

PCRnet: 使用 PointNet 方法的点云配准网络

Vinit Sarode^{1*} Xueqian Li^{1*} Hunter Goforth³ Yasuhiro Aoki²
Rangaprasad Arun Srivatsan⁴ Simon Lucey^{1,3} Howie Choset¹
¹Carnegie Mellon University ²Fujitsu Laboratories Ltd. ³Argo AI. ⁴Apple

摘要

PointNet 最近已成为非结构化点云数据的流行表示形式, 允许将深度学习应用于目标检测、分割和形状完成等任务。然而, 最近的文献工作表明, 点网表示法对姿势错位非常敏感。本文提出了一个新的框架, 该框架使用点网表示来对齐点云, 并在跟踪、三维重建和姿态估计等应用中执行配准。我们开发了一个框架, 用于比较模板和源点云的点网功能, 以找到精确对齐它们的转换。根据点云形成的物体形状的先验信息, 我们的框架可以生成特定于形状或一般于看不见形状的方法。特定于形状的方法使用带有完全连接 (FC) 层的暹罗体系结构, 并且对噪声和数据中的初始偏差具有鲁棒性。我们进行了大量的模拟和真实实验, 以验证我们的方法的有效性, 并将其性能与最先进的方法进行了比较。

1. 引言

由于激光雷达、立体相机和结构光传感器的发展, 三维点云如今无处不在。因此, 人们对开发直接使用点云进行分类、分割、跟踪、映射等的算法越来越感兴趣。然而, 固有的结构缺陷给在深度学习体系结构中直接使用点云带来了困难。最近的一些发展, 如

PointNet[28]及其变体[29], 有助于克服其中一些困难, 从而产生了对目标检测和分割任务[27, 39]最先进的解决方法。

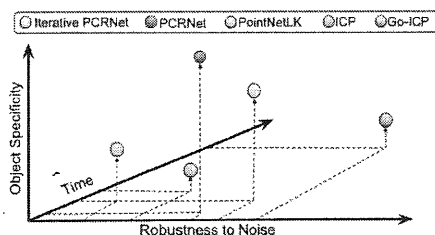


图 1: 基于不同配准方法对噪声的鲁棒性和与对象特异性相关的计算时间的比较。PCRNet 的迭代版本利用对象特定性来产生准确的结果。没有迭代的 PCRNet 计算速度更快, 但在精度上有点折衷。PointNetLK[2]具有良好的通用性, 但对噪声不鲁棒。ICP[7]与物体形状无关, 对于大型点云来说速度较慢, 而 Go ICP[37]的计算成本较高。之前的工作[39, 2]已经观察到, 点网的健壮性能要求点云相对于标准坐标系的偏差最小。虽然这在 ModelNet40[35]等合成数据集中存在, 但现实世界的数据很少与某个规范坐标系对齐。受迭代变压器网络 (IT Net) [39] 和 PointNetLK[2]的最新研究成果启

发,这项工作引入了点云配准网络(PCRNet),这是一个使用点网作为编码函数来估计两个点云之间不对齐的框架。值得注意的是,我们的方法可以直接处理点云以完成配准任务,而不需要手工制作的特征[31, 13]、体素化[25, 22]或网格生成[34]。根据对点云层形成的形状、噪声的存在和计算要求的先验知识,我们的框架为每种情况提供了一种非常适合的方法。我们的框架还为基于 PointNetLK 的一系列配准算法中的 PointNetLK (见图 1) 提供了额外的背景。

我们的方法使用暹罗体系结构中的 PointNet 将模板和源点云的形状信息编码为特征向量,并使用数据驱动技术估计将这两个特征对齐的姿势。与迭代最近点法(ICP)[7]及其变体[30]等形状不可知方法相比,在训练阶段使用形状特定的先验信息可以使我们对数据中的噪声具有鲁棒性。此外,我们发现 PointNetLK 方法使用经典对齐技术,如 Lucas Kanade (LK) 算法[21, 5]来对齐点网特征,对训练中看不到的形状具有良好的泛化性,但对噪声不具有鲁棒性。与传统的配准方法(如 ICP)不同,我们的方法不需要昂贵的最近点对应计算,从而提高了计算效率和对噪声的鲁棒性。此外,该方法是完全可微的,允许与其他深度网络轻

松集成,并且可以直接在 GPU 上运行,无需任何 CPU 计算。

总之,我们的贡献是:(1)提出了两种新的点云对齐算法,它们利用点网表示进行有效配准;(2)对这两种方法进行了彻底的实验验证,包括在模拟和真实数据上与点网 LK、ICP 和 GoICP 进行比较。

2. 相关工作

经典配准。迭代最近点(ICP)[7]仍然是最流行的点云配准技术之一,因为它易于实现,在许多场景中都能产生足够的结果。ICP 的扩展增加了一些功能,如提高计算效率[20, 9, 4]或提高精度[37]。然而,几乎所有 ICP 变量都依赖于最近点对应的显式计算,这一过程与点的数量的比例很差。此外,ICP 是不可微的(因为需要找到离散点对应关系),因此无法集成到端到端的深度学习管道中,从而抑制了将学习到的描述符应用于对齐的能力。

