# AutoURDF: Unsupervised Robot Modeling from Point Cloud Frames Using Cluster Registration

# AutoURDF:基于点云帧的无监督机器人建模与聚类配准

Jiong Lin , Lechen Zhang , Kwansoo Lee , Jialong Ning , Judah Goldfeder , Hod Lipson Columbia University

Jiong Lin , Lechen Zhang , Kwansoo Lee , Jialong Ning , Judah Goldfeder , Hod Lipson 哥伦比亚大学

https://autourdf.github.io/

# Abstract

# 摘要

Robot description models are essential for simulation and control, yet their creation often requires significant manual effort. To streamline this modeling process, we introduce AutoURDF, an unsupervised approach for constructing description files for unseen robots from point cloud frames. Our method leverages a cluster-based point cloud registration model that tracks the 6-DoF transformations of point clusters. Through analyzing cluster movements, we hierarchically address the following challenges: (1) moving part segmentation, (2) body topology inference, and (3) joint parameter estimation. The complete pipeline produces robot description files that are fully compatible with existing simulators. We validate our method across a variety of robots, using both synthetic and real-world scan data. Results indicate that our approach outperforms previous methods in registration and body topology estimation accuracy, offering a scalable solution for automated robot modeling.

机器人描述模型对于仿真和控制至关重要，但其创建通常需要大量手动工作。为了简化这一建模过程，我们提出了AutoURDF，一种从点云帧构建未知机器人描述文件的**无监督方法**。我们的方法利用**基于聚类的点云配准模型**，**跟踪点云簇的6自由度变换**。通过分析簇的运动，我们**分层**解决了以下挑战:**(1)运动部件分割，(2)身体拓扑推断，以及(3)关节参数估计。**完整的流程生成的机器人描述文件与现有仿真器完全兼容。我们在多种机器人上验证了我们的方法，使用了合成数据和真实扫描数据。结果表明，我们的方法在配准和身体拓扑估计准确性上优于以往方法，为自动化机器人建模提供了可扩展的解决方案。

# 1. Introduction

# 1. 引言

Accurate and structured representations of robots are essential for applications such as real-time control, motion planning, and physics-based simulation. Among these representations, robot description files, such as the Unified Robot Description Format (URDF) [43], explicitly capture robot geometry, kinematics, and dynamic properties. Over time, new formats such as MJCF for MuJoCo[47] and USD for NVIDIA’s Isaac Gym [28] have been introduced to enable scene descriptions and parallel simulations. Despite these advancements, customizing basic robot models still demands manual effort, often involving CAD model conversions [4] or tedious XML file modifications. This challenge has driven robotics and computer vision researchers to explore data-driven methods to automate the robot modeling process.

准确且结构化的机器人表示对于实时控制、运动规划和基于物理的仿真等应用至关重要。在这些表示中，机器人描述文件，如统一机器人描述格式(URDF)[43]，明确捕捉了机器人的几何、运动学和动态特性。随着时间的推移，新的格式如MuJoCo的MJCF[47]和NVIDIA的Isaac Gym的USD[28]被引入，以实现场景描述和并行仿真。尽管有这些进展，定制基本机器人模型仍然需要手动工作，通常涉及CAD模型转换[4]或繁琐的XML文件修改。这一挑战促使机器人和计算机视觉研究人员探索数据驱动的方法来自动化机器人建模过程。

**Robot self-modeling** [19] has emerged as a vital approach in enabling robots to autonomously discover their own body kinematics. Previous works in this field utilize sensorimotor data, integrating depth images [6], RGB images [12, 23] or IMU sensors [20] with motor signals to achieve accurate self-modeling. Another related field, - ticulated Object Modeling [21], focuses on reconstructing the kinematic structure of articulated rigid bodies from visual data. Prior research primarily targeted everyday objects -such as scissors, laptops, and drawers-which typically have only a few moving parts and relatively simple kinematic structures. Robots, however, consist of serially connected joints and, in some cases, multibranched links. Additionally, large-scale datasets representing diverse robot morphologies are still lacking. These factors make it challenging to effectively apply supervised learning methods developed for articulated objects to more complex robotic structures.

**机器人自建模**[19]已成为**使机器人自主发现其自身身体运动学**的重要方法。该领域的先前工作利用感觉运动数据，将深度图像[6]、RGB图像[12, 23]或IMU传感器[20]与电机信号集成，以实现准确的自建模。另一个相关领域， - 铰接物体建模[21]，专注于从视觉数据重建铰接刚体的运动学结构。先前的研究主要针对日常物体 -如剪刀、笔记本电脑和抽屉-这些物体通常只有少数运动部件和相对简单的运动学结构。然而，机器人由串联连接的关节组成，在某些情况下还有多分支连杆。此外，代表**多样化机器人形态的大规模数据集仍然缺乏**。这些因素使得将针对铰接物体开发的监督学习方法有效应用于更复杂的机器人结构具有挑战性。

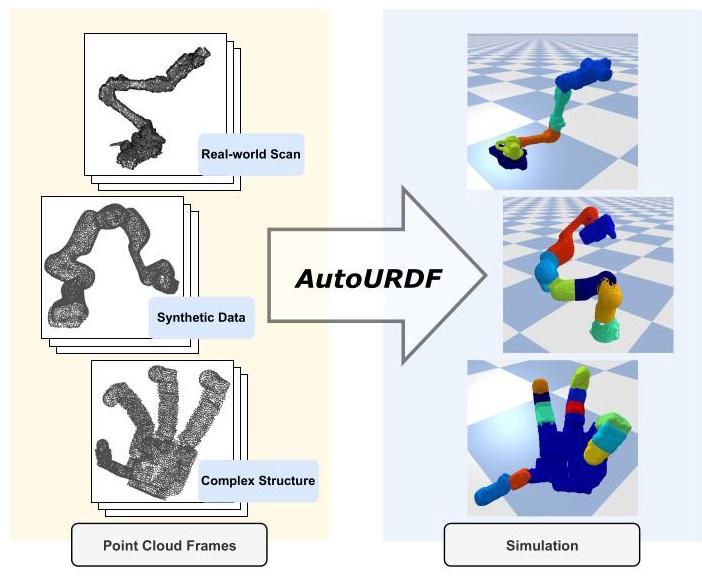


Figure 1. We present AutoURDF, a novel framework that derives robot description files from time-series point cloud frames. We validate our method using a diverse set of robots, evaluated on both synthetic data and real-world scan data..

图1. 我们提出了AutoURDF，一种从时间序列点云帧中导出机器人描述文件的新框架。我们在多种机器人上验证了我们的方法，使用了合成数据和真实扫描数据进行评估。

Here, we present AutoURDF, an unsupervised approach that constructs robot description files, specifically URDFs, from time-series point cloud frames. Rather than relying on sensorimotor data, our method derives the robot structure purely from visual data. Furthermore, it does not require ground-truth annotations or manual intervention, making it scalable to a wide range of robots. The key objective of this work is illustrated in Figure 1.

在这里，我们提出了AutoURDF，一种从时间序列点云帧构建机器人描述文件(特别是URDF)的无监督方法。我们的方法不依赖于感觉运动数据，而是纯粹从视觉数据中推导出机器人结构。此外，它不需要真实标注或手动干预，使其可扩展到广泛的机器人。这项工作的关键目标如图1所示。

Our method proceeds in a hierarchical manner. It starts by clustering the point cloud from the first frame based on point positions. Next, we compute a series of 6-DoF transformations across all time steps by registering the initial point clusters to each frame. From the motion patterns of these clusters, we segment them into distinct moving parts. Using the minimum spanning tree (MST) [18] and segmented parts, we identify link connections and infer the body topology. We then calculate transformation matrices for each child part relative to its parent, deriving the joint parameters. Finally, we generate the URDF file and validate the reconstructed structure in the PyBullet [9] simulator.

我们的方法以分层方式进行。它首先基于点位置对第一帧的点云进行聚类。接下来，我们通过将初始点云簇配准到每一帧，计算所有时间步长的一系列6自由度变换。从这些簇的运动模式中，我们将它们分割成不同的运动部件。使用最小生成树(MST)[18]和分割的部件，我们识别连杆连接并推断身体拓扑。然后，我们计算每个子部件相对于其父部件的变换矩阵，推导出关节参数。最后，我们生成URDF文件并在PyBullet[9]仿真器中验证重建的结构。

Our contributions are as follows:

我们的贡献如下:

* We propose a novel framework, AutoURDF, that constructs robot description models from point cloud frames without requiring prior knowledge of robot structure or kinematics.
* 我们提出了一种新框架AutoURDF，它从点云帧构建机器人描述模型，而不需要预先了解机器人结构或运动学。
* We design a cluster registration neural network to address point cloud registration for articulated objects. The extracted 6-DoF transformations are subsequently used for part segmentation, body topology inference, and joint estimation.
* 我们设计了一个聚类配准神经网络，用于解决铰接物体的点云配准问题。提取的6自由度变换随后用于部件分割、身体拓扑推断和关节估计。
* We implement a description file generation pipeline that outputs URDF files, including XML formatting and mesh file creation.
* 我们实现了一个描述文件生成管道，输出URDF文件，包括XML格式化和网格文件创建。

# 2. Related Work

# 2. 相关工作

Robot self-modeling. Task-agnostic self-modeling [19] enables robots to autonomously construct and update their own kinematic models from interaction data, without requiring specific task information. Recent advances in neural implicit representations have shown promising results in reconstructing dynamic 3D shapes [26, 39]. By applying spatial implicit functions to robot modeling, Chen et al. [6] demonstrated effective shape prediction by modeling robot morphology as a spatial query model conditioned on motor commands. Liu et al. [23] further extended this approach with a fully differentiable robot body rendering model. However, both methods require robot control information for model supervision, limiting their autonomy. Implicit function-based robot models are powerful for visual supervision and shape prediction, but they are not directly compatible with physics-based simulators, and accurately and efficiently calculating physical interactions with them remains challenging. On the other hand, Ledezma et al. [20] take a data-driven approach to robot modeling, explicitly representing the robot’s morphology through topology and kinematic parameters. However, their method requires IMU sensors on each moving part, which greatly increases setup complexity as the robot’s degrees of freedom (DoF) increase. Our method is motivated by similar goals to task-agnostic self-modeling approaches, aiming to enable autonomous robot modeling without specific task constraints. Unlike these approaches, our method relies solely on raw point cloud frames, without needing forward kinematics, degrees of freedom (DoF), or motor control data. Additionally, our approach maintains an explicit representation of the robot’s kinematics and geometry, ensuring direct compatibility with robot simulators.

机器人自我建模。任务无关的自我建模[19]使机器人能够从交互数据中自主构建和更新其运动学模型，而无需特定任务信息。神经隐式表示的最新进展在重建动态3D形状[26, 39]方面显示出有希望的结果。通过将空间隐函数 应用于机器人建模，Chen等人[6]通过将机器人形态建模为基于运动命令的空间查询模型，展示了有效的形状预测。Liu等人[23]进一步扩展了这一方法，提出了完全可微分的机器人身体渲染模型。然而，这两种方法都需要机器人控制信息进行模型监督，限制了其自主性。基于隐函数的机器人模型在视觉监督和形状预测方面非常强大，但它们与基于物理的模拟器不直接兼容，准确高效地计算物理交互仍然具有挑战性。另一方面，Ledezma等人[20]采用数据驱动的方法进行机器人建模，通过拓扑和运动学参数明确表示机器人的形态。然而，他们的方法需要在每个运动部件上安装IMU传感器，随着机器人自由度(DoF)的增加，这大大增加了设置的复杂性。我们的方法受到任务无关自我建模方法的类似目标的启发，旨在实现无需特定任务约束的自主机器人建模。与这些方法不同，我们的方法仅依赖于原始点云帧，不需要前向运动学、自由度(DoF)或运动控制数据。此外，我们的方法保持了机器人运动学和几何的明确表示，确保与机器人模拟器的直接兼容性。

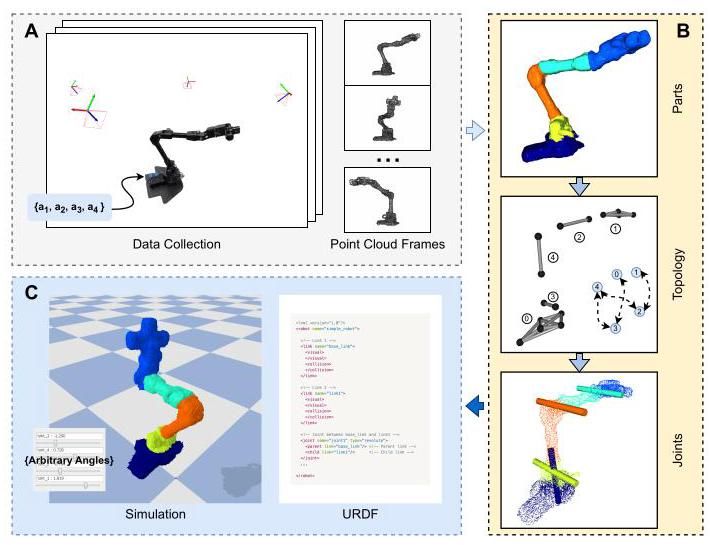


Figure 2. AutoURDF overview. Our method provides a complete pipeline for the automated construction of robot description files. (A). Data Collection: By commanding robots with randomly sampled motor angle sequences, we capture the corresponding time-series point cloud frames. (B). Three Substeps: We tackle the problem in three substeps: 1) part segmentation, 2) body topology inference, and 3) joint parameter estimation. (C). Description File Generation: The final output is a URDF file that defines the robot’s links, joints, and collision properties. We successfully build and simulate the description model for a WX200 robot arm from real-world scan data.

图2. AutoURDF概述。我们的方法提供了一个完整的自动化构建机器人描述文件的流程。(A) 数据收集:通过命令机器人执行随机采样的运动角度序列，我们捕获相应的时间序列点云帧。(B) 三个子步骤:我们通过三个子步骤解决问题:1) 部件分割，2) 身体拓扑推断，3) 关节参数估计。(C) 描述文件生成:最终输出是一个URDF文件，定义了机器人的链接、关节和碰撞属性。我们成功地从真实世界扫描数据中构建并模拟了WX200机械臂的描述模型。

Articulated objects modeling. Sapien dataset [31] has advanced research in supervised learning for articulated objects modeling. Most existing studies focus on everyday objects with only one or a few degrees of freedom , . Few works have explored the challenging task of reconstructing complex kinematic structures with more than 10 degrees of freedom. Among these, Real2Code[29] and URDFormer[7] are recent work that reconstructs the kinematic structure of actuated objects through code generation. However, these works assume moving parts are connected to a single parent part, which is not applicable to robotic structures with a series of joints for each kinematic branch. Watch-It-Move[35] and Reart[24] are two notable works that reconstruct the kinematic structure for multiple kinds of robots. However, these methods implement their own interaction mechanisms to move the object parts, which are not directly applicable to existing simulators. In contrast to the aforementioned methods, our approach does not rely on ground-truth annotations for training supervision and produces robot descriptions in commonly used formats.

