡量点云间对应关系时展现出更高的准确性和全面性。

4 实验与分析

本文采用ModelNet40[20]标准数据集和室外数据集KITTI odometry[21]来训练网络模型，此外，还进行对比实验、消融实验验证网络有效性。

4.1 实验环境与参数设置

本文实验在Ubuntu 16.04操作系统中进行，在Python3.8、CUDA11.1环境下对选定数据集进行训练与测试。实验硬件平台搭载了高性能的NVIDIA RTX 3080 GPU，其24GB的显存为大规模数据处理提供了坚实的支撑。实验程序在PyTorch 1.10框架中编写实现，为了保证实验的可重复性和科学性，设置了一个固定的随机种子，确保每次实验生成的变换矩阵完全相同。采用随机梯度下降（Stochastic Gradient Descent，SGD）优化器对网络进行训练，通过精细调整其参数——初始学习率设为0.001，权重衰减设为0.000 001，以及动量参数设为0.98，实现对模型参数的快速且稳定的优化。在ModelNet40数据集上，设置200次迭代作为训练周期，训练集和测试集的批次大小均设置为20。在KITTI odometry数据集上使用0.6m和5°作为阈值。硬阈值置信度得分的阈值设置为0.6。

4.2 ModelNet40数据集

ModelNet40数据集包含9 843个训练示例和2 468个测试示例。每个点云数据由大量的三维点坐标构成，这些点捕捉了对象的形状和结构信息。通过对这些点云数据进行处理和分析，可以提取出对象的特征，进而实现对象的分类和识别。ModelNet40分为40类，包括飞机、床、椅子等。对于给定的模型，以最远点采样方法得到1 024个点形成原始点云，随机生成[0°，45°]的旋转角度和[−0.5，0.5]的平移量作用于点云，形成目标点云。

（1）网络性能的评价指标　MSE、RMSE和MAE是点云配准任务中常用的评价指标。其中，$MSE(R)$、$RMSE(R)$、$MAE(R)$分别表示旋转角度真实值与预测值之差的平方、平方根和绝对值。$MSE(t)$、$RMSE(t)$、$MAE(t)$分别表示平移距离真实值与预测值之差的平方、平方根和绝对值。这些误差越小，表示点云配准精度越高。

（2）整体结果　在本文中，所构建的模型在各项评估指标下的训练与测试性能曲线如图4所示。这些曲线呈现出高度的相似性，体现了模型在训练与验证阶段的一致性表现。具体而言，误差指标在最初的75个迭代周期内经历了显著的下降过程，标志着模型快速学习并适应数据分布。随着训练步数进一步增加至大约100步时，误差曲线趋于平稳，表明模型已经收敛，进入全局最优状态。

本文所构建的模型通过精确预测旋转矩阵与平移向量，对源点云实施了高效且准确的变换操作，成功实现了点云的精细配准。为了直观展现配准过程的成效，我们进一步将模板点云、原始源点云及配准后的点云进行了可视化呈现，如图5所示。在图中，采用鲜明的颜色进行区分，蓝色代表源点云，绿色代表模板点云，而红色则代表经过配准后的点云。通过观察可视化结果，可以看到，无论是具有规则对称结构的点云，还是结构复杂的点云，本文方法均展现出了卓越的配准效果。

（3）对比实验　1）精度测试。评估本文方法与同类方法的性能，结果见表1。本文方法具有最低的$MSE(R)$，$RMSE(R)$，$MAE(R)$，$MSE(t)$，$RMSE(t)$，MAE（t）；通过与迭代最近点ICP算法、Go-ICP算法、PointNetLK算法、PCRNet算法和AMGNet算法做对比实验，分析点云配准算法的配准效能。对比PCRNet方法，

在 6 个评价指标上分别带来了 30.3%、16.5%、24.3%、56.0%、33.6% 和 25.3% 的提升。这一显著成果不仅验证了本文提出的配准方法在精准度层面的卓越优势，更将其精准性能推向了新的高度。不仅如此，通过本文方法预测的旋转矩阵与平移向量，其与实际地面真相之间的误差指标达到了最低阈值，进一步突出了其在点云配准任务中的高精度表现。总体来看，本文的创新方法在效能上显著超越当前主流技术，实现了配准精度的提升。