兴趣点方法计算并比较局部描述符以估计对齐[13, 15, 16]。兴趣点方法具有计算上有利的优势,然而,它们的使用通常限于具有可识别和独特特征的特征点云数据,这些特征在正在配准的点云之间是持久的[23, 26, 31]。

全局最优方法寻求找到最优解,而这种最优解无法通过迭代技

术（如 ICP）可靠地找到 [18, 19, 24]。我们用作基线的一个代表性例子是 Go ICP [37]，这是一种使用分支定界优化的技术。这些技术的特点是计算时间延长，这在很大程度上妨碍了它们在需要实时速度的应用中的使用。

PointNet. PointNet [28] 是第一个直接处理点云的深层神经网络，与物体的 2D 图像投影 [36, 8, 14]、体素表示 [25, 35, 41] 或图形表示 [34] 等替代表示不同。在更大的网络架构中，PointNet 已被证明对分类、语义分割、对象检测 [27] 和完成部分点云 [40] 等任务非常有用。[29] 中描述了用于估计局部特征描述符的 PointNet 扩展。文涛等人介绍了迭代变压器网络（IT 网）[39]，它使用点网来估计点云的标准方向，以提高分类和分割精度。[1] 中使用 PointNet 的全局描述符从 3D 数据中识别位置。深度网络中用于点云处理的损失函数是一个重要的考虑因素，我们将在第 3 节中详细讨论。[12] 中引入了地球移动器距离（EMD）和倒角距离（CD），而 [2] 中使用了估计和地面真值转换矩阵之间差异的 Frobenius 范数。

学习配准。 判别优化 [32] 和最近的逆合成判别优化 [33] 结合了手工制作的特征向量和学习的地

图集，用于点云配准任务。这些方法的缺点是点的数量是二次复杂的，并且由于特征向量和配准映射都被学习而缺乏泛化。[10] 中，深度自动编码器用于提取局部描述符，以配准大型室外点云。在 [38] 中，设计了一个学习兴趣点检测和描述符计算的神经网络，用于描述符匹配配准方法。Wang 等人使用局部邻域图对连接相邻点对的边执行卷积运算 [34]。PointNetLK [2] 通过最小化 PointNet 生成的固定长度全局描述符之间的距离来执行任意点云的配准，它与我们的工作关系最为密切，并作为基线。

3. 方法

点云是高度非结构化的，排列顺序不明确。在使用 PointNet 执行分类时，使用了对称池函数（如 maxpool）来提供输入置换的不变性。对称函数的输出向量称为全局特征向量。我们将表示模板点云 PT 和源 PS，以及点网函数 ϕ 。由于全局特征向量包含点云的几何信息和方向信息，因此可以通过比较特征向量来获得两个点云之间的变换。换句话说，我们计算刚体变换 $T \in SE(3)$ ，使 $\phi(PS)$ 和 $\phi(PT)$ 之间的差异最小化。

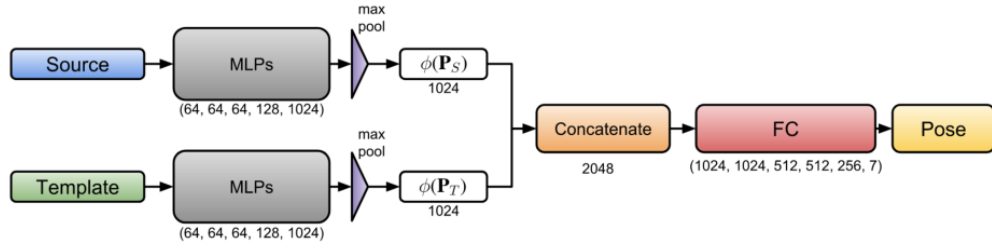


图 2:PCRNet 架构：该模型由五个大小为（64、64、64、128、1024）的 MLP 组成。源点云和模板点云通过一组以暹罗体系结构排列的 MLP 作为输入发送。使用最大池函数，我们获得全局特征。权重在 MLP 之间共享。这些功能被连接起来，并作为输入提供给五个完全连接的层 1024、1024、512、512、256 和一个大小为 7 的输出层。前三个输出值表示平移，标准化后的最后四个输出值表示旋转四元数。

3.1 PCRnet

本节介绍 PCRNet 体系结构。该体系结构的框图如图 2 所示。从传感器获得的点云数据称为源，与待配准对象的已知模型相对应的点云称为模板。该模型由五个多层感知器（MLP）组成，类似于大小为 64、64、64、128 和 1024 的点网架构。MLP 的布局类似于暹罗建筑[17]。源 \mathbf{P}_S 和模板 \mathbf{P}_T 都作为 mlp 的输入，mlp 以连体结构排列，并使用对称最大池函数查找全局特征向量 $\phi(\mathbf{P}_S)$ 和 $\phi(\mathbf{P}_T)$ 。权重在用于源和模板的 MLP 之间共享。