关节物体建模。Sapien数据集[31]推动了关节物体建模的监督学习研究。大多数现有研究集中在仅有一个或几个自由度 ， 的日常物体上。很少有工作 探索了重建具有超过10个自由度的复杂运动学结构的挑战性任务。其中，Real2Code[29]和URDFormer[7]是最近通过代码生成重建驱动物体运动学结构的工作。然而，这些工作假设运动部件连接到单个父部件，这不适用于每个运动分支有一系列关节的机器人结构。Watch-It-Move[35]和Reart[24]是两个显著的工作，重建了多种机器人的运动学结构。然而，这些方法实现了自己的交互机制来移动物体部件，这些机制不直接适用于现有的模拟器。与上述方法相比，我们的方法不依赖于训练监督的真实标注，并以常用格式生成机器人描述。

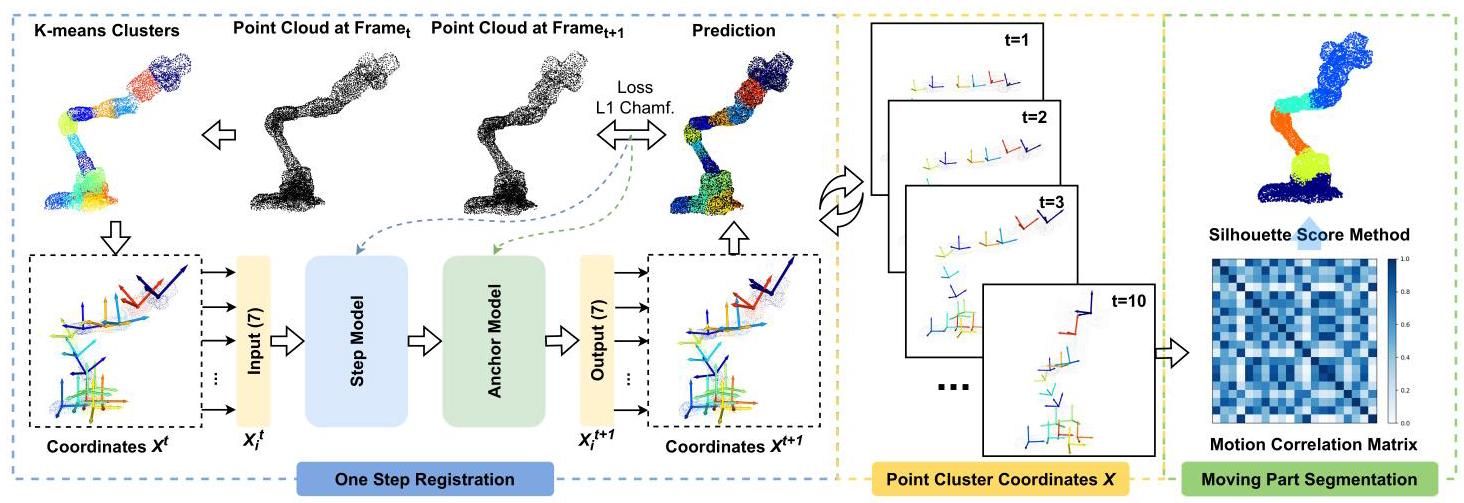


Figure 3. Point Cluster Registration and Part Segmentation. We tackle the part segmentation problem by extracting the 6-DoF transformations from the registration of a set of point clusters. For each pair of consecutive frames, we begin by initializing point clusters using the K-means clustering algorithm. Each point cluster is assigned a 6-DoF coordinate, , predicted from the previous registration step. Next, we employ a shared neural network with a positional encoder as the Step Model to predict the target frame’s point cluster coordinates, . Following the rotation representation study [5,10,52], we represent it as a 7-dimensional vector, combining Cartesian coordinates and quaternion orientation. The predicted coordinates are passed through an Anchor Model, which aligns the transformations to the initial coordinates. This process is repeated across frames in the sequence, tracing the motion of a set of 6-DoF coordinate frames. Finally, based on the time-series 6-DoF transformations, we compute a correlation matrix between the point clusters to segment distinct moving parts.

图3. 点簇配准和部件分割。我们通过从一组点簇的配准中提取6自由度变换来解决部件分割问题。对于每对连续帧，我们首先使用K-means聚类算法初始化点簇。每个点簇被分配一个6自由度坐标 ，该坐标从前一步的配准步骤中预测。接下来，我们采用一个带有位置编码器的共享神经网络作为步骤模型，预测目标帧的点簇坐标 。根据旋转表示研究[5,10,52]，我们将其表示为7维向量，结合笛卡尔坐标和四元数方向。预测的坐标通过锚点模型传递，该模型将变换对齐到初始坐标。这个过程在序列中的帧之间重复，追踪一组6自由度坐标系的运动。最后，基于时间序列的6自由度变换，我们计算点簇之间的相关矩阵以分割不同的运动部件。

To the best of our knowledge, our method is the first to construct functional description files for complex robotic structures (with up to 18 degrees of freedom and serially connected links) from unlabeled point cloud data.

据我们所知，我们的方法是首个从无标签点云数据中为复杂机器人结构(具有多达18个自由度和串联连接的连杆)构建功能描述文件的方法。

# 3. Methodology

# 3. 方法论

Our goal is to extract the information to create a robot description file from a series of point cloud frames. This process involves segmenting the robot’s links, identifying the connections between them, determining joint parameters-such as the orientation of the rotation axis and the location of the joint center-and representing the geometry of each link as mesh files. In this paper, we focus on revolute joints, the most common type in robotic structures, and assume a tree-like structure without closed kinematic loops.

我们的目标是从一系列点云帧中提取信息以创建机器人描述文件。此过程包括分割机器人的连杆、识别它们之间的连接、确定关节参数(如旋转轴的方向和关节中心的位置)以及将每个连杆的几何形状表示为网格文件。在本文中，我们专注于旋转关节，这是机器人结构中最常见的类型，并假设为树状结构，没有闭合的运动学环路。

# 3.1. Problem Formulation and Method Overview

# 3.1. 问题表述与方法概述

For each robot reconstruction task, we collected a time series of whole-body point cloud frames, represented as , where is the point cloud at time and is the number of points in each frame. Our goal is to extract three components:

对于每个机器人重建任务，我们收集了全身点云帧的时间序列，表示为 ，其中 是时间 的点云， 是每帧中的点数。我们的目标是提取三个组成部分:

* Segmented parts: denoted as link point clouds , where is the point cloud of the -th link, and is the total number of links.
* 分割部分:表示为连杆点云 ，其中 是第 个连杆的点云， 是连杆的总数。
* Body topology: represented as a graph , where is the set of link indices and is the set of edges connecting the parent and child links.
* 身体拓扑:表示为图 ，其中 是连杆索引的集合， 是连接父连杆和子连杆的边的集合。
* Joint parameters: represented as , where contains the 6-DoF parameters of the -th joint, and is the number of joints. For the tree-like structure, we have .
* 关节参数:表示为 ，其中 包含第 个关节的6自由度参数， 是关节的数量。对于树状结构，我们有 。

To achieve this goal, we design a rigid body registration algorithm to track the motion of a set of point clusters, initialized at time step . We define the set of point clusters as , where is a hyperpa-rameter that determines the number of clusters. The motion of the clusters is represented by a set of position and orientation coordinates, , where represents the Cartesian center of the point cluster, and is the quaternion orientation of the cluster. These coordinates are used to compute a correlation matrix between point clusters. Based on this correlation matrix, , we segment the clusters into distinct moving parts, denoted as the predicted links . We then apply the MST algorithm on the position components of to infer the body topology . Finally, we estimate the joint parameters by computing the homogeneous transformation matrices for pairs of clusters associated with connected links.

为实现这一目标，我们设计了一种刚体配准算法，用于跟踪在时间步 初始化的一组点簇的运动。我们将点簇集合定义为 ，其中 是一个超参数，用于确定簇的数量。簇的运动由一组位置和方向坐标 表示，其中 表示点簇的笛卡尔中心， 是簇的四元数方向。这些坐标用于计算点簇之间的相关矩阵。基于该相关矩阵 ，我们将簇分割为不同的运动部分，表示为预测的链接 。然后，我们在 的位置分量上应用MST算法，以推断身体拓扑 。最后，我们通过计算与连接链接相关联的簇对的齐次变换矩阵来估计关节参数 。

In summary, we derive from , by solving for and .

总之，我们通过求解 和 ，从 推导出 。

# 3.2. Cluster Registration and Part Segmentation

# 3.2. 簇配准与部分分割

Traditional point cloud registration algorithms [2] typically focus on rigid body alignment, while non-rigid registration methods handle more complex deformation scenarios, such as those involving the human body, cloth, or soft objects. In these cases, 3D coordinates (i.e., x, y, z) and graph representations are often used to model the object’s deformation. Robots, however, are articulated rigid bodies with 1-DoF joints, which can be viewed as a special case of non-rigid bodies. We aim to leverage the characteristics of coherent motion within the same rigid parts to extract accurate 6-DoF transformations for each point cluster.

传统的点云配准算法[2]通常专注于刚体对齐，而非刚性配准方法则处理更复杂的变形场景，例如涉及人体、布料或软物体的场景。在这些情况下，通常使用3D坐标(即x、y、z)和图表示来建模物体的变形。然而，机器人是具有1自由度关节的铰接刚体，可以视为非刚体的一种特殊情况。我们旨在利用同一刚体部分内连贯运动的特性，为每个点簇提取精确的6自由度变换。

As shown in Figure 3, we use a Step Model, denoted as , to update the 6-DoF coordinates and an Anchor Model, denoted as , to align the transformation to the initial coordinates. Inspired by PointNet [40, 41], we designed a lightweight neural network that computes coordinates using a shared MLP module. We extend the input from 3- dimensional Cartesian coordinates to a 7-dimensional vector to align with our cluster pose representation. We provide details on our model architecture and the comparison of 6- DoF pose representations in the appendix. The loss function in Eq. 1 consists of a single differentiable L1 Chamfer distance term between the predicted point clusters and the target whole-body point cloud. This loss function serves to optimize the network parameters, which in turn updates the output coordinates.

如图3所示，我们使用一个称为 的步进模型来更新6自由度坐标，并使用一个称为 的锚定模型将变换对齐到初始坐标。受PointNet[40, 41]的启发，我们设计了一个轻量级神经网络，该网络使用共享的MLP模块计算坐标。我们将输入从3维笛卡尔坐标扩展到7维向量，以与我们的簇姿态表示对齐。我们在附录中提供了模型架构的详细信息以及6自由度姿态表示的比较。公式1中的损失函数由预测点簇与目标全身点云之间的单个可微L1 Chamfer距离 项组成。该损失函数用于优化网络参数，从而更新输出坐标。

Where is the predicted point cloud for the entire body, integrated from the set of point clusters in the world coordinate frame at time step is the target whole body point cloud at the same time step. Step registration includes a resampling module, where the predicted coordinates are used as the center for clustering the next frame’s point cloud . Consequently, the step matching model performs registration over one frame interval. The Anchor Model takes as input the predicted coordinates from Step Registration and the point cluster from the first step initialization. It refines the coordinates to capture the transformation from the first step to the target step.

其中 是整个身体的预测点云，从世界坐标系中的点集群集合在时间步 整合而来， 是同一时间步的目标全身点云。步骤配准包括一个重采样模块，其中预测坐标 用作聚类下一帧点云 的中心。因此，步骤匹配模型在一帧间隔内执行配准。锚点模型 将步骤配准的预测坐标和第一步初始化的点集群作为输入，并细化坐标以捕捉从第一步到目标步骤的变换。

We demonstrate our registration pipeline, whose input is a whole body point cloud video and output are the clustered point cloud set and the correlated coordinates , in algorithm 1 . In the clustering and Chamfer distance calculations, we use point clusters in the world coordinate frame. For optimization in the Step Model and Anchor Model, we work with point clusters in the local coordinate system, represented by the tracked 6-DoF coordinates associated with each cluster. We employ two clustering algorithms: K-means[27] in the resampling module, where cluster centers are provided as input, and K-means++[1] in the initialization step, where the hyperparameter (the number of clusters) is specified.

我们展示了我们的配准流程，其输入是全身点云视频 ，输出是聚类点云集 和相关 坐标 ，如算法1所示。在聚类和Chamfer距离计算中，我们使用世界坐标系中的点集群。在步骤模型和锚点模型的优化中，我们使用局部坐标系中的点集群，由每个集群关联的跟踪6自由度坐标表示。我们采用两种聚类算法:重采样模块中的K-means[27]，其中聚类中心作为输入提供，以及初始化步骤中的K-means++[1]，其中指定了超参数 (聚类数量)。

Algorithm 1 Point Cluster Registration

算法1 点集群配准

Input: Point Cloud Frames

Output: Point Clusters , Coordinates

Initialize: at step

Add to in local coordinate frame)

for do

Load from

Step Registration:

where

Find

Update with

Resample:

where the is the position term of

Transfer

Anchor Alignment:

Add to

end for

As shown in equation 2 and equation 3, we compute the correlation matrix between the point clusters based on the transformation coordinates . The distance is calculated as the Euclidean distance between positions and the geodesic distance between orientations. Here, is a scaling parameter automatically calculated from the bounding box of the point cloud. The result is normalized to a range between 0 and 1 to facilitate alignment across multiple sequences.

如方程2和方程3所示，我们基于变换坐标 计算点集群之间的相关矩阵 。距离计算为位置之间的欧几里得距离和方向之间的测地距离。这里， 是从点云的边界框自动计算的缩放参数。结果归一化到0到1之间，以便于跨多个序列的对齐。

Clusters with highly correlated coordinate sequences (indicating minimal pose differences) are grouped into the same moving parts. We then merge the point clusters within each moving part to form the predicted links , which represent sets of point clouds that capture the shape of the robot links.

具有高度相关坐标序列(表示最小姿态差异)的集群被分组到相同的运动部分中。然后我们合并每个运动部分内的点集群以形成预测链接 ，这些链接表示捕捉机器人链接形状的点云集。

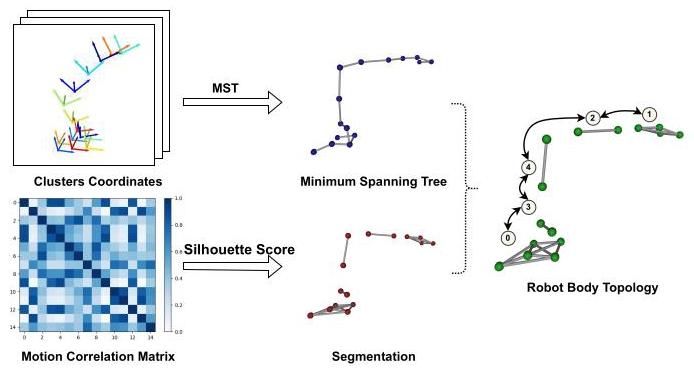


Figure 4. Topology Inference. We infer the robot’s body topology by merging the result of the MST algorithm on position terms and the segmented point clusters. In the segmentation graph, the nodes represent the center positions of the clusters; the edges represent the motion correlated between the connected nodes.