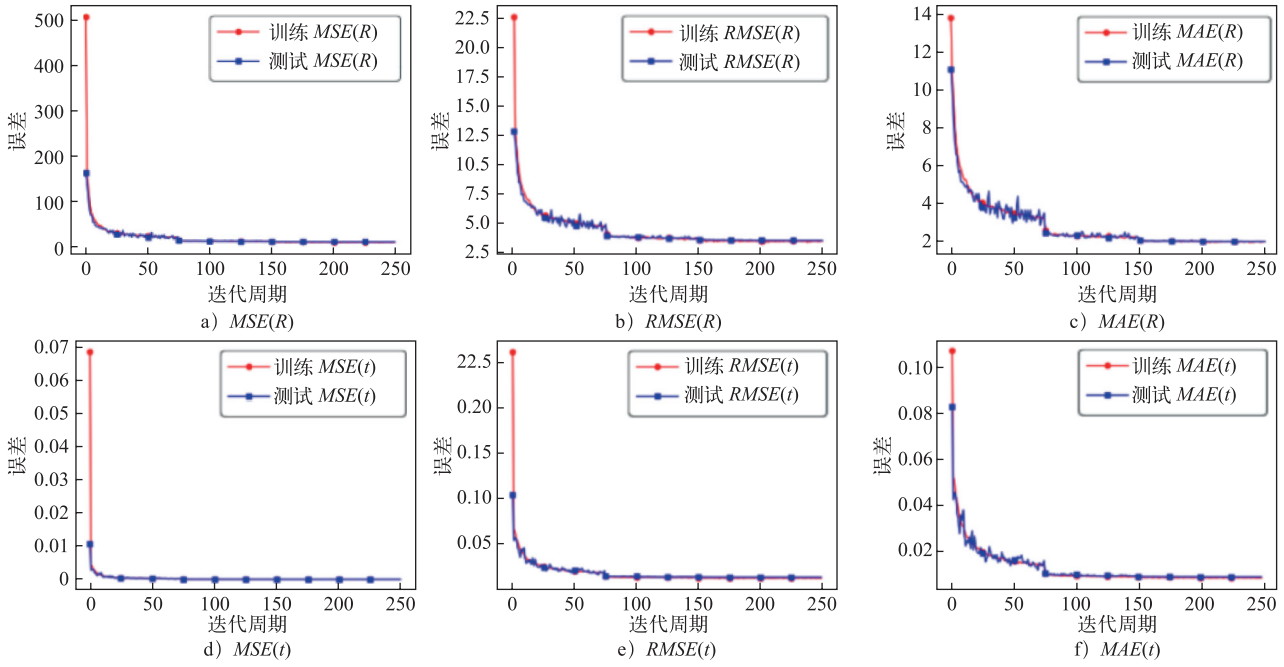
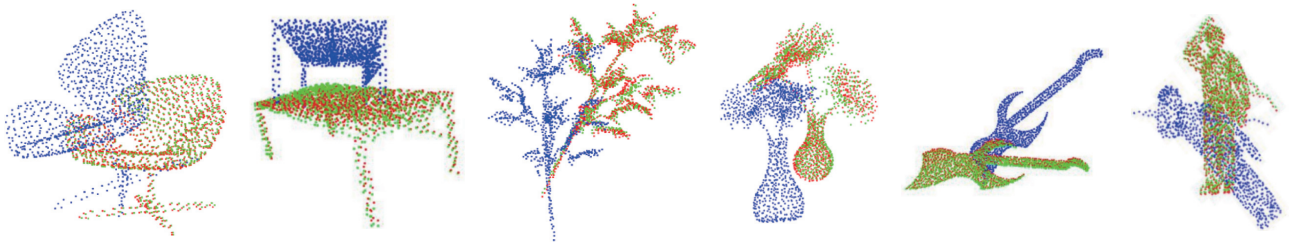