全局特征被连接起来，并作为输入提供给多个完全连接的层。在这项工作中，我们选择了五个完全

连接的层，因为它们似乎足以实现健壮的性能。我们尝试使用较少数量的 FC 层，但网络性能较差。

图 2 中红色块所示的 FC 层具有五个隐藏层，1024、1024、512、512、256，以及大小为 7 的输出层，其参数将表示估计的变换 T 。我们用来表示平移向量 t 的前三个输出值 $\in \mathbb{R}^3$ 和最后四个表示旋转四元数 $q \in \mathbb{R}^4$ ， $q^T q = 1$ 。通过这种方式，通过网络的单次向前传递或单次激发来估计对齐 $\phi(\mathbf{P}_S)$ 和 $\phi(\mathbf{P}_T)$ 的变换 T 。单发设计特别适合高速应用，这将在第 4 节中进一步讨论。

请注意，如果我们用传统的对齐算法（如 Lucas Kanade[21, 5]）替换网络中的 FC 层，那么最终的实现将类似于 PointNetLK[2]。

3.2 迭代 PCRNet

在本节中，我们介绍了一个网络，其迭代方案类似于 ICP 和 Lucas Kanade，用于图像对齐，如图 3 所示。我们保留了结构，但修改了单发 PCRNet 的层数。对于迭代实现，完全连接的层有三个大小为 1024、512、256 的隐藏层，以及一个大小为 7 的输出层。此外，在输出层之前还有一个额外的漏层，以避免过度拟合。我们根据经验观察到，与 PCRNet 相比，引入迭代使我们能够使用更少的隐藏层，同时获得稳健的性能。

在第一次迭代中，原始源点云和模板点云被提供给 PCRNet，PCRNet 预测它们之间的初始偏差 $T(1)$ 。在下一次迭代中，将 $T(1)$ 应用于源点云，然后将变换后的源点云和原始模板点云作为 PCRNet 的输入。在执行 n 次迭代后，我们通过组合每次迭代中的所有姿势，找到原始源点云和模板点云之间的整体变换：

$$T = T(n) \times T(n-1) \times \dots \times T(1) \quad (1)$$

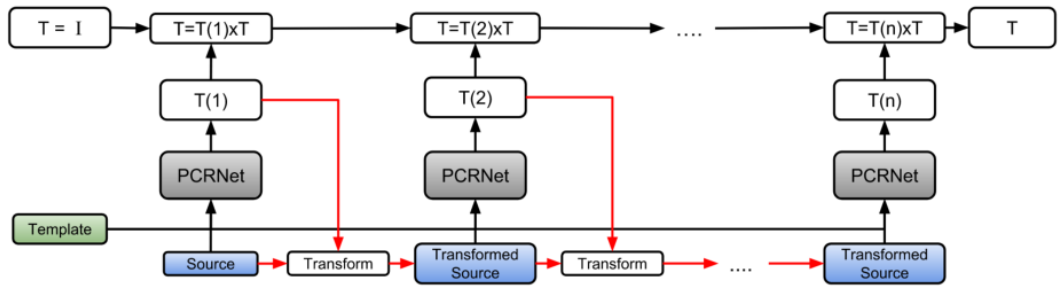


图 3: 迭代 PCRNet 架构: 迭代 PCRNet 使用图 2 中描述的 PCRNet 的修改形式，并迭代地改进 PCRNet 的估计。在第一次迭代中，源点云和模板点云被提供给 PCRNet，PCRNet 预测初始失准 $T(1)$ 。在下一次迭代中，使用 $T(1)$ 变换源点云，并将原始模板作为 PCRNet 的输入。在执行 n 次迭代后，我们将每个迭代的姿势组合起来，以找到原始源和模板之间的整体转换。

3.3 损失函数

用于训练配准网络的损失函数的目标应该是最小化源点云和模板点云中对应点之间的距离。该距

离可使用土方工程距离（EMD）功能计算，

$$\text{EMD}(\mathbf{P}_S^{\text{est}}, \mathbf{P}_T) = \min_{\psi: \mathbf{P}_S^{\text{est}} \rightarrow \mathbf{P}_T} \frac{1}{|\mathbf{P}_S^{\text{est}}|} \sum_{x \in \mathbf{P}_S^{\text{est}}} \|x - \psi(x)\|_2, \quad (2)$$

式中， \mathbf{P}_T 是模板点云， \mathbf{P}_S 是源点云，通过公式 1 中的估计变换 T 进行变换。该函数找到一个双射 ψ ，并基于 ψ 最小化对应点之间的距离。虽然损失函数还有很多其他选择，包括 Frobenius 范数[2]和 PoseLoss[36]，但我们发现 EMD 损失对于学习第 4 节中描述的迭代和单次 PCRNet 的训练数据最有效。

3.4 训练

在这项工作中，我们使用 ModelNet40 数据集[35]来训练网络。该数据集包含 40 个不同对象类别的 CAD 模型。我们根据人脸面积统一采样点，然后使用最远点算法[11]得到完整的点云。我们使用以下三种不同类型的数据集对网络进行训练：（1）多个类别的对象和每个类别的多个模型，（2）特定类别的多个模型，（3）特定类别的单个模型。我们选择这三个案例来展示基于点网的方法在不同对象特定性级别的数据上的性能。