图4. 拓扑推断。我们通过合并位置项 上的MST算法结果和分割的点集群来推断机器人的身体拓扑。在分割图中，节点表示集群的中心位置；边表示连接节点之间的运动相关性。

We use the Silhouette Score method [44] to determine the optimal number of segments based on the Motion Correlation Matrix.

我们使用轮廓评分方法[44]基于运动相关矩阵确定最佳分割数量。

# 3.3. Topology Inference

# 3.3. 拓扑推断

We infer the body topology based on two graphs, 1). the segmented point clusters, which we represented as clusters index and the edges between the clusters, and 2). The minimum spanning tree is conditioned on the accumulated position distance between the point clusters. We demonstrate the process in Figure 4, and detailed algorithm in Appendix. The MST algorithm is applied to the point clusters, where the weight of each edge is determined by the accumulated position distance between the point clusters. We search the edges in the MST that connect the point clusters from different moving parts, which represent the connections between the robot links. The resulting graph represents the body topology of the robot, with the nodes as the link index and the edges as the connections between the links.

我们基于两个图推断身体拓扑，1). 分割的点集群，我们表示为集群索引和集群之间的边，以及2). 最小生成树基于点集群之间的累积位置距离。我们在图4中展示了该过程，并在附录中详细描述了算法。MST算法应用于点集群，其中每条边的权重由点集群之间的累积位置距离确定。我们搜索MST中连接来自不同运动部分的点集群的边，这些边表示机器人链接之间的连接。结果图 表示机器人的身体拓扑，节点作为链接索引，边作为链接之间的连接。

To comply with the URDF file format, we assign the base link of the robot as the root node. The links are ranked by their accumulated pose variation across consecutive frames, with the link exhibiting the least variation identified as the root node. Starting from this base link, we traverse the graph as a directed tree to establish the parent-child relationships between the links.

为了符合URDF文件格式，我们将机器人的基础链接指定为根节点。链接按其跨连续帧的累积姿态变化排序，具有最小变化的链接被识别为根节点。从这个基础链接开始，我们将图遍历为有向树，以建立链接之间的父子关系。

# 3.4. Joint Estimation

# 3.4. 关节估计

Referring to the merged point clusters, we compute the average coordinates of each link, which we denote as , where is the number of links. In joint estimation, we transfer the pose coordinates to the homogeneous transformation matrix . The base link is assigned the identity matrix , thus is the transformation matrix of the -th link in the world coordinate system. Therefore, the transformation matrix of the -th link in the coordinate system of the parent link is computed as , where is the transformation matrix from the base link to the -th link, and is the transformation matrix from the base link to the parent link of this -th link.

参考合并的点簇，我们计算每个链路的平均 坐标，记为 ，其中 是链路的数量。在联合估计中，我们将姿态坐标 转换为齐次变换矩阵 。基础链路被分配为单位矩阵 ，因此 是世界坐标系中第 个链路的变换矩阵。因此，第 个链路在父链路坐标系中的变换矩阵计算为 ，其中 是从基础链路到第 个链路的变换矩阵， 是从基础链路到第 个链路的父链路的变换矩阵。

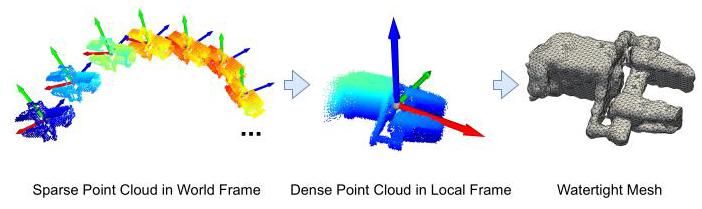


Figure 5. Point Cloud to Mesh. For each segmented link, we integrate the sparse point cloud data from the world coordinate system at each time step to form a dense point cloud in the local frame. The dense point cloud is constructed by combining 10 frames of data. We then apply the marching cubes[25] algorithm to convert this dense point cloud into a mesh file. The resulting mesh files are high-quality and watertight, as demonstrated by the example of the WX200 robot arm’s end-effector shown in the figure.

图5. 点云到网格。对于每个分割的链路，我们在每个时间步从世界坐标系中整合稀疏点云数据，以在局部坐标系中形成密集点云。密集点云通过结合10帧数据构建。然后我们应用行进立方体[25]算法将此密集点云转换为网格文件。生成的网格文件质量高且封闭，如图中WX200机器人臂末端执行器的示例所示。

According to the body topology graph, we locate every parent-child pair of links and compute the time-series transformation matrix of each child link in the coordinate system of the parent link. For each pair of links that connect at a 1-DoF revolute joint, the local transformation matrix yields a rotation matrix and a static translation vector . We extract the rotation axis orientation and joint center position from the rotation matrix and translation vector, respectively.

根据身体拓扑图，我们定位每对父子链路，并计算每个子链路在父链路坐标系中的时间序列变换矩阵。对于每个通过一自由度旋转关节连接的链路对，局部变换矩阵生成一个旋转矩阵 和一个静态平移向量 。我们分别从旋转矩阵和平移向量中提取旋转轴方向和关节中心位置。

# 3.5. Description File Generation

# 3.5. 描述文件生成

The URDF contains an XML file that defines the properties of the robot’s links and joints, along with a set of mesh files representing the geometric details of each link. Based on the predicted body topology and joint parameters, we generate the XML file specifying the links and joints of the robot. The pose coordinates of each link, , represent the transformation from the link’s local coordinate system to the world coordinate system, allowing us to record the link’s point cloud in the same coordinate system across all frames. This enables us to integrate the sparse point clouds to form a dense point cloud in the local frame. As shown in Figure 5, we combine data from 10 frames to construct a dense point cloud, which we then process using the marching cubes algorithm[25] to convert the dense representation into a high-quality, watertight mesh. The resulting mesh files are used to define the geometric properties of each link in the URDF file.

URDF包含一个XML文件，定义了机器人链路和关节的属性，以及一组表示每个链路几何细节的网格文件。基于预测的身体拓扑和关节参数，我们生成指定机器人链路和关节的XML文件。每个链路的姿态坐标 表示从链路的局部坐标系到世界坐标系的变换，使我们能够在所有帧中在同一坐标系中记录链路的点云。这使我们能够整合稀疏点云以在局部坐标系中形成密集点云。如图5所示，我们结合10帧数据构建密集点云，然后使用行进立方体算法[25]将密集表示转换为高质量且封闭的网格。生成的网格文件用于定义URDF文件中每个链路的几何属性。

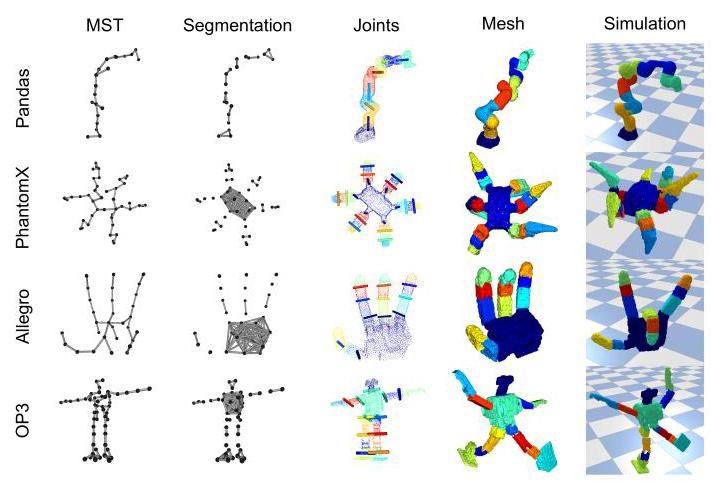


Figure 6. Qualitative Results. We illustrate the core stages of our method for various robots, including the minimum spanning tree, part segmentation, joint estimation results, mesh, and constructed URDFs for simulation. The colors of the joints indicate the parent link they are connected to.

图6. 定性结果。我们展示了我们方法在各种机器人中的核心阶段，包括最小生成树、部分分割、关节估计结果、网格和构建的用于仿真的URDF。关节的颜色表示它们连接的父链路。

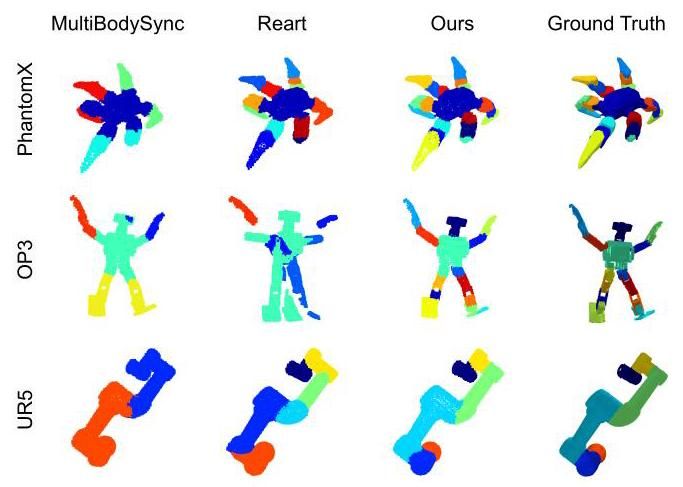


Figure 7. Qualitative comparison with MultiBodySync and Reart. Given the same input sequence (10 point cloud frames), we compare our method with MultiBodySync[13] and Reart[24] for part segmentation and point cloud registration results at the last frame.

图7. 与MultiBodySync和Reart的定性比较。给定相同的输入序列(10帧点云)，我们在最后一帧比较我们方法与MultiBodySync[13]和Reart[24]的部分分割和点云配准结果。

# 4. Experiments

# 4. 实验

# 4.1. Experiments Setup

# 4.1. 实验设置

Dataset. The validation dataset includes a variety of robots, including single-branch link robot arms (WX200, Franka Panda, UR5e) and multi-branch link robots (Bolt, Solo8, PhantomX, Allegro Hand, OP3 Humanoid), with degrees of freedom ranging from 5 to 18 . We evaluate our method on one point cloud video of a real-world scan of the WX200 robot arm and five point cloud videos of eight robots in simulated settings. Each video consists of 10 frames of point cloud data, down-sampled to 5,000 points per frame. For data collection, we fixed certain joints on the Allegro Hand and OP3 Humanoid and removed the end-effectors for the Franka Panda and UR5e. Further details on the data collection pipeline are provided in the Appendix.

数据集。验证数据集包括多种机器人，包括单分支连杆机械臂(WX200、Franka Panda、UR5e)和多分支连杆机器人(Bolt、Solo8、PhantomX、Allegro Hand、OP3 Humanoid)，自由度范围从5到18。我们在WX200机械臂的真实扫描点云视频和模拟环境中的八个机器人的五个点云视频上评估了我们的方法。每个视频由10帧点云数据组成，每帧下采样到5000个点。在数据收集过程中，我们固定了Allegro Hand和OP3 Humanoid的某些关节，并移除了Franka Panda和UR5e的末端执行器。数据收集流程的更多细节见附录。

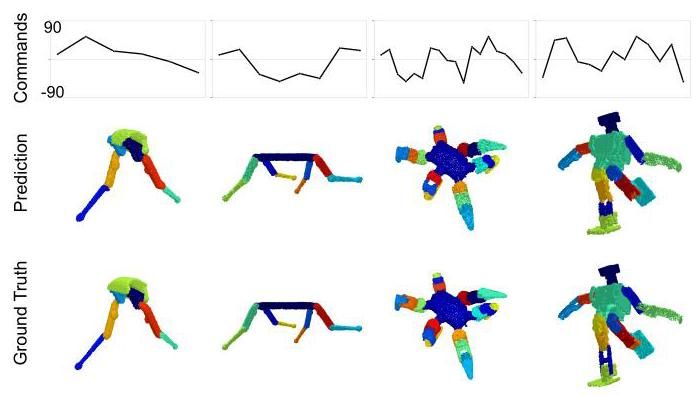


Figure 8. Qualitative Results for Simulating New Configurations We directly compare the predicted URDFs with the ground truth URDFs by applying the same motor commands. Au-toURDF generates accurate description files, producing control results closely matching those of the ground truth URDFs. We use the same colormap for predicted URDF and the ground-truth URDF.

图8。模拟新配置的定性结果。我们通过应用相同的电机命令，直接将预测的URDF与真实URDF进行比较。AutoURDF生成了准确的描述文件，产生的控制结果与真实URDF非常匹配。我们为预测的URDF和真实URDF使用了相同的颜色映射。

Metrics. We evaluate our method across four metrics: registration, body topology inference, joint estimation, and repose. Registration is measured using the average L1 Chamfer Distance (in millimeters), denoted as in our tables, which quantifies the alignment between the transformed whole-body point cloud from the first frame and each subsequent frame. Body topology inference is evaluated by comparing the predicted kinematic tree with the ground-truth kinematic tree using the tree editing distance , denoted as , to assess the accuracy of the inferred body topology. We measure joint estimation by calculating the average angular difference between joints (in degrees) and the average normal distance between the predicted and ground-truth joint axes (in millimeters). We denote these two metrics as and , unit in millimeters and degrees, respectively. Repose accuracy, denoted as , is evaluated by commanding two URDFs with random new motor angles and computing the surface point cloud L1 Chamfer distance. It is important to note that Chamfer distance depends on the robot’s scale and the density of the point cloud. In this work, we calculate the L1 Chamfer distance using the original scale of the robots, with 5,000 points sampled for alignment in all experiments.

指标。我们在四个指标上评估了我们的方法:配准、身体拓扑推断、关节估计和重新姿态。配准使用平均L1 Chamfer距离(以毫米为单位)进行测量，在我们的表格中表示为 ，它量化了从第一帧到后续每一帧的变换后的全身点云之间的对齐情况。身体拓扑推断通过使用树编辑距离 (表示为 )比较预测的运动学树与真实运动学树来评估推断的身体拓扑的准确性。我们通过计算关节之间的平均角度差(以度为单位)和预测与真实关节轴之间的平均法向距离(以毫米为单位)来测量关节估计。我们将这两个指标分别表示为 和 ，单位分别为毫米和度。重新姿态准确性(表示为 )通过命令两个URDF使用随机的新电机角度并计算表面点云的L1 Chamfer距离来评估。需要注意的是，Chamfer距离取决于机器人的尺度和点云的密度。在本工作中，我们使用机器人的原始尺度计算L1 Chamfer距离，在所有实验中对齐时采样5000个点。

# 4.2. Qualitative Experiments

# 4.2. 定性实验

Figure 6 illustrates the core stages of AutoURDF, including the MST, segmentation results represented as cluster skeletons, joint prediction results, mesh quality, and simulation examples in the PyBullet simulator. In Figure 7, we qualitatively compare our segmentation and registration results with MultibodySync [13] and Reart [24] using the same point cloud sequences. Our method produces more accurate segmentation and matching results across all three robot sequences. Figure 8 provides a direct comparison between the generated URDFs and the ground truth URDFs under the same new motor configurations. We also demonstrate the calculation of the Silhouette Score 9 based on the Motion Correlation matrices. The highest score indicates the optimal number of distinct moving parts, and therefore the number of degrees of freedom. Finally, we tested our method on real scanned point cloud data of a WX200 robot arm, commanding on all 5 servo motors, shown in Figure 10. Despite the high noise in the scan and misalignment of the coordinate systems across frames, our method successfully derives a functional URDF file for this robot with only 10 frames data, achieving relatively accurate kinematics and morphology computation.