图 4　不同评价指标下的训练曲线和测试曲线对比图



a）转椅　b）桌子　c）树枝　d）花瓶　e）吉他　f）人

图 5　在 ModelNet40 数据集上配准效果

表 1　在 ModelNet40 原始数据集上进行测试

模型	*MSE*（*R*）	*RMSE*（*R*）	*MAE*（*R*）	*MSE*（*t*）	*RMSE*（*t*）	*MAE*（*t*）
ICP[2]	894.897 339	29.914 835	23.544 817	0.084 643	0.290 935	0.248 755
Go-ICP[3]	140.477 25	11.852 313	2.588 463	0.000 659	0.025 665	0.007 092
PointNetLK[6]	227.870 331	15.095 374	4.225 304	0.000 487	0.022 065	0.005 404
PCRNet[9]	17.447 983	4.177 078	2.628 345	0.000 241	0.015 526	0.010 579
AMGNet[13]	15.014 486	3.874 853	2.365 829	0.000 192	0.013 869	0.009 693
Ours	12.160 731	3.487 224	1.988 427	0.000 106	0.010 312	0.007 902

2）泛化性比较。为了全方位评估各模型的泛化能力，选用未曾参与模型训练的未验证类别数据集作为测试基准。见表 2，本文提出的配准方法不仅在 *MSE*（*R*）、*RMSE*（*R*）、*MSE*（*t*）、*RMSE*（*t*）、*MAE*（*R*）和 *MAE*（*t*）

这六大关键性能指标上均展现出最低值，而且其预测的旋转矩阵与平移向量与真实地面值的偏差在所有这些指标中均达到了最小值。充分证明了在面对全新、未经训练的类别数据时，本文方法依然能够维持极高的配准精度，彰显了其优越的泛化能力与创新优势。

表 2　对未验证类别数据集进行测试

模型	*MSE*（*R*）	*RMSE*（*R*）	*MAE*（*R*）	*MSE*（*t*）	*RMSE*（*t*）	*MAE*（*t*）
ICP	890.283 866	29.837 625	23.643 827	0.088 499	0.297 488	0.251 147
Go-ICP	163.144 190	12.772 791	2.915 846	0.000 538	0.023 214	0.006 472
PointNetLK	267.883 894	16.367 159	5.275 341	0.000 791	0.028 127	0.005 404
PCRNet	28.564 995	5.344 623	2.694 435	0.000 196	0.013 997	0.010 124
AMGNet	15.229 873	3.902 547	2.385 265	0.000 193	0.013 885	0.009 726
Ours	12.652 519	3.557 038	1.995 311	0.000 155	0.012 458	0.008 592

3）噪声鲁棒性比较。为了测试噪声对各模型性能的影响，在输入点云数据中引入了期望值为 0、方差设定为 0.01 的高斯噪声进行测试，其结果见表 3。本文所提出的配准方法在 *MSE*（*R*）、*RMSE*（*R*）、*MAE*（*R*）、*MSE*（*t*）和 *RMSE*（*t*）这五项核心指标上均显著优于其他对比方法，彰显了其强大的数据处理能力和抗噪性能。唯一不足的 *MAE*（*t*）指标，也仍优于传统的 ICP 方法和先进的 PCRNet 方法。从整体表现来看，本文方法在面对高斯噪声干扰时，依然能够保持较高的配准精度和稳定性，充分证明了其对于噪声环境的强鲁棒性。

表 3　对高斯噪声数据集的检验

模型	*MSE*（*R*）	*RMSE*（*R*）	*MAE*（*R*）	*MSE*（*t*）	*RMSE*（*t*）	*MAE*（*t*）
ICP	882.889 646	29.713 459	23.556 923	0.084 546	0.290 767	0.249 113
Go-ICP	131.498 121	11.467 263	2.535 739	0.000 483	0.023 154	0.004 279
PointNetLK	256.490 603	16.015 324	4.585 624	0.000 487	0.021 985	0.005 663
PCRNet	17.735 620	4.211 368	2.675 309	0.000 236	0.015 362	0.010 613
AMGNet	15.958 123	3.994 762	2.386 679	0.000 192	0.013 852	0.009 744
Ours	12.811 403	3.579 302	2.012 458	0.000 178	0.013 352	0.008 815

（4）消融实验　为了验证混合注意力机制和自适应权重网络在增强基础网络 PCRNet 性能方面的有效性，设计了消融实验。具体而言，我们分别评估了仅采用自适应权重网络用于特征融合的网络模型，仅采用混合注意力机制进行特征提取的网络模型，以及融合了混合注意力机制与自适应权重网络的综合模型进行了全面对比。在 ModelNet40 数据集上结果见表 4。