我们在训练期间用 8 次迭代来训练迭代 PCRNet，观察到超过 8

次对结果几乎没有改善。在一些实验中，训练数据被高斯噪声破坏，这将在第二节详细讨论。这些网络接受了 300 个训练，学习率为 10%—每 3×10^6 步，批量为 32，指数衰减率为 0.7。网络参数在单个 NVIDIA GeForce GTX 1070 GPU 和 4.0GHz 的 Intel Core i7 CPU 上使用 Adam Optimizer 进行更新。

4. 结果

在本节中，我们将比较我们的网络在测试数据上的性能，测试数据包括多个对象类别、一个特定对象类别、一个来自训练数据集的特定对象以及在训练中看不到的对象。我们使用 ModelNet40 数据集[35]中的模型进行以下实验。模板点云标准化为一个单位框，然后将其平均值移动到原点。我们随机选择 5070 个欧拉角在 $[-45^\circ, 45^\circ]$ 以及 $[-1, 1]$ 单位范围内的转换值。我们在模板点云上应用这些刚性变换来生成源点云。在执行测试时，我们允许迭代 PCRNet 和 PointNetLK 的最大迭代次数为 20 次，而 ICP 的最大迭代次数选择为 100 次。除了最大迭代次数外，我

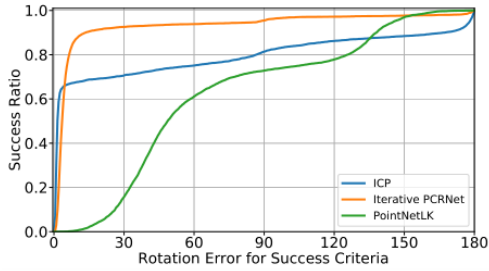
们 还 使 用 收 敛 准 则

$$\|\mathbf{T}_i \mathbf{T}_{i-1}^{-1} - \mathbf{I}\|_F < \epsilon,$$

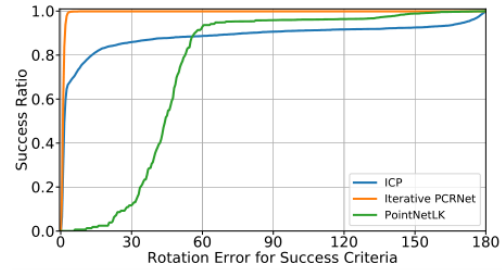
$\mathbf{T}_i \in \text{SE}(3)$ 是否在当前和以前的迭代中预测了转换, 以及转换的值被选为 10^{-7} 。

为了评估配准算法的性能, 我们使用的一个度量是曲线下面积 (AUC)。为 ICP、迭代 PCRNet 和 PointNetLK 生成了显示旋转误差成功率与成功标准 (以度为单位)

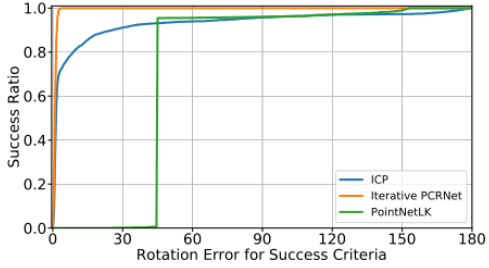
的曲线图。图 4 显示了这些曲线的示例。这些曲线图中曲线下方的面积除以 180, 在 0 和 1 之间进行归一化, 定义为 AUC 1。AUC 表示配准成功的度量, 因此 AUC 值越高, 网络性能越好。我们测量预测变换和地面真值变换之间的偏差, 并用轴角度表示, 我们将角度报告为旋转误差。至于翻译错误, 我们报告了地面真值和估计翻译向量之间差异的 L2 范数。



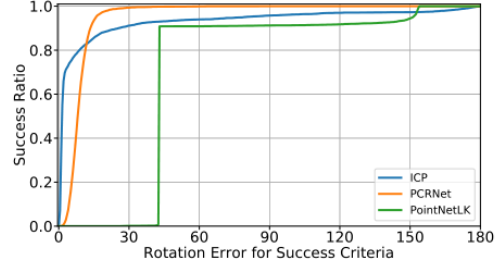
(a) Training and testing: Multiple object categories with noise.



(b) Training and testing: Multiple models of a category with noise



(c) Training and testing: Only one model with noise



(d) Trained on one model without noise and tested on data with noise

图 4: 第 4.2 节的结果。y 轴是成功实验的比率, x 轴显示最大旋转误差的值, 该值证明估算是成功的。(a), (b) 和 (c) 显示了使用三种不同类型的数据集将迭代 PCRNet 与 ICP 和 PointNetLK 进行比较的结果。我们观察到迭代 PCRNet 的优越性能, 因为我们的网络具有更多特定于模型/类别的信息。(d) PCRNet 在训练过程中没有看到噪声, 但使用噪声数据进行测试, 也显示出良好的性能, 比 ICP 和 PointNetLK 更快。第 2 节讨论了速度方面的考虑。

4.1. 概括性与特异性

在第一个实验中，迭代 PCRNet 和 PointNetLK 在 ModelNet40 的 20 个不同对象类别上进行训练，共有 5070 个模型。我们使用从 5 个对象类别中选择的 100 个模型进行测试，这些对象类别不在训练数据中（称为看不见的类别），在点云中没有噪声。我们确保使用同一对源点云和模板点云测试所有算法，以便进行公平比较。