图6展示了AutoURDF的核心阶段，包括MST、表示为聚类骨架的分割结果、关节预测结果、网格质量以及PyBullet模拟器中的模拟示例。在图7中，我们使用相同的点云序列，将我们的分割和配准结果与MultibodySync [13]和Reart [24]进行了定性比较。我们的方法在所有三个机器人序列上产生了更准确的分割和匹配结果。图8提供了在相同新电机配置下生成的URDF与真实URDF之间的直接比较。我们还展示了基于运动相关矩阵的轮廓分数9的计算。最高分数表示最佳的不同运动部件数量，因此也是自由度的数量。最后，我们在WX200机械臂的真实扫描点云数据上测试了我们的方法，命令所有5个伺服电机，如图10所示。尽管扫描中存在高噪声且帧间坐标系未对齐，我们的方法仅用10帧数据成功推导出了该机器人的功能性URDF文件，实现了相对准确的运动学和形态计算。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Metrics | Methods | WX200 | Panda | UR5e | Bolt | Solo | PhantomX | Allegro | OP3 | Mean Std |
| CD | Reart[24] | 9.33 | 18.81 | 15.86 | 10.39 | 11.14 | 14.73 | 6.38 | 44.95 |  |
| Ours | 7.49 | 13.56 | 12.84 | 8.41 | 9.77 | 10.88 | 5.80 | 8.30 |  |
| TED↓ | MultiBodySync[13] | 3.33 | 5.00 | 3.40 | 3.80 | 4.40 | 14.60 | 8.60 | 10.00 |  |
| Reart[24] | 0.83 | 2.40 | 4.40 | 3.20 | 4.00 | 13.00 | 6.00 | 11.60 |  |
| Ours | 0.33 | 1.40 | 0.60 | 1.75 | 0.00 | 4.00 | 4.00 | 6.00 |  |
| 指标 | 方法 | WX200 | 熊猫 | UR5e | 螺栓 | 独奏 | 幻影X | 快板 | OP3 | 均值 标准差 |
| CD | Reart[24] | 9.33 | 18.81 | 15.86 | 10.39 | 11.14 | 14.73 | 6.38 | 44.95 |  |
| 我们的 | 7.49 | 13.56 | 12.84 | 8.41 | 9.77 | 10.88 | 5.80 | 8.30 |  |
| TED↓ | 多体同步[13] | 3.33 | 5.00 | 3.40 | 3.80 | 4.40 | 14.60 | 8.60 | 10.00 |  |
| Reart[24] | 0.83 | 2.40 | 4.40 | 3.20 | 4.00 | 13.00 | 6.00 | 11.60 |  |
| 我们的 | 0.33 | 1.40 | 0.60 | 1.75 | 0.00 | 4.00 | 4.00 | 6.00 |  |

Table 1. Baseline Comparison

表1. 基线对比

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Metrics | Methods | WX200 | Panda | UR5e | Bolt | Solo | PhantomX | Allegro | OP3 | Mean Std |
| CD | w/o anchor m | 7.95 | 14.67 | 14.19 | 9.74 | 10.73 | 12.31 | 6.38 | 9.86 |  |
| w/o step m | 7.52 | 13.59 | 12.90 | 8.42 | 9.80 | 11.03 | 5.79 | 8.42 |  |
| Ours | 7.55 | 13.52 | 12.84 | 8.39 | 9.78 | 10.93 | 5.79 | 8.32 |  |
| TED↓ | w/o ori | 3.17 | 3.20 | 2.20 | 3.00 | 7.20 | 20.60 | 8.80 | 9.80 |  |
| w/o pos | 1.33 | 0.60 | 0.80 | 1.40 | 6.20 | 17.40 | 7.60 | 12.00 |  |
| Ours | 0.33 | 1.40 | 0.60 | 1.75 | 0.00 | 4.00 | 4.00 | 6.00 |  |
| 指标 | 方法 | WX200 | 熊猫 | UR5e | 螺栓 | 独奏 | 幻影X | 快板 | OP3 | 均值 标准差 |
| CD | 无锚点 m | 7.95 | 14.67 | 14.19 | 9.74 | 10.73 | 12.31 | 6.38 | 9.86 |  |
| 无步骤 m | 7.52 | 13.59 | 12.90 | 8.42 | 9.80 | 11.03 | 5.79 | 8.42 |  |
| 我们的 | 7.55 | 13.52 | 12.84 | 8.39 | 9.78 | 10.93 | 5.79 | 8.32 |  |
| TED↓ | 无方向 | 3.17 | 3.20 | 2.20 | 3.00 | 7.20 | 20.60 | 8.80 | 9.80 |  |
| 无位置 | 1.33 | 0.60 | 0.80 | 1.40 | 6.20 | 17.40 | 7.60 | 12.00 |  |
| 我们的 | 0.33 | 1.40 | 0.60 | 1.75 | 0.00 | 4.00 | 4.00 | 6.00 |  |

Table 2. Ablation Experiment

表2. 消融实验

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Metrics | Methods | WX200 | Panda | UR5e | Bolt | Solo | PhantomX | Allegro | OP3 | Mean Std |
|  | 1 Sequence | 11.51 | 42.82 | 35.76 | 17.86 | 13.41 | 14.76 | 8.66 | 32.25 |  |
| 5 Sequences | 11.10 | 45.60 | 21.24 | 12.39 | 13.03 | 13.85 | 8.30 | 21.85 |  |
|  | 1 Sequence | 1.16 | 4.46 | 3.32 | 5.44 | 5.34 | 4.39 | 6.20 | 11.60 |  |
| 5 Sequences | 1.17 | 3.95 | 2.17 | 4.94 | 1.61 | 2.50 | 3.97 | 7.37 |  |
|  | 1 Sequence | 1.91 | 7.37 | 7.39 | 4.13 | 3.59 | 3.91 | 7.85 | 15.81 |  |
| 5 Sequences | 1.13 | 4.16 | 2.88 | 2.44 | 1.86 | 1.74 | 3.68 | 7.11 |  |
| 指标 | 方法 | WX200 | 熊猫 | UR5e | 螺栓 | 独奏 | 幻影X | 快板 | OP3 | 均值 标准差 |
|  | 1个序列 | 11.51 | 42.82 | 35.76 | 17.86 | 13.41 | 14.76 | 8.66 | 32.25 |  |
| 5个序列 | 11.10 | 45.60 | 21.24 | 12.39 | 13.03 | 13.85 | 8.30 | 21.85 |  |
|  | 1个序列 | 1.16 | 4.46 | 3.32 | 5.44 | 5.34 | 4.39 | 6.20 | 11.60 |  |
| 5个序列 | 1.17 | 3.95 | 2.17 | 4.94 | 1.61 | 2.50 | 3.97 | 7.37 |  |
|  | 1个序列 | 1.91 | 7.37 | 7.39 | 4.13 | 3.59 | 3.91 | 7.85 | 15.81 |  |
| 5个序列 | 1.13 | 4.16 | 2.88 | 2.44 | 1.86 | 1.74 | 3.68 | 7.11 |  |

Table 3. Multi-Sequence Merging Experiment

表3. 多序列合并实验

# 4.3. Quantitative Results and Ablation Study

# 4.3. 定量结果与消融研究

Table 1 shows a quantitative comparison between our method and two previous works [13, 24] on our validation dataset. We compare our method with Reart for point cloud registration, using average end-point error measured in L1 Chamfer distance (CD). For robot body topology inference, we compare our method with both Reart and Multi-BodySync by computing the tree editing distance (TED) between the ground truth kinematic tree and the predicted ones. Our method consistently achieves lower scores on both metrics, indicating improved registration accuracy and body topology inference accuracy across all types of robots.

表1展示了我们的方法与之前两项工作[13, 24]在验证数据集上的定量比较。我们使用L1 Chamfer距离(CD)测量的平均端点误差，将我们的方法与Reart进行点云配准比较。对于机器人身体拓扑推断，我们通过计算真实运动学树与预测树之间的树编辑距离(TED)，将我们的方法与Reart和Multi-BodySync进行比较。我们的方法在两项指标上均取得了更低的分数，表明在所有类型的机器人上，配准精度和身体拓扑推断精度均有所提高。

To validate the contributions of the Step Model and Anchor Model in our point cluster registration algorithm, we compare the matching results of our complete pipeline with results using only the Step Model and only the Anchor Model. The complete method achieves superior performance, as shown in Table 2. Additionally, to validate the contributions of the two distance terms in Equation 3, we conduct ablation experiments comparing the complete equation with versions that include only the position term or only the orientation term. As shown in Figure 11, using only the position distance struggles to produce accurate segmentation, while using only the orientation term is easily misled by links with similar rotational motions. Three links of the Solo8 robot are segmented into the same group due to their similar rotational motion within this point cloud sequence.

为了验证Step Model和Anchor Model在我们的点簇配准算法中的贡献，我们将完整管道的匹配结果与仅使用Step Model和仅使用Anchor Model的结果进行比较。完整方法表现出更优的性能，如表2所示。此外，为了验证公式3中两个距离项的贡献，我们进行了消融实验，比较完整公式与仅包含位置项或仅包含方向项的版本。如图11所示，仅使用位置距离难以产生准确的分割，而仅使用方向项则容易被具有相似旋转运动的连杆误导。Solo8机器人的三个连杆由于在该点云序列中具有相似的旋转运动，被分割到同一组中。

Table 3 presents a comparison of our method, extending the input point cloud data from 1 sequence ( 10 frames) to 5 sequences (50 frames). We evaluate the reposed robot shape error using the metric, the joint axis normal distance with the metric, and the joint angle error with the metric. As the dataset scales, AutoURDF improves performance in 7 out of 8 robots for repose and joint distance and in all robots for joint angle accuracy. With 50 frames of point cloud data from the WX200 robot arm, we achieve an average joint angle difference of 1.13 degrees and a joint normal distance of across its 5 joint axes.

表3展示了我们的方法在将输入点云数据从1个序列(10帧)扩展到5个序列(50帧)时的比较。我们使用 指标评估重新定位的机器人形状误差，使用 指标评估关节轴法线距离，使用 指标评估关节角度误差。随着数据集的扩展，AutoURDF在8个机器人中的7个在重新定位和关节距离上表现提升，在所有机器人中关节角度精度均有所提高。使用WX200机械臂的50帧点云数据，我们在其5个关节轴上实现了平均1.13度的关节角度差异和 的关节法线距离。

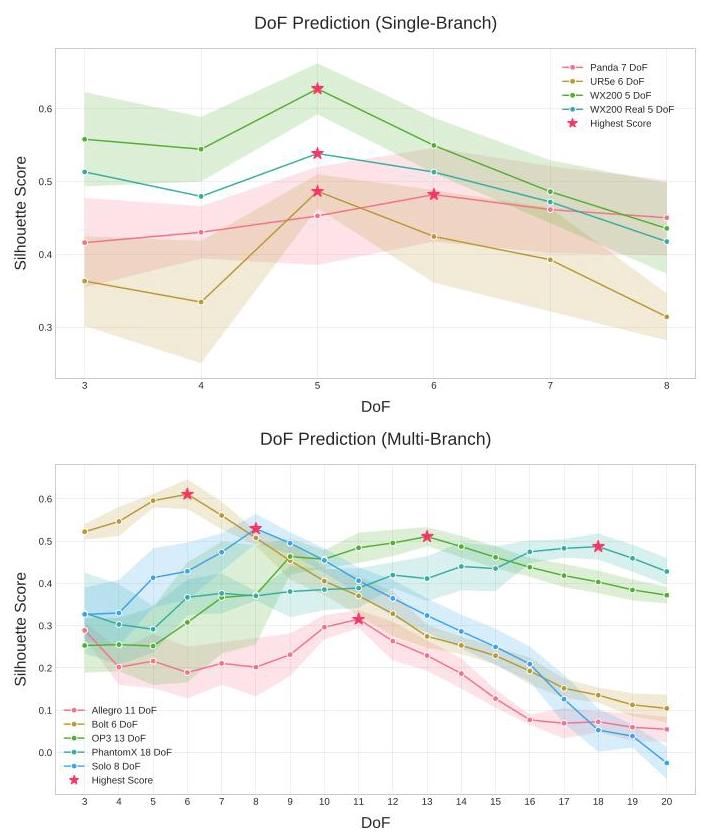


Figure 9. DoF Prediction. We plot the Silhouette Score for eight different robots across five point cloud sequences. The highest score indicates the optimal number of distinct moving parts, and thus the number of DoF. By averaging the Silhouette Scores across sequences, our method accurately predicts the correct number of DoF for all eight robots.

图9. 自由度预测。我们绘制了八个不同机器人在五个点云序列上的轮廓分数。最高分数表示最优的独立运动部件数量，即自由度数量。通过对序列的轮廓分数进行平均，我们的方法准确预测了所有八个机器人的正确自由度数量。

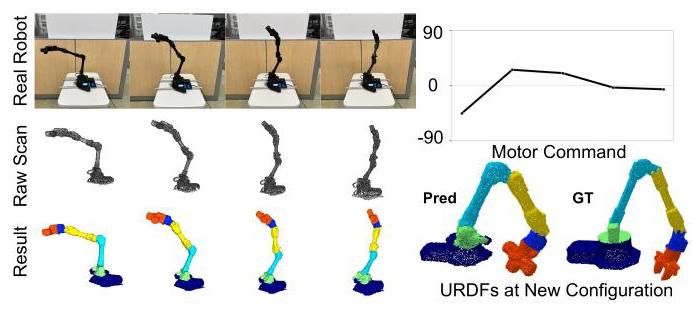


Figure 10. Real World Experiment. The first row shows the WX200 robot arm at four different time steps under random motion commands. The second row displays the corresponding whole-body point cloud frames. The third row presents alignment results with meshes. The right column shows the constructed URDF compared to the ground truth under a random new configuration.

图10. 真实世界实验。第一行展示了WX200机械臂在随机运动命令下的四个不同时间步。第二行展示了相应的全身点云帧。第三行展示了与网格的对齐结果。右列展示了构建的URDF与随机新配置下的真实情况的比较。

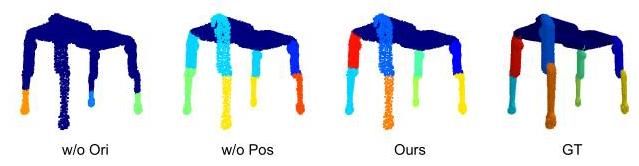


Figure 11. Ablation study on Correlation Matrix. We compare the distance equation 3 under three conditions: without the orientation term , without the position term , and with the complete calculation.