实验数据表明，当在 PCRNet 中单独使用混合注意力机制提取特征，六项关键评估指标均实现了显著提升，增幅分别高达 26.1%、14.1%、21.7%、46.9%、29.2% 及 24.0%，显现了该机制在捕捉数据深层次特性上的非凡能力。在 PCRNet 中单独使用自适应权重网络融合特征，也带来了可观的性能跃升，各项指标的增长率分别达到了 24.0 %、12.8%、20.2%、40.7%、22.9% 及 19.5%，验证了其依据数据特性智能调节融合权重的有效性与必要性。在 PCRNet 中同时使用混合注意力机制提取特征和自适应权重网络融合特征，其效果实现了 1+1>2 的飞跃。六项关键评估指标不仅均超过了单独使用任一技术时的提升幅度，更以 30.3%、16.5%、24.3%、56.0%、33.6% 及 25.3% 的卓越表现，显示了这两大创新技术在提升点云配

准精度上的协同增效作用。综上所示，混合注意力机制和自适应权重网络各自都能对网络起到实质的作用，从而提高点云配准的精度。

表 4　不同模块改进 PCRNet 模型的比较

模型	*MSE*（*R*）	*RMSE*（*R*）	*MAE*（*R*）	*MSE*（*t*）	*RMSE*（*t*）	*MAE*（*t*）
PCRNet	17.447 983	4.177 078	2.628 345	0.000 241	0.015 526	0.010 579
加自适应权重	13.254 151	3.640 625	2.098 704	0.000 143	0.011 975	0.008 520
加混合注意力	12.886 068	3.589 717	2.057 626	0.000 128	0.010 996	0.008 041
Ours	12.160 731	3.487 224	1.988 427	0.000 106	0.010 312	0.007 902

4.3　室外数据集 KITTI odometry

KITTI odometry 数据集作为室外评估数据集，该数据集包含 21 个户外场景，其中只有前 11 个场景记录了 ground-truth 刚性变换。使用场景 0 ～ 5 进行训练，场景 6 ～ 7 进行验证，场景 8 ～ 10 进行测试。

（1）网络性能的评价指标　相对平移误差（Relative Translation Error，RTE）是实际平移量与估计平移量之间的差异。它通常用平移向量的欧氏距离来表示，即估计平移向量 $\boldsymbol{t}$ 和真实平移向量 $\boldsymbol{t}^*$ 的差的模：

$$RTE = \left\| \boldsymbol{t} - \boldsymbol{t}^* \right\|_2 \tag{11}$$

相对旋转误差（Relative Rotation Error，RRE）衡量的是实际旋转角度与估计旋转角度之间的差异，即估计的旋转矩阵 $\boldsymbol{R}$ 和真实旋转矩阵 $\boldsymbol{R}^*$ 之间的差值：

$$RRE = \arccos\left(\frac{trace\left(\left(\boldsymbol{R}^* \right)^{\mathrm{T}} \boldsymbol{R} \right) - 1}{2} \right) \tag{12}$$

RR 为配准召回率（Registration Recall，RR），即旋转误差和平移误差均小于预定义阈值的配准比例。

（2）配准结果　本文提出的点云配准算法在 KITTI odometry 数据集上配准结果的可视化呈现如图 6 所示。

图 6　在 KITTI odometry 数据集上配准效果

（3）对比实验　本文算法在基本上所有指标上都比其他方法取得了更好的结果，见表 5，除了 *RRE*，它比 AMGNet 落后 0.06 个百分点。结果表明，本文算法可以在周围更具判别性的重叠区域中检测到相应的点，取得了较好的配准效果。

表 5　在 KITTI odometry 数据集上进行测试

模型	*RTE*/cm	*RRE*/（°）	*RR*（%）
ICP[2]	25.9	0.72	96.0
Go-ICP[3]	9.5	0.57	96.6
PointNetLK[6]	12	0.53	97.5
PCRNet[9]	9.9	0.46	98.1
AMGNet[13]	6.8	0.30	98.8
Ours	6.2	0.32	98.8