我们使用多个对象类别训练迭代 PCRNet 和 PointNetLK，并使用不在训练数据中的对象类别对它们进行测试。在本实验的训练和测试期间，源数据中没有噪声。通过这些测试，我们发现 ICP 的 AUC 为 0.802，迭代 PCRNet 为 0.682，PointNetLK 为 0.998。

通过使用与测试数据相同类别的对象对网络进行训练，重复实验后，我们观察到迭代 PCRNet 的 AUC

大幅提高，从 0.682 提高到 0.972。ICP 和 PointNetLK 的 AUC 与之前相似，分别为 0.862 和 0.998，PCRNet 的 AUC 为 0.998。

这些结果强调，迭代 PCRNet 和 PCRNet 在使用特定于对象的信息进行再培训时，与 ICP 以及使用多个类别进行培训版本相比，可以提供更好的配准结果。当使用特定于对象的信息进行训练时，它们的性能与 PointNetLK 相当。然而，PointNetLK 在各种对象类别中比迭代 PCRNet 表现出更好的泛化能力，并且与 ICP 相比具有更好的性能（正如[2]所观察到的）。我们将这归因于 PCRNet 学习能力的固有局限性，即形状变化较大，而 PointNetLK 只需要学习点云表示，而不需要学习对齐任务。然而，在接下来的一组实验中，我们展示了 PCRNet 相对于 PointNetLK 和其他基线的明确优势，尤其是在存在噪声数据的情况下。

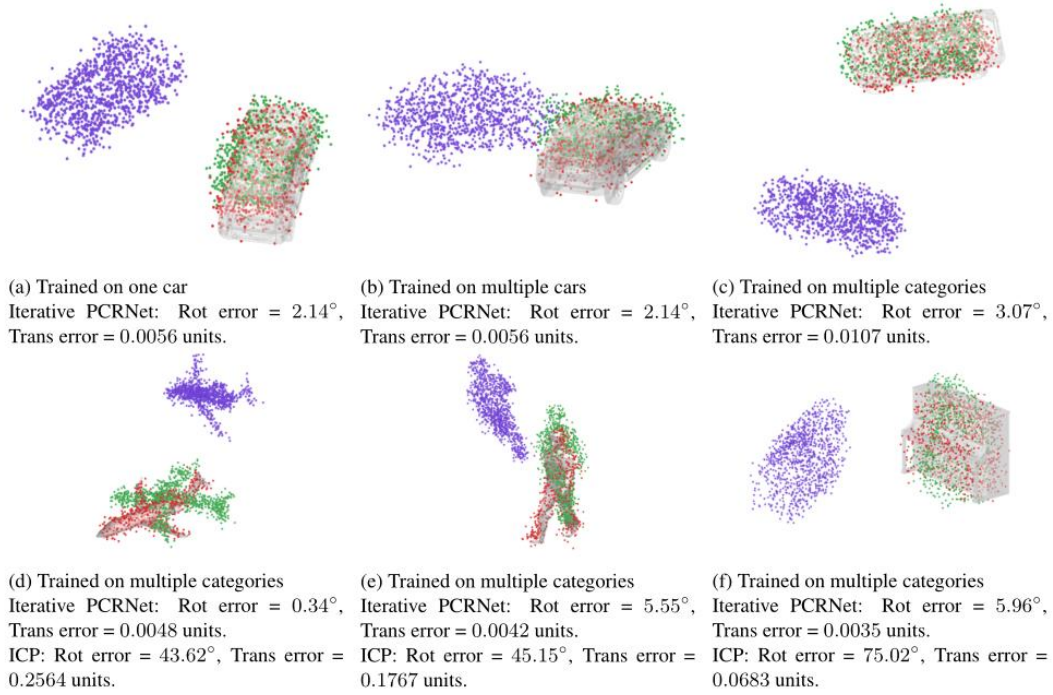


图 5：第 4.2 节的定性结果。对于每个示例，模板由灰色渲染的 CAD 模型显示，紫色点显示源的初始位置，红色点显示基于噪声数据的迭代 PCRNet 的收敛结果，绿色点显示 ICP 的结果。（a），（b），（c），（d）和（e）显示已看到类别的结果，而（f）显示未看到类别的结果。

4.2. 高斯噪声

为了评估我们的网络对噪声的鲁棒性，我们在源点用高斯噪声进行了实验。在我们的第一个测试中，我们使用第节中描述的数据集。4.1. 我们从高斯分布中为源点云中的每个点采样噪声，平均值为 0，标准偏差在 0 到 0.04 个单位的范围内变化。对于这些结果，我们训练了一个迭代 PCRNet 和一个 PointNetLK，它们使用 20 种不同

的对象类别和总共 5070 个模型，使用噪声源点云进行训练。

在测试过程中，我们将 ICP、PointNetLK 和迭代 PCRNet 与每个算法的源数据中的噪声进行比较。我们确保数据集具有相同的源点云和模板点云对，以便进行公平比较。图 4a 显示了结果。我们观察到，与 ICP 和 PointNetLK 相比，我们的迭代 PCRNet 具有更高的成功测试案例数和更小的旋转误差，这表明我们的迭代 PCRNet 对高斯噪声具有鲁棒性。值得注意的是，

