IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Prepare the Chair for the Bear! Robot Imagination of Sitting Affordance to Reorient Previously Unseen Chairs

Xin Meng[1], Hongtao Wu[2], Sipu Ruan[1] and Gregory S. Chirikjian[1]

***Abstract*—In this letter, a paradigm for the classification and manipulation of novel objects is established and demonstrated with the example of chairs. Our approach leverages the robot's understanding of object stability, perceptibility, and affordance to prepare previously unseen and randomly oriented chairs on which a teddy bear is to be seated. The teddy bear is a proxy for an elderly person, hospital patient, or child. By autonomously reconstructing a complete model of the object and inserting it into a physical simulator (*i.e.*, the robot's "imagination"), the robot assesses whether or not the object is a chair and, if it is, determines how to reorient it properly to be used. Experimental results show that our method achieves a high success rate on the real robot task of chair preparation. Also, it outperforms several baseline methods on the task of upright pose prediction for chairs. The same methodology can be easily transferred to a wide variety of application scenarios, and illustrates a broader paradigm in affordance-based reasoning.**

***Index Terms*—AI-Enabled Robotics, Simulation and Animation, Manipulation Planning**

I. INTRODUCTION

As robots begin to enter human life, we expect them to interact with unknown objects intelligently to help humans with daily household tasks. This requires robots to understand: 1) *what* the potential functionality an object possesses, 2) *where* the object can afford such functionality, and 3) *how* the functionality can be afforded. The concept of object affordance [1] describes how a human might interact with an object in a particular environment so as to achieve a goal. It encompasses the variation in functionalities [2] an object can have in different scenarios. For example, a chair possesses sitting functionality, and it can only afford such functionality when it is placed upright in an open space. That is, when it is flipped over or blocked by obstacles, humans cannot sit on it, and thus the chair cannot afford the sitting functionality.

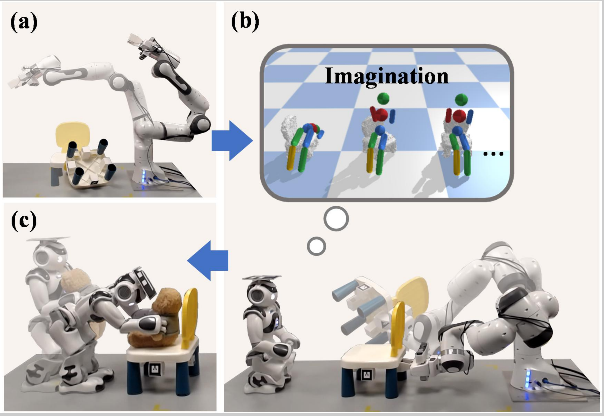

Fig. 1. Overview. (a) The robot places the object into different poses to reveal the occluded part for reconstruction. (b) The robot imagines the sitting affordance of the object to determine if the object is a chair, the functional pose of the object, and how a humanoid figure can be seated. Then the robot rotates the object to the imagined pose for sitting. (c) The robot seats a teddy bear on the object according to the imagination. Video demo and more details are available at https://chirikjianlab.github.io/preparechair/.

To reason about object affordances for robot-object interaction, we define objects from a robot-centric perspective via their interaction-based definition (IBD) [3]. For chairs, the IBD is given by: "an object which can be stably placed on a flat horizontal surface in such a way that a typical human is able to *sit*[1] on it stably above the ground. " This defines an object based on how it can be used. It helps the robot to classify the object more intelligently based on *what* its potential functionality is. We define the pose that enables the object to afford the functionality as the *functional pose* (*i.e.*, upright pose for chairs), which answers the *where* question. When an agent sits on a chair, the body configuration and pose show *how* the chair can be used for sitting. We define the body configuration and pose as the *sitting configuration* and *sitting pose* of the chair.