图11. 相关矩阵的消融研究。我们比较了公式3在三种条件下的距离:不包含方向项 、不包含位置项 以及完整计算。

Among all these improvements, AutoURDF stands out for its speed as an unsupervised method. On the WX200 robot data, registration takes 50 seconds, and URDF construction takes 12 seconds on an NVIDIA 3090 GPU. In comparison, Reart[24] refinement training, with the flow model pre-trained, takes 35.5 minutes (2,120 seconds) on the same machine.

在所有这些改进中，AutoURDF作为一种无监督方法，其速度尤为突出。在WX200机器人数据上，配准耗时50秒，URDF构建在NVIDIA 3090 GPU上耗时12秒。相比之下，Reart[24]的细化训练，在预训练流模型的情况下，在同一台机器上耗时35.5分钟(2,120秒)。

# 5. Discussion and Conclusion

# 5. 讨论与结论

# 5.1. Limitations and Future Work

# 5.1. 局限性与未来工作

Our method has three key limitations. First, we collected collision-free robot motion sequences, there is no self-collision or interaction with the environment, so our method does not learn dynamic information, and the resulting URDF files lack mass and moment of inertia data. Second, segmenting more complex structures requires longer point cloud sequences. For example, with the OP3 humanoid robot, our method does not produce clean segmentation results due to the structure’s complexity. Extending the length of the point cloud video could improve segmentation accuracy for such complex robots. Third, this work focuses on robots with revolute joints and tree-like structures. Future work could extend the method to handle non-revolute joints, such as prismatic or spherical joints, and robots with parallel or closed-loop structures.

我们的方法有三个主要局限性。首先，我们收集了无碰撞的机器人运动序列，没有自碰撞或与环境的交互，因此我们的方法没有学习动态信息，生成的URDF文件缺乏质量和惯性矩数据。其次，分割更复杂的结构需要更长的点云序列。例如，对于OP3人形机器人，由于结构的复杂性，我们的方法未能产生清晰的分割结果。延长点云视频的长度可以提高此类复杂机器人的分割精度。第三，本工作专注于具有旋转关节和树状结构的机器人。未来的工作可以扩展该方法以处理非旋转关节，如棱柱或球形关节，以及具有平行或闭环结构的机器人。

# 5.2. Conclusion

# 5.2. 结论

In this paper, we present a novel approach for constructing robot description files directly from point cloud data. Our method employs 6-DoF cluster registration to segment moving parts, infer body topology, and estimate joint parameters, achieving accurate joint direction estimation and whole-body shape reposing. The output URDF-formatted robot description file is compatible with widely used robot simulators. Experimental results show that our method outperforms previous approaches in the accuracy of robot point cloud registration and body topology estimation, offering an efficient and scalable solution for automated robot modeling from visual data.

本文提出了一种直接从点云数据构建机器人描述文件的新方法。我们的方法采用6自由度(6-DoF)聚类配准来分割运动部件、推断身体拓扑结构并估计关节参数，从而实现精确的关节方向估计和全身形状重构。输出的URDF格式机器人描述文件与广泛使用的机器人模拟器兼容。实验结果表明，我们的方法在机器人点云配准和身体拓扑结构估计的准确性上优于以往的方法，为基于视觉数据的自动化机器人建模提供了一种高效且可扩展的解决方案。

References

参考文献

[1] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006. 4

David Arthur 和 Sergei Vassilvitskii. k-means++:谨慎初始化的优势. 技术报告, 斯坦福大学, 2006. 4

[2] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):239-256, 1992. 4

P.J. Besl 和 Neil D. McKay. 一种3D形状配准的方法. IEEE模式分析与机器智能汇刊, 14(2):239-256, 1992. 4

[3] Gunilla Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. IEEE Transactions on pattern analysis and machine intelligence, 10(6):849-865, 1988. 4

Gunilla Borgefors. 分层倒角匹配:一种参数化边缘匹配算法. IEEE模式分析与机器智能汇刊, 10(6):849-865, 1988. 4

[4] Stephen Brawner. Solidworks to urdf exporter. https:// github.com/ros/solidworks\_urdf\_exporter, 2024. Accessed: 2024-11-13. 1

Stephen Brawner. Solidworks到URDF导出工具. https://github.com/ros/solidworks\_urdf\_exporter, 2024. 访问日期: 2024-11-13. 1

[5] Romain Brégier. Deep regression on manifolds: a 3d rotation case study. In 2021 International Conference on 3D Vision (3DV), pages 166-174. IEEE, 2021. 3

Romain Brégier. 流形上的深度回归:3D旋转案例研究. 2021年国际3D视觉会议(3DV), 166-174页. IEEE, 2021. 3

[6] Boyuan Chen, Robert Kwiatkowski, Carl Vondrick, and Hod Lipson. Fully body visual self-modeling of robot morphologies. Science Robotics, 7(68):eabn1944, 2022. 1, 2

Boyuan Chen, Robert Kwiatkowski, Carl Vondrick, 和 Hod Lipson. 机器人形态的全身视觉自建模. 科学机器人, 7(68):eabn1944, 2022. 1, 2

[7] Zoey Chen, Aaron Walsman, Marius Memmel, Kaichun Mo, Alex Fang, Karthikeya Vemuri, Alan Wu, Dieter Fox, and Abhishek Gupta. Urdformer: A pipeline for constructing articulated simulation environments from real-world images. arXiv preprint arXiv:2405.11656, 2024. 2

Zoey Chen, Aaron Walsman, Marius Memmel, Kaichun Mo, Alex Fang, Karthikeya Vemuri, Alan Wu, Dieter Fox, 和 Abhishek Gupta. Urdformer:从真实世界图像构建关节仿真环境的流程. arXiv预印本 arXiv:2405.11656, 2024. 2

[8] Ruihang Chu, Zhengzhe Liu, Xiaoqing Ye, Xiao Tan, Xi-aojuan Qi, Chi-Wing Fu, and Jiaya Jia. Command-driven articulated object understanding and manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8813-8823, 2023. 2

Ruihang Chu, Zhengzhe Liu, Xiaoqing Ye, Xiao Tan, Xiaojuan Qi, Chi-Wing Fu, 和 Jiaya Jia. 命令驱动的关节物体理解与操作. IEEE/CVF计算机视觉与模式识别会议(CVPR)论文集, 8813-8823页, 2023. 2

[9] Erwin Coumans. Bullet physics simulation. In ACM SIG-GRAPH 2015 Courses, page 1. ACM, 2015. 2

Erwin Coumans. Bullet物理模拟. ACM SIGGRAPH 2015课程, 第1页. ACM, 2015. 2

[10] A René Geist, Jonas Frey, Mikel Zobro, Anna Levina, and Georg Martius. Learning with rotations, a hitchhiker’s guide to so (3). arXiv preprint arXiv:2404.11735, 2024. 3

A René Geist, Jonas Frey, Mikel Zobro, Anna Levina, 和 Georg Martius. 学习 旋转，SO(3)的旅行者指南. arXiv预印本 arXiv:2404.11735, 2024. 3

[11] Nick Heppert, Muhammad Zubair Irshad, Sergey Zakharov, Katherine Liu, Rares Andrei Ambrus, Jeannette Bohg, Ab-hinav Valada, and Thomas Kollar. Carto: Category and joint agnostic reconstruction of articulated objects. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21201-21210, 2023. 2

Nick Heppert, Muhammad Zubair Irshad, Sergey Zakharov, Katherine Liu, Rares Andrei Ambrus, Jeannette Bohg, Abhinav Valada, 和 Thomas Kollar. Carto:类别和关节无关的关节物体重建. IEEE/CVF计算机视觉与模式识别会议论文集, 21201-21210页, 2023. 2

[12] Yuhang Hu, Jiong Lin, and Hod Lipson. Teaching robots to build simulations of themselves. arXiv preprint arXiv:2311.12151, 2023. 1

Yuhang Hu, Jiong Lin, 和 Hod Lipson. 教机器人构建自身的模拟. arXiv预印本 arXiv:2311.12151, 2023. 1

[13] Jiahui Huang, He Wang, Tolga Birdal, Minhyuk Sung, Federica Arrigoni, Shi-Min Hu, and Leonidas Guibas. Multi-bodysync: Multi-body segmentation and motion estimation via 3d scan synchronization. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 7104-7114. IEEE, 2021. 1, 6, 7

Jiahui Huang, He Wang, Tolga Birdal, Minhyuk Sung, Federica Arrigoni, Shi-Min Hu, 和 Leonidas Guibas. Multi-bodysync:通过3D扫描同步进行多体分割和运动估计. 2021年IEEE/CVF计算机视觉与模式识别会议(CVPR), 7104-7114页. IEEE, 2021. 1, 6, 7

[14] Ajinkya Jain, Rudolf Lioutikov, Caleb Chuck, and Scott Niekum. Screwnet: Category-independent articulation model estimation from depth images using screw theory. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pages 13670-13677. IEEE, 2021. 2

Ajinkya Jain, Rudolf Lioutikov, Caleb Chuck, 和 Scott Niekum. Screwnet:基于螺旋理论的深度图像类别无关关节模型估计. 2021年IEEE国际机器人与自动化会议(ICRA), 13670-13677页. IEEE, 2021. 2

[15] Zhenyu Jiang, Cheng-Chun Hsu, and Yuke Zhu. Ditto: Building digital twins of articulated objects from interaction.

蒋振宇、许承春和朱玉科。Ditto:从交互中构建关节物体的数字孪生。

In Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 1, 2

在计算机视觉与模式识别会议(CVPR)上，2022年。1, 2

[16] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42 (4), 2023. 2, 1

伯恩哈德·克布、乔治奥斯·科帕纳斯、托马斯·莱姆库勒和乔治·德雷塔基斯。 高斯泼溅用于实时辐射场渲染。ACM图形学汇刊，42(4)，2023年。2, 1

[17] Diederik P Kingma. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014. 3

迪德里克·P·金马。Adam:一种随机优化方法。arXiv预印本 arXiv:1412.6980，2014年。3

[18] Joseph B Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of the American Mathematical society, 7(1):48-50, 1956. 2

约瑟夫·B·克鲁斯卡尔。关于图的最短生成子树和旅行商问题。美国数学学会会刊，7(1):48-50，1956年。2

[19] Robert Kwiatkowski and Hod Lipson. Task-agnostic self-modeling machines. Science Robotics, 4(26):eaau9354, 2019. 1, 2

罗伯特·克维亚特科夫斯基和霍德·利普森。任务无关的自我建模机器。科学机器人学，4(26):eaau9354，2019年。1, 2

[20] Fernando Díaz Ledezma and Sami Haddadin. Machine learning-driven self-discovery of the robot body morphology. Science Robotics, 8(85):eadh0972, 2023. 1, 2

费尔南多·迪亚斯·莱德斯马和萨米·哈达丁。机器学习驱动的机器人身体形态自我发现。科学机器人学，8(85):eadh0972，2023年。1, 2

[21] Xiaolong Li, He Wang, Li Yi, Leonidas J Guibas, A Lynn Abbott, and Shuran Song. Category-level articulated object pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 3706-3715, 2020. 1

李晓龙、王赫、易立、莱昂尼达斯·J·吉巴斯、A·林恩·阿博特和宋书然。类别级关节物体姿态估计。在IEEE/CVF计算机视觉与模式识别会议论文集上，第3706-3715页，2020年。1

[22] Jiayi Liu, Ali Mahdavi-Amiri, and Manolis Savva. Paris: Part-level reconstruction and motion analysis for articulated objects. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 352-363, 2023. 2

刘佳怡、阿里·马赫达维-阿米里和马诺利斯·萨瓦。Paris:关节物体的部件级重建和运动分析。在IEEE/CVF国际计算机视觉会议论文集上，第352-363页，2023年。2

[23] Ruoshi Liu, Alper Canberk, Shuran Song, and Carl Von-drick. Differentiable robot rendering, 2024. 1, 2

刘若石、阿尔珀·坎伯克、宋书然和卡尔·冯德里克。可微分机器人渲染，2024年。1, 2

[24] Shaowei Liu, Saurabh Gupta, and Shenlong Wang. Building rearticulable models for arbitrary objects from point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21138- 21147, 2023. 2, 6, 7, 8, 4, 5

刘绍伟、索拉布·古普塔和王申龙。从 点云构建任意 物体的可重关节模型。在IEEE/CVF计算机视觉与模式识别会议论文集上，第21138-21147页，2023年。2, 6, 7, 8, 4, 5

[25] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH ’87), pages 163-169. ACM, 1987. 5

威廉·E·洛伦森和哈维·E·克莱恩。行进立方体:一种高分辨率3D表面构建算法。在第14届计算机图形与交互技术年会(SIGGRAPH ’87)论文集上，第163-169页。ACM，1987年。5

[26] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In

乔纳森·卢伊滕、乔治奥斯·科帕纳斯、巴斯蒂安·莱布和德瓦·拉马南。动态3D高斯:通过持久动态视图合成进行跟踪。在

[27] J MacQueen. Some methods for classification and analysis of multivariate observations. In Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability/University of California Press, 1967. 4

J·麦克奎恩。一些用于分类和分析多元观测值的方法。在第5届伯克利数学统计与概率研讨会论文集/加州大学出版社，1967年。4

[28] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. Isaac gym: High performance gpu-based physics simulation for robot learning, 2021. 1

维克托·马科维奇丘克、卢卡什·瓦夫日尼亚克、郭云荣、米歇尔·卢、基尔·斯托里、迈尔斯·麦克林、大卫·霍勒、尼基塔·鲁丁、亚瑟·奥尔希尔、安库尔·汉达和加夫里尔·斯泰特。Isaac gym:用于机器人学习的高性能GPU物理模拟，2021年。1

[29] Zhao Mandi, Yijia Weng, Dominik Bauer, and Shuran Song. Real2code: Reconstruct articulated objects via code generation. arXiv preprint arXiv:2406.08474, 2024. 1, 2

赵曼迪、翁一佳、多米尼克·鲍尔和宋书然。Real2code:通过代码生成重建关节物体。arXiv预印本 arXiv:2406.08474，2024年。1, 2

[30] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1):99-106, 2021. 2

本·米尔登霍尔(Ben Mildenhall)、普拉图尔·P·斯里尼瓦桑(Pratul P Srinivasan)、马修·坦西克(Matthew Tancik)、乔纳森·T·巴伦(Jonathan T Barron)、拉维·拉马莫尔西(Ravi Ramamoorthi)和任昂(Ren Ng)。NeRF:将场景表示为神经辐射场以进行视图合成。ACM通讯，65(1):99-106, 2021. 2

[31] Kaichun Mo, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019. 2, 5

莫凯春，朱士林，Angel X. Chang，李毅，Subarna Tripathi，Leonidas J. Guibas，苏浩。PartNet:用于细粒度和层次化部件级3D物体理解的大规模基准。发表于IEEE计算机视觉与模式识别会议(CVPR)，2019年。2, 5