PointNetLK 的性能最差，并且对噪声数据非常敏感。上述测试结果强调，迭代 PCRNet 在源数据中存在噪声的情况下工作得相当好，如果对象类别已知，其性能优于所有其他方法。

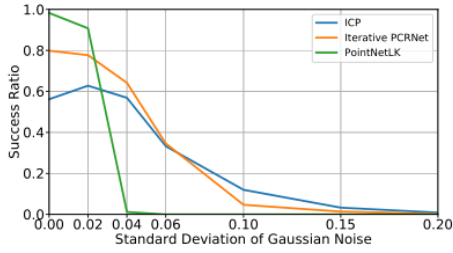


图 6: Sec 的结果。4.2. 迭代 PCRNet 和 PointNetLK 在高斯噪声的多个对象类别上进行训练，标准偏差的最大值等于 0.04。x 轴显示测试中使用的噪声的不同标准偏差值。PointNetLK 在没有噪声的情况下最准确，而迭代 PCRNet 对训练期间观察到的噪声水平（0.02-0.06）附近的噪声具有鲁棒性。

表 1: 第 4.3 节的结果。用于记录噪声数据的精度和计算时间比较。请注意，这两种 PCRNet 模型实现的 AUC 与 Go ICP 几乎相同，但速度要快几个数量级。

Algorithm	Rot. Error (deg)		Traj. Mean
	Mean	Std. Dev.	
PCRNet	8.82	4.82	0.0077
Iterative PCRNet	1.03	2.56	0.0085
PointNetLK [2]	51.80	29.63	0.8783
ICP [6]	11.87	31.87	0.0282
Go-ICP [37]	0.45	0.19	0.0016

对于第二个测试，我们使用了第 4.1 节中描述的数据集。4.1 和如上所述在源点云中添加高斯噪声。我们在特定的对象类别上训练网络，并使用 150 个车型在同一类别上测试它们。在源点云中的训练和测试过程中存在高斯噪声。图 4b 中的结果表明，迭代 PCRNet 的性能最好，并且有更多成功的测试用例。

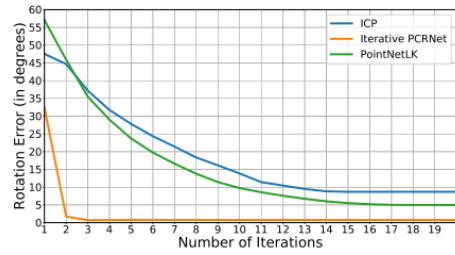


图 7: y 轴是预测和地面真值转换之间的旋转误差，x 轴显示了找到转换所执行的迭代次数。迭代 PCRNet 显示了在更少的迭代中对齐源点云和模板点云的能力。

我们比较了仅在一个噪声模型上进行训练和测试时网络的成功率（见图 4c）。迭代 PCRNet 再次显示出较高的成功率，优于 ICP 和 PointNetLK。最后，我们将无噪声训练并在噪声数据上测试的 PCRNet 与 ICP 和 PointNetLK 进行了比较。虽然没有 ICP 那么好，但我们的结果仍然具有竞争力，并且性能比 PointNetLK 好得多（见图 4d）。

我们使用在多个数据集上训练的迭代 PCRNet 和噪声数据测试，在图 5 中给出了定性结果。正如预期的那样，当在同一模型上训练迭代 PCRNet 时，其精度最高。然而，当在多个模型和多个类别上进行训练时，准确度只会略有下降，只要训练中有一些测试数据的表示，

就可以表现出良好的泛化。此外，如图 5 (c, f) 所示，对于一些看不见的类别，结果也是准确的，这表明了迭代 PCRNet 的可推广性。

图 6 显示了成功率与测试期间添加到源点云的噪声量的变化。迭代 PCRNet 和 PointNetLK 都在多个目标类别上进行训练，高斯噪声的最大标准偏差为 0.04。当噪声的标准偏差增加到 0.02 以上时，我们观察到 PointNetLK 性能突然下降。另一方面，迭代 PCRNet 在其训练的噪声范围 (0.02-0.06) 附近表现最佳，产生的结果与超出该噪声水平的 ICP 相当。这表明，与 PointNetLK 相比，我们的网络对

噪声的鲁棒性更强。

图 7 显示了不同方法的旋转误差与迭代次数的关系。请注意，与其他需要 15 次以上迭代的方法相比，迭代 PCRNet 只需要 3 次迭代就可以接近收敛。

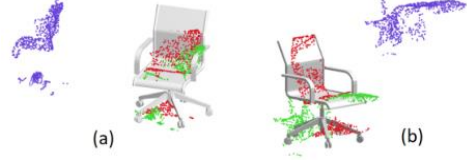


图 8：来自斯坦福 S3DIS 室内数据集的椅子点云配准[3]。CAD 模型显示来自 ModelNet40 的模板数据，紫色点来自 S3DIS 数据集，红色点代表迭代 PCRNet 估计，而绿色点代表 ICP 估计。