5　结束语

本文提出混合注意力机制自适应权重的点云配准算法，来解决点云配准中的特征提取和特征融合问题，采用混合注意力机制来进行特征提取，采用自适应权重网络进行特征融合，通过自动调整权重实现更精确的数据融合，更好的完成点云配准。实验结果表明，本文所提出的方法在点云配准领域具有良好的配准效果，在泛化性和噪声鲁棒性方面也呈现出显著优势，怎样更好地计算姿态变换是下一步需要研究的问题。IM

参考文献

[1] LU W，WAN G，ZHOU Y，et al. Deepvcp：An end-to-end deep neural network for point cloud registration[C]// Proceedings of the IEEE/CVF international conference on computer vision，2019：12-21.

[2] SAFIZADEH N，AHMADI S H R. A comprehensive analysis of the false positive problem in SVD-based image watermarking[J]. Multimedia Tools and

Applications，2023，82 (14)：22275-22295.

[3] YANG J，LI H，CAMPBELL D，et al. Go-Icp：a globally optimal solution to 3D ICP point-set registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，38 (11)：2241-2254.

[4] 荆路，武斌，李先帅. 基于 SAC-IA 和 NDT 融合的点云配准方法 [J]. 大地测量与地球动力学，2021，41 (4)：378-381.

[5] SHEN Z，MA X，LI Y. A hybrid 3D descriptor with global structural frames and local signatures of histograms[J]. IEEE Access，2018 (6)：39261-39272.

[6] WANG Y，SOLOMON J M. Deep closest point：learning representation for point cloud registration[C]// Seoul：IEEE/CVF International Conference on Computer Vision，2019：3523-3532.

[7] WANG Y，SOLOMON J M. Prnet：self-supervised learning for partial to-partial registration[J]. Advances in Neural Information Processing Systems，2019 (32)：8814-8826.

[8] WU B，MA J，CHEN G，et al. Feature interactive representation for point cloud registration[C]// Montreal：IEEE/CVF International Conferenc on Computer Vision，2021：5530-5539.

[9] MEI Q，WANG F，TONG C，et al. PACNet：A high-precision point cloud registration network based on deep learning[C]//2021 3th International Conference on Wireless Communications and Signal Processing (WCSP)，IEEE，2021：1-5.

[10] CHOY C，DONG W，KOLTUN V. Deep global registration[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition，2020：2514-2523.

[11] YEW Z J，LEE G H. Rpm-net：Robust point matching using learned features[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition，2020：11824-11833.

[12] ZHU X，LYU S，WANG X，et al. TPH-YOLOv5：Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]// Proceedings of the IEEE/CVF international conference on computer vision，2021：2778-2788.

[13] 施成浩，陈谢沅澧，郭瑞斌，等. 面向点云配准和地点识别的多头旋转注意力网络 [J]. 控制理论与应用，2023，40 (12)：2187-2197.

[14] 刘旭珩，柏正尧，许祝，等. 结合注意力机制的多重引导点云配准网络 [J]. 计算机科学，2024，51 (2)：142-150.

[15] LIU S，HUANG D，WANG Y. Learning spatial fusion for single-shot object detection[J]. 2019 (25)：1-10.

[16] GUO Y，ZHAO L，SHI Y，et al. Adaptive weighted robust iterative closest point[J]. Neurocomputing，2022 (508)：225-241.

[17] MUNGOLI，NEELESH. Adaptive feature fusion：enhancing generalization in deep learning models[J]. 2023 (4)：1-5.

[18] 李杰，张洛维，王晓燕，等. 基于视锥距离和自适应权重卡尔曼滤波的多传感器融合算法 [J]. 中国公路学报，2024，37 (3)：194-203.

[19] ZHANG H. Multimodal sentiment analysis method based on hierarchical adaptive feature fusion network[J]. International Journal on Semantic Web and Information Systems (IJSWIS)，2024，20 (1)：1-23.

[20] WU Z，SONG S，KHOSLA A，et al. 3d shapenets：A deep representation for volumetric shapes[C]// Proceedings of the IEEE conference on computer vision and pattern recognition，2015：1912-1920.

[21] GEIGER A，LENZ P，URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//2012 IEEE conference on computer vision and pattern recognition，IEEE，2012：3354-3361.

收稿日期：2024-09-23