[32] Jiteng Mu, Weichao Qiu, Adam Kortylewski, Alan Yuille, Nuno Vasconcelos, and Xiaolong Wang. A-sdf: Learning disentangled signed distance functions for articulated shape representation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 13001-13011, 2021. 2

穆继腾, 邱伟超, Adam Kortylewski, Alan Yuille, Nuno Vasconcelos, 王晓龙. A-sdf: 学习解耦的符号距离函数用于关节形状表示. 在IEEE/CVF国际计算机视觉会议论文集, 第13001-13011页, 2021. 2

[33] Inc. Niantic. Scaniverse. https://scaniverse.com/, 2024. Mobile app. 1, 2

公司。Niantic。Scaniverse。https://scaniverse.com/, 2024. 移动应用程序。1, 2

[34] Neil Nie, Samir Yitzhak Gadre, Kiana Ehsani, and Shu-ran Song. Structure from action: Learning interactions for articulated object structure discovery. arXiv preprint arXiv:2207.08997, 2022. 2

尼尔·聂(Neil Nie)、萨米尔·伊茨哈克·加德雷(Samir Yitzhak Gadre)、基亚娜·埃萨尼(Kiana Ehsani)和宋淑然(Shu-ran Song)。从动作中学习结构:用于关节物体 结构发现的交互学习。arXiv预印本 arXiv:2207.08997, 2022. 2

[35] Atsuhiro Noguchi, Umar Iqbal, Jonathan Tremblay, Tatsuya Harada, and Orazio Gallo. Watch it move: Unsupervised discovery of joints for re-posing of articulated objects. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3667-3677. IEEE, 2022. 2

野口敦弘(Atsuhiro Noguchi)、乌马尔·伊克巴尔(Umar Iqbal)、乔纳森·特伦布莱(Jonathan Tremblay)、原田达也(Tatsuya Harada)和奥拉齐奥·加洛(Orazio Gallo)。观看其运动:无监督发现的 关节用于铰接物体的重新定位。在2022年IEEE/CVF计算机视觉与模式识别会议(CVPR)上，第3667-3677页。IEEE，2022年。2

[36] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 165-174, 2019. 2

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 165-174, 2019. 2

[37] Mateusz Pawlik and Nikolaus Augsten. Efficient computation of the tree edit distance. ACM Transactions on Database Systems (TODS), 40(1):1-40, 2015. 6

马特乌什·帕夫利克(Mateusz Pawlik)和尼古劳斯·奥格斯滕(Nikolaus Augsten)。树编辑距离的高效计算。ACM数据库系统汇刊(TODS)，40(1):1-40，2015年。6

[38] Mateusz Pawlik and Nikolaus Augsten. Tree edit distance: Robust and memory-efficient. Information Systems, 56:157- 173, 2016. 6

马特乌什·帕夫利克(Mateusz Pawlik)和尼古劳斯·奥格斯滕(Nikolaus Augsten)。树编辑距离:鲁棒且内存高效。《信息系统》(Information Systems)，56卷:157-173页，2016年。6

[39] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10318-10327, 2021. 2

阿尔伯特·普马罗拉(Albert Pumarola)、恩里克·科罗纳(Enric Corona)、杰拉德·庞斯-莫尔(Gerard Pons-Moll)和弗朗西斯科·莫雷诺-诺格尔(Francesc Moreno-Noguer)。D-NeRF:动态场景的神经辐射场(Neural Radiance Fields for Dynamic Scenes)。在IEEE/CVF计算机视觉与模式识别会议(Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition)中，第10318-10327页，2021年。2

[40] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 652-660, 2017.4

查尔斯·R·齐、苏浩、莫凯春和莱昂尼达斯·J·吉巴斯。Pointnet:用于 分类和分割的点集深度学习。在IEEE计算机视觉与模式识别会议论文集，第652-660页，2017年4月。

[41] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30, 2017. 4

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30, 2017. 4

[42] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating deep learning with pytorch3d. arXiv:2007.08501, 2020. 4

Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, 和 Georgia Gkioxari。使用 PyTorch3D 加速深度学习。arXiv:2007.08501, 2020. 4

[43] ROS Wiki. Unified robot description format (urdf), 2023. Accessed: 2024-09-24. 1

ROS Wiki. 统一机器人描述格式(Unified Robot Description Format, URDF), 2023. 访问日期: 2024-09-24. 1

[44] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20:53-65, 1987. 5, 2

彼得·J·鲁索(Peter J Rousseeuw)。轮廓图:一种辅助解释和验证聚类分析的图形工具。计算与应用数学杂志，20:53-65, 1987. 5, 2

[45] Chaoyue Song, Jiacheng Wei, Chuan Sheng Foo, Guosheng Lin, and Fayao Liu. Reacto: Reconstructing articulated objects from a single video. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5384-5395, 2024. 2

宋超越, 魏家成, 符传盛, 林国盛, 刘发耀. Reacto: 从单一视频重建关节物体. 在IEEE/CVF计算机视觉与模式识别会议论文集, 页码5384-5395, 2024. 2

[46] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ra-mamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. NeurIPS, 2020. 3

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, 和 Ren Ng. 傅里叶特征让网络在低维域中学习高频函数。NeurIPS, 2020. 3

[47] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 5026-5033. IEEE, 2012. 1

Emanuel Todorov, Tom Erez, 和 Yuval Tassa. Mujoco: 基于模型控制的物理引擎。在2012年IEEE/RSJ国际智能机器人与系统会议上，页码5026-5033。IEEE, 2012. 1

[48] Xiaogang Wang, Bin Zhou, Yahao Shi, Xiaowu Chen, Qin-ping Zhao, and Kai Xu. Shape2motion: Joint analysis of motion parts and attributes from 3d shapes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8876-8884, 2019. 2

王晓刚, 周斌, 石亚豪, 陈晓武, 赵钦平, 和 徐凯. Shape2motion: 从3D形状中联合分析运动部件和属性。在IEEE/CVF计算机视觉与模式识别会议论文集上，页码8876-8884, 2019. 2

[49] Zhan Xu, Yang Zhou, Evangelos Kalogerakis, Chris Lan-dreth, and Karan Singh. Rignet: Neural rigging for articulated characters. arXiv preprint arXiv:2005.00559, 2020. 2

徐展, 周阳, Evangelos Kalogerakis, Chris Landreth, 和 Karan Singh. Rignet: 用于关节角色的神经绑定。arXiv预印本 arXiv:2005.00559, 2020. 2

[50] Li Yi, Haibin Huang, Difan Liu, Evangelos Kalogerakis, Hao Su, and Leonidas Guibas. Deep part induction from articulated object pairs. ACM Trans. Graph., 37(6), 2018. 1

李毅, 黄海斌, 刘迪凡, Evangelos Kalogerakis, 苏浩, 和 Leonidas Guibas. 从关节对象对中深度诱导部件。ACM图形学汇刊, 37(6), 2018. 1

[51] Harry Zhang, Ben Eisner, and David Held. Flowbot++: Learning generalized articulated objects manipulation via articulation projection. In Conference on Robot Learning, pages 1222-1241. PMLR, 2023. 2

张哈里, Ben Eisner, 和 David Held. Flowbot++: 通过学习关节投影实现广义关节对象操作。在机器人学习会议上，页码1222-1241。PMLR, 2023. 2

[52] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5745-5753, 2019. 3

周毅, Connelly Barnes, 卢静婉, 杨继梅, 和 李浩. 关于神经网络中旋转表示的连续性。在IEEE/CVF计算机视觉与模式识别会议论文集上，页码5745-5753, 2019. 3

# AutoURDF: Unsupervised Robot Modeling from Point Cloud Frames Using Cluster Registration

# AutoURDF: 使用聚类配准从点云帧中进行无监督机器人建模

Supplementary Material

补充材料

Overview. The supplementary material is structured into the following sections:

概述。补充材料分为以下几个部分:

* Data Collection A: This section describes the synthetic data generation process and real-world datasets used for evaluating the proposed method;
* 数据收集 A: 本节描述了用于评估所提出方法的合成数据生成过程和真实世界数据集；
* Method Details B: This section provides a detailed explanation of the topology inference algorithm, model architecture, and comparison between different pose representations;
* 方法细节 B: 本节详细解释了拓扑推断算法、模型架构以及不同姿态表示之间的比较；
* Additional Experiments C: This section presents experiments analyzing the impact of varying frames, cameras, clusters, and parameters, along with supplementary visualization results.
* 额外实验 C: 本节展示了分析不同帧、相机、聚类和参数影响的实验，以及补充的可视化结果。

# A. Data Collection

# A. 数据收集

# A.1. Synthetic Data

# A.1. 合成数据

We simulate the data collection process with Pybullet by positioning several cameras around the robot to capture depth images. Each camera’s intrinsic parameters are defined within the simulator, and depth images are rendered at a resolution of from their respective viewpoints. Point cloud data is created by merging depth images captured from various viewpoints. The robot is controlled by randomly sampled motor angle sequences, and the corresponding point cloud frames are captured during this process. In the simulation, we collect 5 video sequences per robot, with each video containing 10 frames of point cloud data. To simulate real-world conditions, random positional noise and per-point noise are added to the point cloud data. In the experimental results presented in the main text, we combine 20 depth map views into a single frame point cloud. Figure 12 illustrates the process of generating a single-frame point cloud from three views of depth images.

我们通过Pybullet模拟数据收集过程，将多个相机放置在机器人周围以捕捉深度图像。每个相机的内参在模拟器中定义，深度图像以 的分辨率从各自的视角渲染。通过合并从不同视角捕捉的深度图像生成点云数据。机器人由随机采样的电机角度序列控制，并在此过程中捕捉相应的点云帧。在模拟中，我们为每个机器人收集5个视频序列，每个视频包含10帧点云数据。为了模拟真实世界条件，向点云数据添加随机位置噪声和每点噪声。在正文的实验结果中，我们将20个深度图视图合并为单帧点云。图12展示了从三个深度图像视图生成单帧点云的过程。

To ensure that the point cloud sequence captures sufficient motion information, we independently sample random targets within the motor angle limits for each motor. Additionally, if the robot detects a self-collision, the sequence is restarted with a new set of target motor angles. Randomly sampled data may include similar rotational motions, making it challenging to identify distinct parts. Our method can merge multiple random motion sequences, improving the reposed shape prediction and joint estimation accuracy.

为了确保点云序列捕捉到足够的运动信息，我们在每个电机的运动角度限制内独立采样随机目标。此外，如果机器人检测到自碰撞，序列将以一组新的目标电机角度重新启动。随机采样的数据可能包含类似的旋转运动，这使得识别不同部件变得具有挑战性。我们的方法可以合并多个随机运动序列，从而提高静止形状预测和关节估计的准确性。

# A.2. Real-world Data

# A.2. 真实世界数据

As shown in Figure 13, we conducted real-world experiments using 10 consecutive point cloud scans of a WX200 robot arm, controlling all five motors. The robot was operated through a ROS2 interface, following a randomly sampled motor angle sequence.

如图13所示，我们使用WX200机器人手臂的10个连续点云扫描进行了真实世界实验，控制所有五个电机。机器人通过ROS2接口操作，遵循随机采样的电机角度序列。

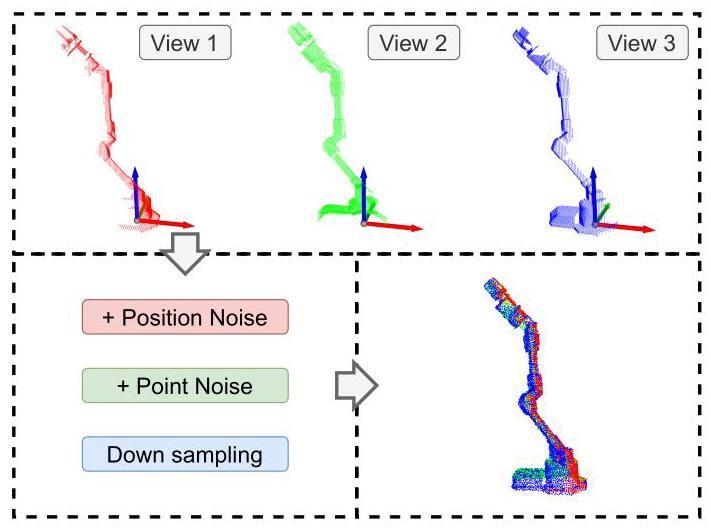


Figure 12. Synthetic data collection. One frame of synthetic data is collected by merging multi-view depth maps into a single-point cloud. Global coordinate noise and per-point noise are applied to simulate realistic conditions. This image shows an example of a point cloud created from three depth images.

图12. 合成数据收集。通过将多视角深度图合并为单点云来收集一帧合成数据。应用全局坐标噪声和每点噪声以模拟真实条件。此图像显示了由三个深度图像创建的点云示例。

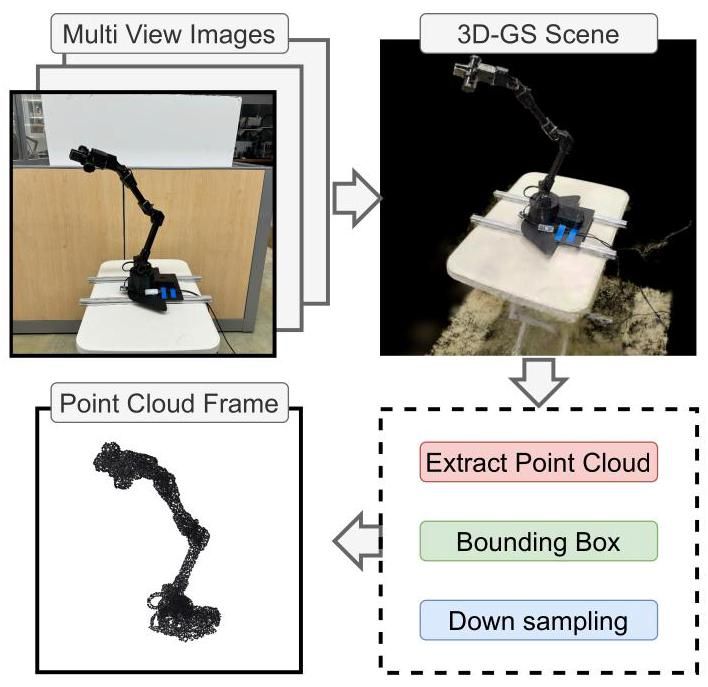


Figure 13. Real-world data collection. For each motion step, we capture a video of the robot arm, reconstruct the 3D scene using a 3D Gaussian splatting application [16, 33], and extract the point cloud. A bounding box removes background points, and down-sampling standardizes the number of points.