图 9：第 5 节的定性结果。斯坦福 S3DIS 室内数据集[3]中办公室场景中椅子的更换。红色皮革椅子显示了 ModelNet40[35] (a) 原始场景中更换的椅子。用 (b) ICP[7]，(c) 混合整数规划[19]和 (d) 迭代 PCRNet 中的配准取代了红色皮革椅子。

4.3. 计算速度比较

在本实验中，我们使用 ModelNet40 数据集中只有一个汽车模型的测试数据集，源数据中含有高斯噪声。我们应用 100 个随机选择的变换，欧拉角在 $[-45^\circ, 45^\circ]$ 以及 $[-1, 1]$ 单位范围内的转换值。这些网络都使用同一类别的多个模型（即汽车）进行训练。我们比较了迭代 PCRNet、PCRNet、PointNetLK、ICP 和 Go ICP 的性

能，如表 1 所示。我们报告了配准、计算时间和 AUC 后的旋转和平移误差。

结果表明，Go-ICP 收敛到全局最优解，旋转误差和平移误差非常小，但所需时间比迭代 PCRNet 多三个数量级，比 PCRNet 多五个数量级。Go ICP 的 AUC 值为 1，这意味着它在所有测试用例中都已收敛，而我们的网络具有次优 AUC 值。该实验表明，迭代 PCRNet 在精度方面与 Go ICP 类似，但计算速度更快，允许在许多实际应用中使使用。此外，虽然 PCRNet 不如迭代 PCRNet 和 Go ICP 精确，但在诸如对象检测和分割等应用中，作为

预对准步骤,精度可能足够好[39]。

5. 使用分段替换模型

为了展示真实数据的定性性能,我们演示了如何使用我们的方法在室内点云数据集中找到姿势并修改模型[3]。我们使用 PointNet[28]中引入的语义分割网络来执行模型替换,以预测斯坦福 S3DIS 室内数据集中每个对象的标签[3]。从场景中选择与椅子对应的点云,并使用迭代 PCRNet 将其配准到 ModelNet40 数据集中的椅子模型中,该 PCRNet 在多个带有噪声的对象类别上进行训练(见图 8)。

然后将迭代 PCRNet 预测的变换应用于从 ModelNet40 中选择的模型,并在原始椅子的位置进行替换,如图 9 所示。原始场景如图 9 (a) 所示。蓝色的椅子替换为与 ModelNet40 数据集不同型号红色椅子。图 9 (b) 显示了使用 ICP 的结果。请注意,ICP 无法将椅子调整到正确的姿势。当我们在 Sec 观察时。4.3 Go ICP 产生最准确的结果,在本例中,Go ICP 没有改善 ICP 的结果。因此,我们报告了另一种使用混合整数规划(MIP)[19]的全局配准方法的结果(见图 9 (c))。请注意,ICP 和 MIP 产生的结果都不如图 9 (d) 中迭代 PCRNet 产生的结果准确。这是因为 ICP、Go ICP 和 MIP 要求模板和源属于同一对象,同一类别对象之间的任何变化都可能导致配准不良。然而,我们的方法对同一类别内形状的变化非常有效,并产生了更好的结果。

6. 讨论和今后的工作

这项工作提出了一种新的数据

驱动框架,用于使用点网表示执行点云配准。该框架提供了一种使用特定于数据的信息生成高精度配准的方法,该方法对噪声和初始失准具有鲁棒性,同时计算速度比现有方法更快。该框架可以以迭代的方式实现,以获得与全球配准方法相当的高精度估计。该框架也可以在没有迭代的情况下实现,但与流行的配准方法相比,它有更深的层,可以产生两到五个数量级的速度改进。该框架说明了如何使用数据驱动技术来学习点云数据(包括噪声数据或类别特异性)中外观变化的分布,并使用这种学习到的先验知识在测试时表现得更好。最后,该框架还将文献中其他最新的基于 PointNetLK 的配准方法(如 PointNetLK)置于上下文中。未来的工作将涉及修改网络以处理部分和遮挡的点云,以及集成到更大的深层神经网络系统中,以完成多目标跟踪、样式转换、映射等任务。未来的工作可能会探索完全连接的配准层的学习能力对数据分布大小的限制。

参考文献

- [1] M. Angelina Uy and G. Hee Lee. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4470 - 4479, 2018. 2
- [2] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey. PointNetLK: Robust & Efficient Point Cloud Registration using PointNet. arXiv preprint arXiv:1903.05711, 2019. 1, 2, 3, 4, 5, 7
- [3] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3d semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2016. 7, 8
- [4] R. Arun Srivatsan, M. Xu, N. Zavallos, and H. Choset. Probabilistic pose estimation using a Bingham distribution-based linear filter. The International Journal of Robotics Research, 37(13-14):1610 - 1631, 2018. 2
- [5] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. International journal of computer vision, 56(3):221 - 255, 2004. 2, 3
- [6] P. Besl and N. D. McKay. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):239 - 256, Feb 1992. 7
- [7] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In Sensor Fusion IV: Control Paradigms and Data Structures, volume 1611, pages 586 - 607. International Society for Optics and Photonics, 1992. 1, 2, 8
- [8] H. Bristow, J. Valmadre, and S. Lucey. Dense semantic correspondence where every pixel is a classifier. In Proceedings of the IEEE International Conference on Computer Vision, pages 4024 - 4031, 2015. 2
- [9] B. Eckart, K. Kim, and J. Kautz. Fast and accurate point cloud registration using trees of gaussian mixtures. arXiv preprint arXiv:1807.02587, 2018. 2
- [10] G. Elbaz, T. Avraham, and A. Fischer. 3d point cloud registration for localization using a deep neural network auto-encoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4631 - 4640, 2017. 2
- [11] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi. The farthest point strategy for progressive image sampling. IEEE Transactions on Image