In this paper, we enable a robot to automatically understand the what, where, and how problems and propose a novel method for the robot to prepare a previously unseen real chair for a teddy bear agent to sit on, regardless of the initial pose of

---

Manuscript received 3 April 2023; accepted 5 August 2023. Date of publication 18 August 2023; date of current version 1 September 2023. This letter was recommended for publication by Associate Editor M. R. Dogar and Editor H. Liu upon evaluation of the reviewers' comments. This work was supported in part by NUS Startup under Grants A-0009059-02-00 and A-0009059-03-00, in part by CDE Board Account under Grant E-465-00-0009-01, and in part by the National Research Foundation, Singapore, under its Medium Sized Centre Programme - Centre for Advanced Robotics Technology Innovation (CARTIN), Sub Award A-0009428-08-00. The work of Hongtao Wu was supported by Gregory S. Chirikjian Internal Discretionary funds at Johns Hopkins University during his Ph.D. studies. *(Corresponding author: Gregory S. Chirikjian.)*

[1] Xin Meng, Sipu Ruan, and Gregory S. Chirikjian are with the Department of Mechanical Engineering, National University of Singapore, Singapore `{mengxin, ruansp, mpegre}@nus.edu.sg`

[2] Hongtao Wu is with the Laboratory for Computational Sensing and Robotics (LCSR), Johns Hopkins University, Baltimore, MD 21218, USA `hwu67@jhu.edu`

Digital Object Identifier (DOI): see top of this page.

[1] To adopt or rest in a posture in which the body is supported on the buttocks and thighs and the torso is more or less upright. https://www.collinsdictionary.com/dictionary/english/sit

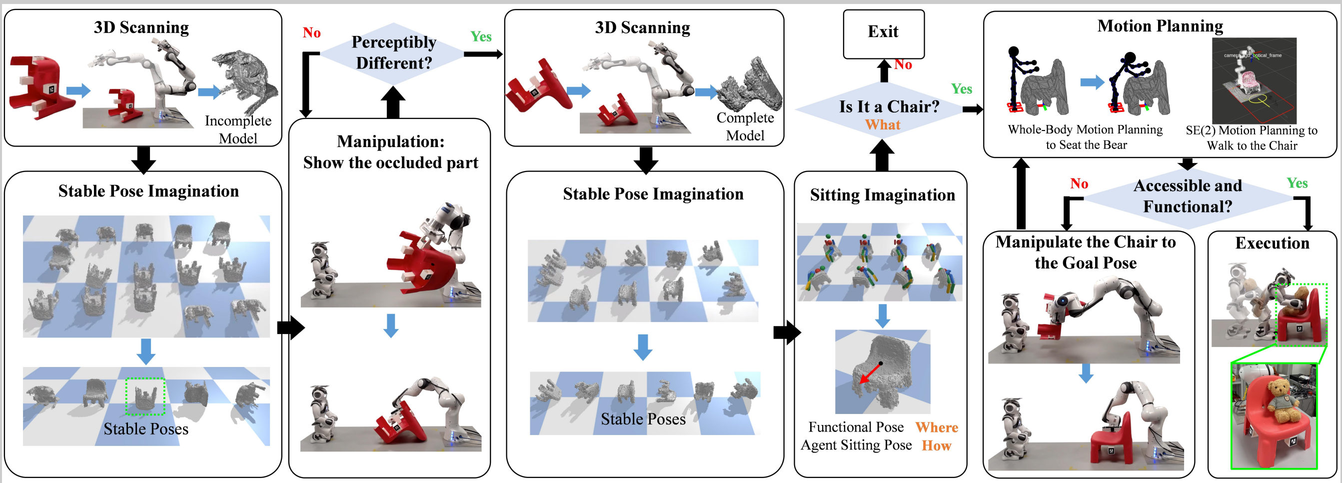

Fig. 2. Pipeline. The robot arm first scans the object and performs stable pose imagination on the reconstructed model. It rotates the object to other stable poses until the object shows the occluded part. It scans the object again and generates the complete model. The robot performs stable pose imagination and sitting imagination on the complete model to determine whether the object is a chair and find a functional pose and sitting pose. The motion of the humanoid robot walking to the chair and seating the bear is planned. The robot uses the planning information to determine if the chair is accessible and/or find an accessible chair pose. The robot arm manipulates the chair to a functional and accessible pose. The humanoid then seats the bear on the chair according to the imagined sitting pose.