图13. 真实世界数据收集。对于每个运动步骤，我们捕获机器人手臂的视频，使用3D高斯泼溅应用程序[16, 33]重建3D场景，并提取点云。边界框去除背景点，下采样标准化点数。

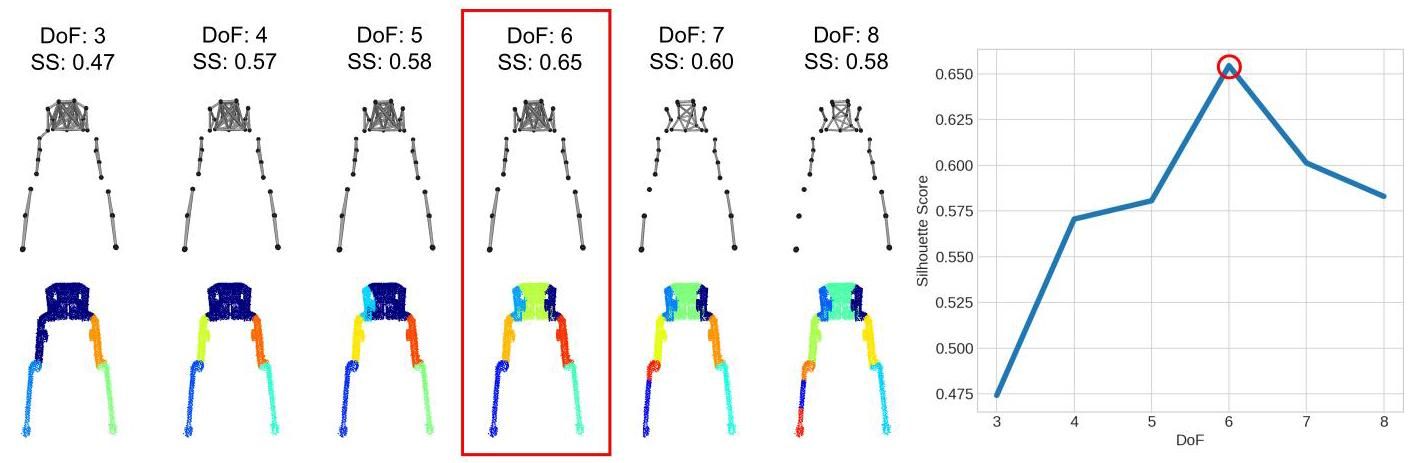


Figure 14. Silhouette Score Method. An example of using the Silhouette Score method to identify distinct moving parts and predict the degrees of freedom(DoF)for the Bolt bipedal robot, with a peak score at . This indicates that segmenting the point clusters into seven parts provides the optimal grouping.

图14. 轮廓评分方法。使用轮廓评分方法识别不同运动部件并预测Bolt双足机器人自由度(DoF)的示例，峰值分数为 。这表明将点簇分割为七个部分提供了最佳分组。

Point cloud data was collected using an iPhone camera and the iOS scanning application Scanverse [33]. A bounding box was applied to isolate the robot arm’s point cloud from the environment. Despite the real scan data containing misaligned coordinates across time-steps and significant surface noise, our method directly processes the raw point cloud data, achieving accurate segmentation and URDF generation.

使用iPhone相机和iOS扫描应用程序Scanverse [33]收集点云数据。应用边界框将机器人手臂的点云与环境隔离。尽管真实扫描数据包含时间步长中未对齐的坐标和显著的表面噪声，我们的方法直接处理原始点云数据，实现了准确的分割和URDF生成。

# B. Method Details

# B. 方法细节

# B.1. Silhouette Score Method for Part Segmentation

# B.1. 用于部件分割的轮廓评分方法

Figure 14 illustrates an example of using the Silhouette Score [44] method to identify the number of links. With DoF ranging from 3 to 8, the averaged Silhouette Score peaks at , which is the correct prediction for the Bolt robot. Given the number of groups for the segmentation algorithm, the Silhouette Score and Coefficient are calculated as equation 4 and equation 5 .

图14展示了使用轮廓评分[44]方法识别链接数量的示例。随着自由度(DoF)从3到8变化，平均轮廓评分在 处达到峰值，这是对Bolt机器人的正确预测。给定分割算法的组数，轮廓评分和系数按公式4和公式5计算。

In Equation 4, denotes the Silhouette Score of the - th node, which, in our case, is a point cluster. The term represents the average distance of the -th node to all other nodes within the same group, while represents the average distance to nodes in the nearest group. The distance is calculated using Equation 3. is the Silhouette Coefficient, which depends on the number of groups, represents the total number of nodes. To determine the optimal number of groups for segmentation, we maximize the average Silhouette Score over all groups.

在公式4中， 表示第 个节点的轮廓评分，在我们的案例中，它是一个点簇。术语 表示第 个节点到同一组内所有其他节点的平均距离，而 表示到最近组节点的平均距离。距离使用公式3计算。 是轮廓系数，它取决于组数， 表示节点总数。为了确定分割的最佳组数，我们最大化所有组的平均轮廓评分。

# B.2. Topology Inference

# B.2. 拓扑推断

As shown in Algorithm 2, topology inference is divided into three main stages to construct the body topology graph . In the first stage, the algorithm identifies connected components from the segmentation graph , grouping clusters into link components . For each component, it determines its neighboring components using the Minimum Spanning Tree , identifying clusters that are directly connected, and therefore the corresponding groups of direct connection. A dictionary is created for each link, containing a unique identifier(Id), a parent link initially set to None, and a list of connected link IDs. These dictionaries are stored in the list Links for further processing. In the second stage, the kinematic tree structure is derived by iteratively traversing the links. Starting with the root link, which is chosen as the first link sorted by ascending center movements, child links are identified by excluding their parent link from the list of connected links. The algorithm updates the parent-child relationships and adds the child links to the next layer for processing. This process continues layer by layer until all links are processed. In the final stage, the body topology graph is constructed, where is the set of link IDs, and consists of directed edges representing parent-child relationships.

如算法2所示，拓扑推断分为三个阶段以构建身体拓扑图 。在第一阶段，算法从分割图 中识别连接组件，将簇分组为链接组件 。对于每个组件，它使用最小生成树 确定其相邻组件，识别直接连接的簇，从而确定直接连接的相应组。为每个链接创建一个字典，包含唯一标识符(Id)、初始设置为None的父链接以及连接的链接ID列表。这些字典存储在列表Links中以供进一步处理。在第二阶段，通过迭代遍历链接导出运动树结构。从根链接开始，根链接选择为按中心运动升序排序的第一个链接，通过从连接链接列表中排除其父链接来识别子链接。算法更新父子关系并将子链接添加到下一层进行处理。此过程逐层继续，直到处理完所有链接。在最后阶段，构建身体拓扑图 ，其中 是链接ID集合， 由表示父子关系的有向边组成。

# B.3. Model Architecture

# B.3. 模型架构

Figure 15 shows the architecture of the registration model. We use a shared neural network to predict incremental updates to the position and rotation of each point cluster. The input dimensions align with the pose representation of the point clusters. A sinusoidal positional encoder enhances

图15展示了配准模型的架构。我们使用共享神经网络来预测每个点簇位置和旋转的增量更新。输入维度与点簇的姿态表示对齐。正弦位置编码器增强了

Algorithm 2 Topology Inference

算法2 拓扑推断

Input: Segmentation , MST

Output: Body Topology

Initialize: Links EmptyList

Initialize: connected\_components

where is a list of cluster indices

// 1. Construct a list of link dictionaries

for in enumerate

Find connected clusters

for in

.neighbors

end for

Find connected links with

Build dictionary: Link

Id: ;

parent: None;

connected\_links:

Links.append(Link)

end for

Sort links by center movements in ascending order

// 2. Derive the kinematic tree

Initialize: current\_layer

repeat

Initialize next\_layer , child\_set

for Link current in current\_layer do

if has parent then

child connected\_links excluding parent

else

child connected\_links

end if

for Link in child do

Update Link child.parent Link current.Id

Add Link to next\_layer

end for

Update child\_set child\_set child

end for

Update current\_layer next\_layer

until child\_set

// 3. Build from Links

where Link.Id (Link.parent, Link.Id)

the network’s ability to capture spatial relationships and patterns[46]. The model includes a fully connected encoder and separate decoders for rotation and position. The network’s output is added to the input coordinates to learn incremental updates directly. Leaky ReLU is used as the activation function to improve learning of rotations, and optimization is performed using the Adam optimizer[17].

网络捕捉空间关系和模式的能力[46]。该模型包括一个全连接编码器和分别用于旋转和位置的解码器。网络的输出被添加到输入坐标中以直接学习增量更新。Leaky ReLU被用作激活函数以改善 旋转的学习，并使用Adam优化器[17]进行优化。

The same model architecture is applied to both the Step Model and the Anchor Model, with learning rates of 0.0001 and 0.00005, respectively. The training loss is calculated using the L1 Chamfer distance between the transformed point cloud and the ground truth. An early stopping mechanism halts optimization if the point cloud error does not decrease after a set number of steps, ensuring efficient training.

相同的模型架构被应用于Step模型和Anchor模型，学习率分别为0.0001和0.00005。训练损失通过变换后的点云与真实值之间的L1 Chamfer距离计算。如果点云误差在一定步数后没有减少，早期停止机制会停止优化，确保训练效率。

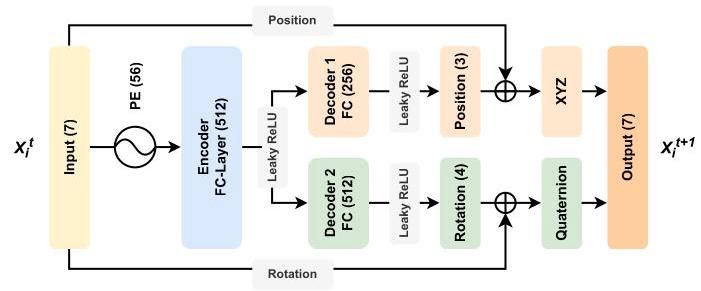


Figure 15. Registration Model Architecture. We developed a lightweight neural network for point cluster registration, employing the same model architecture for both the Step Model and Anchor Model.

图15. 配准模型架构。我们开发了一个轻量级神经网络用于点簇配准，Step模型和Anchor模型均采用相同的模型架构。

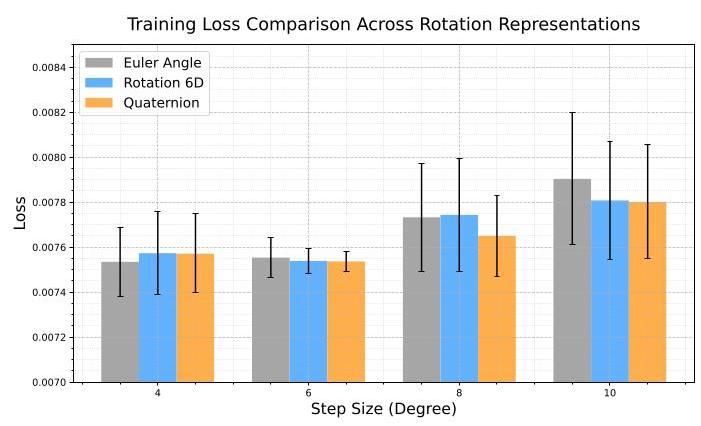


Figure 16. Training Loss Comparison of Rotation Representations Across Step Sizes.

图16. 不同步长下旋转表示的训练损失对比。

# B.4. Rotation Representation

# B.4. 旋转表示

To efficiently and robustly learn the rotation of clusters To optimize the learning of inter-frame cluster rotations in our registration model, we investigate the efficacy of three rotation representations: Euler Angles, Quaternions, and 6D Rotation representations [52]. We conduct extensive experiments across ten diverse sequences, comparing these representations regarding their training stability and convergence properties. As shown in 16, both Quaternions and 6D Rotation representations demonstrate superior robustness as the angular step size increases from to . The empirical results suggest that these continuous representations maintain consistent performance even under larger rotational variations, while Euler Angles show increased instability at higher angles. This aligns with previous findings regarding the advantages of continuous rotation representations in deep learning frameworks. Based on these results, our implementation supports both Quaternion and 6D Rotation representations, with Quaternion as the default configuration.

为了高效且稳健地学习簇的旋转，以优化我们配准模型中帧间簇旋转的学习，我们研究了三种旋转表示的效果:欧拉角、四元数和6D旋转表示[52]。我们在十个不同的序列上进行了广泛的实验，比较了这些表示在训练稳定性和收敛性方面的表现。如图16所示，随着角步长从 增加到 ，四元数和6D旋转表示都表现出更强的鲁棒性。实验结果表明，这些连续表示即使在较大的旋转变化下也能保持一致的性能，而欧拉角在较大角度下表现出更高的不稳定性。这与之前关于深度学习框架中连续旋转表示优势的研究结果一致。基于这些结果，我们的实现支持四元数和6D旋转表示，并以四元数为默认配置。

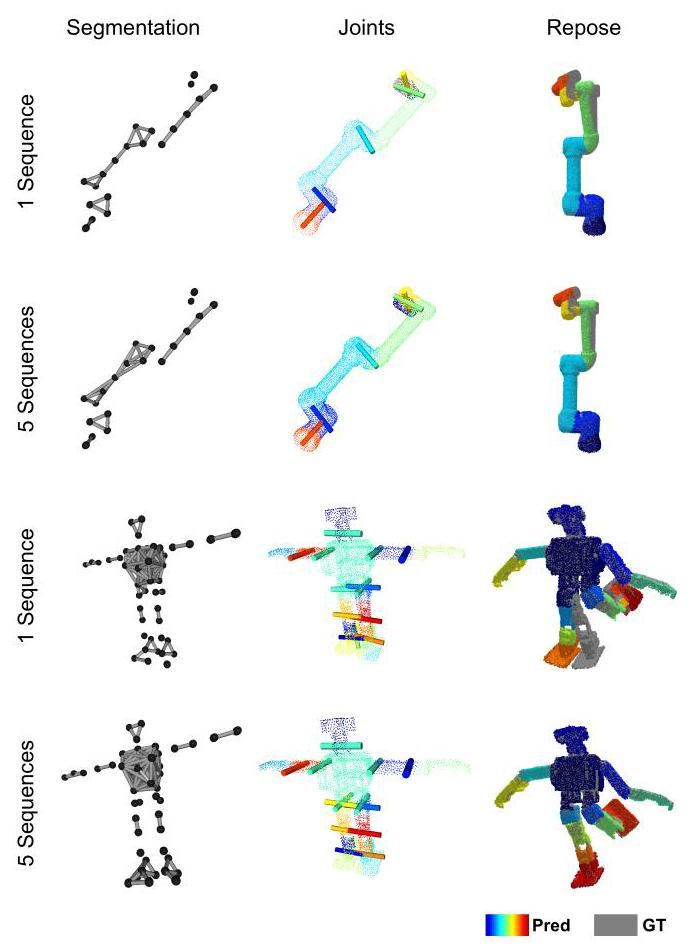


Figure 17. Qualitative Comparison on Different Number of Input Sequences.