- Processing, 6(9):1305 – 1315, 1997. 4
- [12] H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 605 – 613, 2017. 2
- [13] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust global registration. In Symposium on geometry processing, volume 2, page 5, 2005. 1, 2
- [14] G. Georgakis, S. Karanam, Z. Wu, and J. Kosecka. Matching RGB Images to CAD Models for Object Pose Estimation. arXiv preprint arXiv:1811.07249, 2018. 2
- [15] J. Glover, G. Bradski, and R. B. Rusu. Monte carlo pose estimation with quaternion kernels and the distribution. In Robotics: Science and Systems, volume 7, page 97, 2012. 2
- [16] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan. 3D object recognition in cluttered scenes with local surface features: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(11):2270 – 2287, 2014. 2
- [17] D. Held, S. Thrun, and S. Savarese. Learning to track at 100 fps with deep regression networks. In European Conference on Computer Vision, pages 749 – 765. Springer, 2016. 3
- [18] M. B. Horowitz, N. Matni, and J. W. Burdick. Convex relaxations of SE(2) and SE(3) for visual pose estimation. In IEEE International Conference on Robotics and Automation (ICRA), pages 1148 – 1154. IEEE, 2014. 2
- [19] G. Izatt, H. Dai, and R. Tedrake. Globally Optimal Object Pose Estimation in Point Clouds with Mixed-Integer Programming. In International Symposium on Robotics Research, 12 2017. 2, 8
- [20] T. Jost and H. Hugli. A multi-resolution scheme ICP algorithm for fast shape registration. In Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission, pages 540 – 543. IEEE, 2002. 2
- [21] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. Proceedings of 7th IJCAI, 1981. 2, 3
- [22] Z. Ma, B. Liu, F. Zhou, and J. Chen. Point reg net: Invariant features for point cloud registration using in image-guided

- radiation therapy. *Journal of Computer and Communications*, 06:116–125, 01 2018. 1
- [23] A. Makadia, A. Patterson, and K. Daniilidis. Fully automatic registration of 3D point clouds. In *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on, volume 1, pages 1297–1304. IEEE, 2006. 2
- [24] H. Maron, N. Dym, I. Kezurer, S. Kovalsky, and Y. Lipman. Point registration via efficient convex relaxation. *ACM Transactions on Graphics (TOG)*, 35(4):73, 2016. 2
- [25] D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS)*, 2015 IEEE/RSJ International Conference on, pages 922–928. IEEE, 2015. 1, 2
- [26] M. Ovsjanikov, Q. Mérigot, F. Mémoli, and L. Guibas. One point isometric matching with the heat kernel. In *Computer Graphics Forum*, volume 29, pages 1555–1564. Wiley Online Library, 2010. 2
- [27] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas. Frustum pointnets for 3d object detection from RGB-D data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 918–927, 2018. 1, 2
- [28] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, 1(2):4, 2017. 1, 2, 8
- [29] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pages 5099–5108, 2017. 1, 2
- [30] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *3dim*, volume 1, pages 145–152, 2001. 2
- [31] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In *IEEE International Conference on Robotics and Automation*, pages 3212–3217. IEEE, 2009. 1, 2
- [32] J. Vongkulbhisal, F. De la Torre, and J. P. Costeira. Discriminative optimization: Theory and applications to point cloud registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4112, 2017. 2
- [33] J. Vongkulbhisal, B. Irastorza Ugalde, F. De la

- Torre, and J. P. Costeira. Inverse composition discriminative optimization for point cloud registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2993 – 3001, 2018. 2
- [34] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph CNN for learning on point clouds. arXiv preprint arXiv:1801.07829, 2018. 1, 2
- [35] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1912 – 1920, 2015. 1, 2, 4, 8
- [36] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox. PoseCNN: A convolutional neural network for 6d object pose estimation in cluttered scenes. arXiv preprint arXiv:1711.00199, 2017. 2, 4
- [37] 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, 38(11):2241 – 2254, 2016. 1, 2, 7
- [38] Z. J. Yew and G. H. Lee. 3dfeat-net: Weakly supervised local 3d features for point cloud registration. In European Conference on Computer Vision, pages 630 – 646. Springer, 2018. 2
- [39] W. Yuan, D. Held, C. Mertz, and M. Hebert. Iterative transformer network for 3d point cloud. arXiv preprint arXiv:1811.11209, 2018. 1, 2, 8
- [40] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert. PCN: Point Completion Network. In 3D Vision (3DV), 2018 International Conference on, 2018. 2
- [41] Y. Zhou and O. Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4490 – 4499, 2018. 2