the chair. In our method, as the robot has no prior knowledge of the object, it first reconstructs the object and reasons its sitting affordance in physics simulations. A big challenge is that part of the object is always occluded when it is placed on a plane. To fully reconstruct the object, the robot needs to rotate the object to a *stable* pose which shows the occluded part, and fuse the images captured from different viewpoints accurately. If the object is identified as a chair, it can be non-upright or inaccessible[2], and thus not sittable. In this case, the robot needs to further rotate the object to a functional and accessible pose that is sittable for the bear if necessary. In particular, we extend the scope of [3] and [4] by leveraging the imagination method to guide the manipulation for object reconstruction and the manipulation for preparing the chair to a functional and accessible pose. Fig. 2 shows the pipeline of our method.

Our method successfully classifies 12 previously unseen objects with diverse shapes and appearances in 45 trials. It achieves a 100% success rate in preparing the chairs initially placed in non-functional or inaccessible poses for sitting. The humanoid robot seats the bear on the chairs according to the imagined sitting poses, with a success rate of 96.7%. Comparing it with baselines on functional pose prediction, we empathize effectiveness of our imagination method. The contributions of this paper mainly include:

- a mathematical formulation of functionally equivalent and perceptibly equivalent poses.
- an automatic object reconstruction method guided by stability and perceptibility.
- a real robot manipulation system for preparing an unseen chair for sitting.

II. RELATED WORKS

Object Affordance Reasoning. Affordance detection is attracting growing interest in both the fields of computer vision [5], [6] and robotics [7]–[9]. It helps robots with grasping [10] and tool using [8], [11]. Learning-based methods are popular for classifying objects [7] and predicting affordance keypoints [11]. But they require a large dataset for training. As an alternative to learning, our method explores object affordances by simulating physical interactions. It is able to provide additional information including object potential affordance in different poses, object-agent interaction information, and integration of physics in affordance reasoning. This exploration of potential functionalities in simulation is what we refer to as *robot imagination*.

Physics Reasoning. Physical reasoning facilitates robot manipulation in a wide range of tasks including pouring [9], [12], bottle opening [13], cutting [14], and *etc*. Zhu *et al.* [15] learn physical concepts from an RGB-D video and pick the best tool for the task. The problem of containability has been studied by physically simulating putting objects or particles into the object [9], [12]. This idea is similar to digital twins [16] that is widely used in industry to predict system performance. Our method goes beyond predicting outcomes and reasoning about interactions. It explores object affordances by simulating objects interacting with humans and leverages the understanding on robot manipulation.

Sitting Affordance Reasoning. Sitting is one of the most common postures of humans. Hinkle and Olson [17] drop spheres onto objects in simulation and classify the objects into chairs, tables, and containers according to the sphere's final configuration. Grabner *et al.* [18] detect sitting affordance by fitting a humanoid mesh into the scene and evaluating the distance between the object and the humanoid. Besides classification, we explore a deeper understanding of sitting including how an object is able to afford sitting and how an

---

[2] A chair can be not sittable even though it is in its functional (upright) pose. For example, it can be blocked by obstacles (*e.g.*, walls) and thus inaccessible for sitting. By accessible, we mean that the chair can be directly sat on without any changes to its pose.

agent could sit on it.

Object Pose Estimation. Pose estimation has been extensively studied with learning-based methods [19], [20] and traditional methods [21]. However, these methods mainly focus on estimating the pose of a known object given its 3D model. To handle novel objects, Aubry *et al.* [22] and Kadam *et al.* [23] propose to align the 2D image and 3D point cloud with the models in a dataset, respectively. But they mainly focus on objects in upright poses. Li *et al.* [24] performs pose estimation of unknown objects from 3D point clouds based on a generalized reference frame. Our method is different in that it explores the upright pose of an unknown object via reasoning the object functionality. It tackles the challenge of estimating the upright pose of an unknown object in an unknown arbitrary pose.