图17. 不同输入序列数量下的定性对比。

# C. Additional Results

# C. 附加结果

# C.1. Experiment on Number of Input Sequences

# C.1. 输入序列数量实验

The Multi-Sequence Merging Experiment presented in our main text compares the performance of our method using a single sequence of point cloud frames against five sequences of point cloud frames. The results indicate that Our method achieves improved performance in 7 out of 8 robots for both repose evaluation and joint distance evaluation while demonstrating improvements across all robots for joint angle evaluation. Additionally, Figure 20 provides a qualitative comparison. With data from five sequences, our method generates segmentations with higher distinction, creating more edges within the correct group of point clusters, as exemplified in the UR5 robot. Furthermore, it achieves higher accuracy in joint estimation and reposed point cloud generation.

我们在主文本中提出的多序列合并实验比较了使用单一点云帧序列和五点云帧序列时我们方法的性能。结果表明，在8个机器人中的7个中，我们的方法在姿态评估和关节距离评估中均表现出改进，同时在所有机器人的关节角度评估中也表现出改进。此外，图20提供了定性对比。使用五个序列的数据，我们的方法生成了更具区分性的分割，在正确的点簇组内创建了更多的边缘，如UR5机器人所示。此外，它在关节估计和重新生成点云方面也达到了更高的准确性。

With the starting motor configurations aligned across different sequences, our method merges multiple sequences by registering them to the same set of point clusters and averaging the resulting motion correlation matrices. We perform repose comparison by repeating the synthetic point cloud collection process (Figure 12) using a new set of random motor configurations applied to both the predicted and ground-truth URDFs.

通过在不同序列中对齐起始电机配置，我们的方法通过将它们配准到同一组点簇并平均生成的运动相关矩阵来合并多个序列。我们通过使用一组新的随机电机配置应用于预测和真实URDF，重复合成点云收集过程(图12)来进行姿态比较。

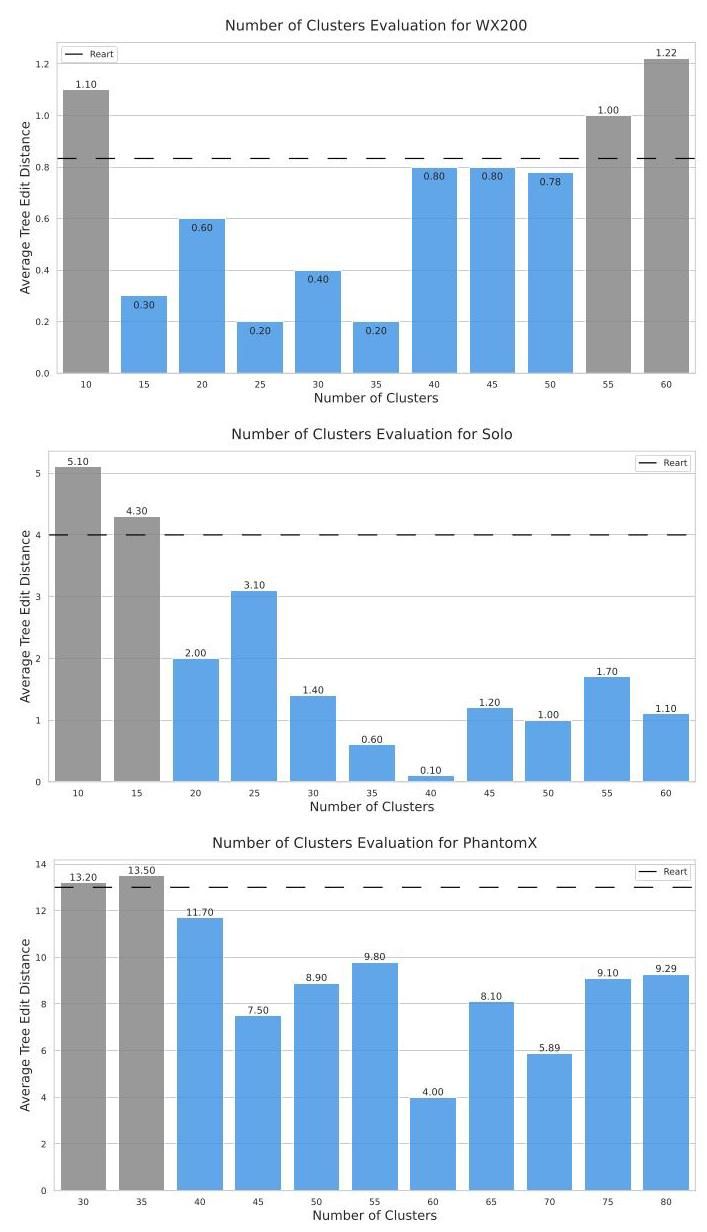


Figure 18. Impact of Cluster Number on Tree Edit Distance. Evaluation across three robot configurations (WX200, Solo, Phan-tomX) with varying structural complexity. Blue bars indicate performance superior to Reart [24], shown as the dashed line.

图18. 簇数量对树编辑距离的影响。在三种具有不同结构复杂性的机器人配置(WX200、Solo、PhantomX)上进行评估。蓝色条表示性能优于Reart[24]，如虚线所示。

# C.2. Experiment on Number of Clusters

# C.2. 簇数量实验

To investigate the sensitivity of our method to the number of clusters hyperparameter, we conduct extensive experiments across robots with varying structural complexity. We evaluate each robot configuration across eleven different cluster quantities, comparing against the Reart baseline [24]. As shown in 17, we observe that while optimal performance occurs at specific cluster number ranges, our method consistently outperforms the baseline (indicated by blue bars) across a wide range of parameter settings, demonstrating the method’s stability across different robot architectures.

为了研究我们的方法对簇数量超参数的敏感性，我们在具有不同结构复杂性的机器人上进行了广泛的实验。我们在十一种不同的簇数量下评估每种机器人配置，并与Reart基线[24]进行比较。如图17所示，我们观察到，虽然在特定簇数量范围内性能最佳，但我们的方法在广泛的参数设置范围内始终优于基线(由蓝色条表示)，展示了该方法在不同机器人架构下的稳定性。

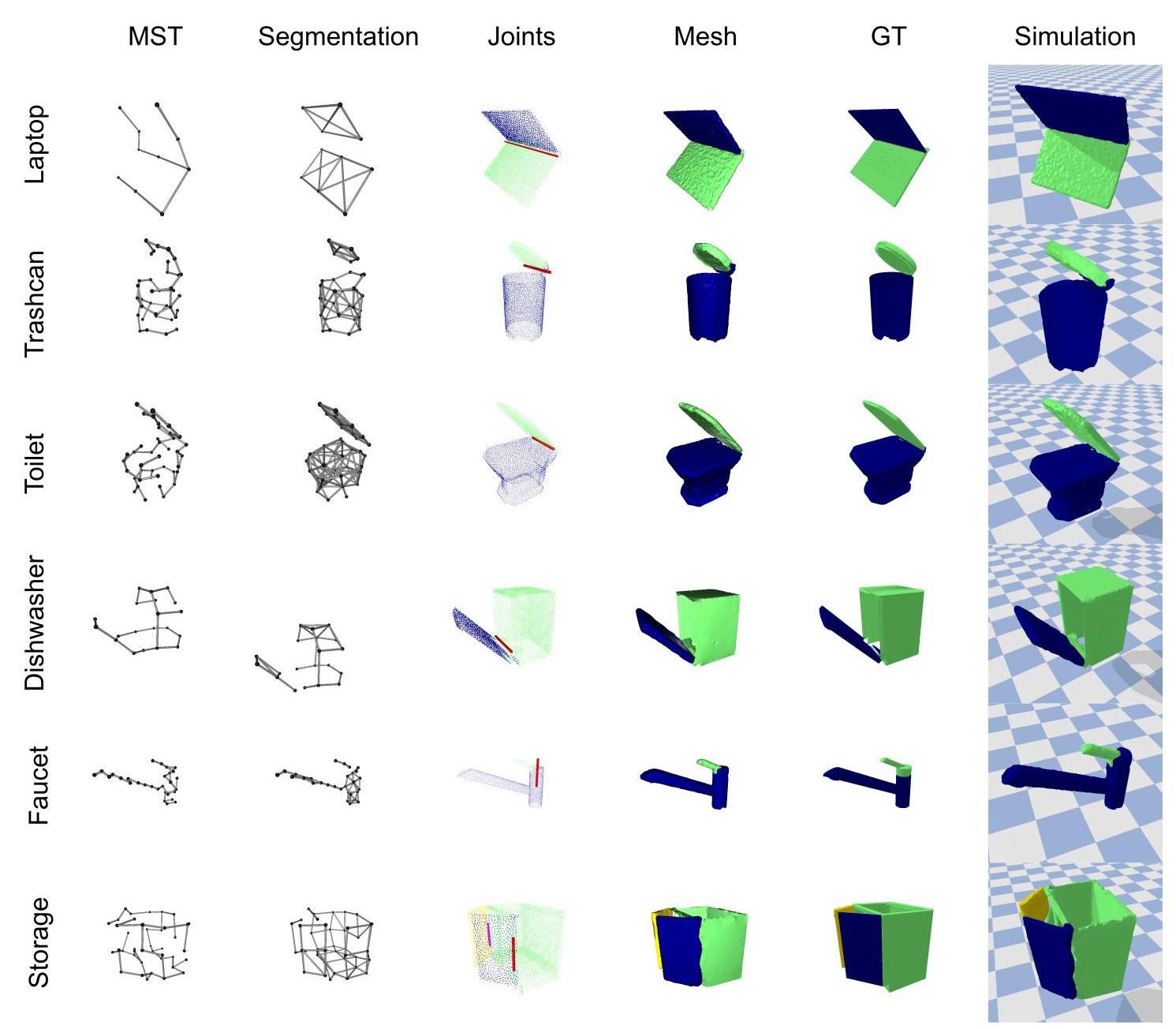


Figure 19. Qualitative Results on Sapien[31] Dataset.

图19. Sapien[31]数据集上的定性结果。

# C.3. Experiment on Sapien Dataset

# C.3. Sapien数据集实验

We evaluate our AutoURDF framework on six common household articulated objects from the Sapien dataset [31], each featuring one or two degrees of freedom: laptop, trash-can, toilet, dishwasher, faucet, and storage cabinet. As shown in 19, we initialize each object in an open configuration to facilitate distinct part clustering. To enhance the initial segmentation of planar components, we incorporate -means clustering with normal information from the point cloud. The framework demonstrates robust performance in both segmentation and joint parameter estimation for objects with predominantly planar structures. While the cylindrical geometry of the faucet presents challenges for precise segmentation, the framework still maintains accurate joint axis prediction, highlighting its robustness to partial segmentation errors.

我们在Sapien数据集[31]中的六个常见家用铰接物体上评估了我们的AutoURDF框架，每个物体具有一个或两个自由度:笔记本电脑、垃圾桶、马桶、洗碗机、水龙头和储物柜。如图19所示，我们将每个物体初始化为开放配置，以便于进行明显的部件聚类。为了增强平面组件的初始分割，我们结合了 -均值聚类和点云中的法线信息。该框架在具有主要平面结构的物体的分割和关节参数估计中表现出稳健的性能。虽然水龙头的圆柱几何形状对精确分割提出了挑战，但该框架仍然保持了准确的关节轴预测，突出了其对部分分割错误的鲁棒性。

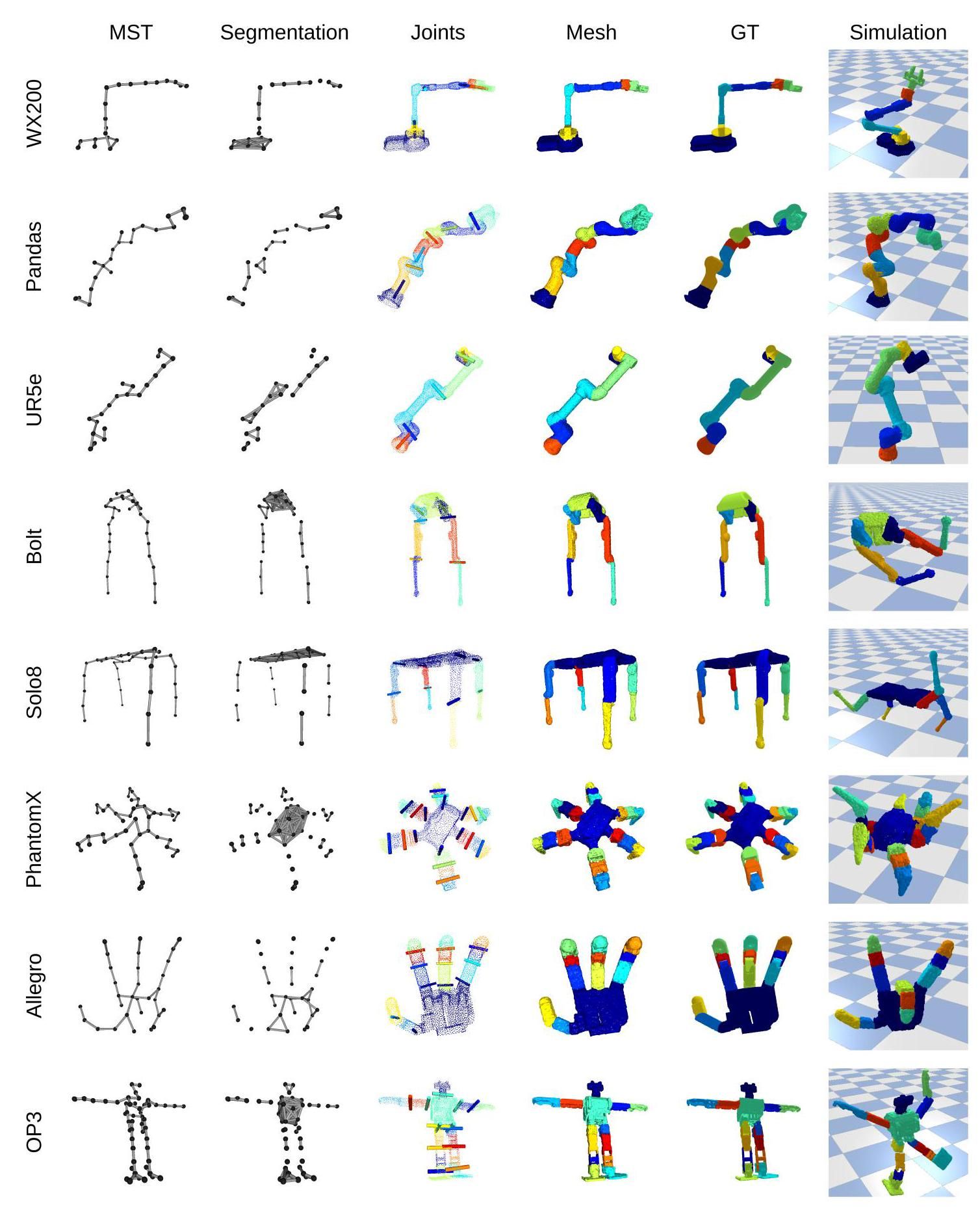


Figure 20. Quantative Results on AutoURDF Dataset.

图20. AutoURDF数据集上的定量结果。