Comparison with Previous Publications. One prior work [3] proposes a method to identify the sitting affordance of an object and find the functional pose. No real robot experiments are performed. In [4], the focus lies on finding the agent sitting pose for an object known as a chair in an upright pose *a priori*. Thus, no robot manipulation of the chair is included. In this paper, we focus on a completely different task that involves a real robot preparing an unknown object in an unknown *arbitrary* initial pose for sitting. That is, the robot has no prior knowledge of the object model, object category, or upright pose information. The absence of prior knowledge brings great challenges to affordance reasoning, especially when the object is a flipped-over chair with its seat being occluded. To this end, we propose a novel manipulation-assisted 3D reconstruction method that allows the robot to acquire the complete object model for affordance reasoning. Overall, we present a comprehensive robot manipulation system that is able to autonomously explore the sitting affordance of an unknown object in an unknown arbitrary pose and prepare it for sitting via robot manipulation.

III. PROBLEM FORMULATION

The pose of a rigid body can be described as $g = (R, \mathbf{p}) \in SE(3)$, where $R \in SO(3)$ is a rotation matrix that can be parameterized with zyx Euler angles $R = R_{ZYX}(\alpha, \beta, \gamma) = R_Z(\alpha) R_Y(\beta) R_X(\gamma)$. α, β, and γ correspond to the yaw, pitch, and roll, respectively. $R_X(\cdot)$, $R_Y(\cdot)$, and $R_Z(\cdot)$ are the rotation matrices representing rotations about the x-, y-, and z-axis of the world frame, respectively. $\mathbf{p} = [x, y, z]^T \in \mathbb{R}^3$ represents the position.

We define an object as a chair if there exists any pose $g \in SE(3)$ that enables the object to afford the functionality of sitting. When such a pose exists, it is a *functional pose* of the object. Given an unseen object in an arbitrary pose, our goal is threefold. **1) Reconstruct a complete model of the object.** When an object is placed on a plane, part of it is always occluded by the plane. To scan the occluded part, we formulate the problem as rotating the chair into a *stable* pose that shows the occluded part. **2) Reason the sitting affordance from the reconstructed object model.** This includes a) classifying whether an object is a chair (*what*), b) finding the functional pose in which it can afford sitting (*where*), and c) finding the sitting pose for the agent (*how*). **3) Prepare the chair for the bear.** We use a teddy bear as a real agent to showcase the understanding of sitting. Thus, the problem becomes manipulating the chair into a functional and accessible pose so that the teddy bear can be automatically seated by a humanoid robot.

IV. METHODS

A rigid body has infinitely many possible poses in $SE(3)$. It is intractable to search the whole $SE(3)$ space to find functional poses. According to the IBD of chairs, a chair should be necessarily stable when an agent is sitting on it. Therefore, our approach first finds a discrete set of stable poses $G_s \subset SE(3)$ via *stable pose imagination* (Sec. IV-B). It then performs *sitting imagination* (Sec. IV-C) on each stable pose $g_s \in G_s$ to check whether it can afford the sitting functionality. If such a pose exists, the object is classified as a chair, and this pose is identified as a functional pose.

The teddy bear agent has almost rigid joints and the body configuration is very close to a sitting configuration. Therefore, we simplify the problem of putting the bear onto a chair as finding the sitting pose $g_{sit} = (R_{sit}, \mathbf{p}_{sit})$ and leave the robot-agent interaction as future work. The IBD of chairs indicates that the agent's torso is more or less upright when sitting. We thus set the rotation of the agent as $R_{sit} = R_Z(\alpha_{sit}) R_0$ in which the initial rotation R_0 puts the agent to an upright orientation. We denote the direction the agent faces as the sitting direction. The problem becomes finding one of the sitting position $\mathbf{p}_{sit}$ and the sitting direction indicated by the yaw angle α_{sit}.

A. Functionally and Perceptibly Equivalent Poses

There are infinitely many stable poses of an object. But we notice that many of them are *equivalent* in terms of functionality and perceptibility. For example, an upright chair can always afford the sitting functionality regardless of any planar motions (translation in the xy-plane and rotation about the z-axis). When an object is placed on a plane, the upper side can be directly perceived by the robot regardless of any planar motions. Planar motions form a subgroup of $SE(3)$:

$$H \doteq \{g \in SE(3) | g = (R, \mathbf{t}), R = R_Z(\alpha), \mathbf{t} = [x, y, 0]^T\} \cong SE(2) \tag{1}$$

in which $\alpha \in [0, 2\pi)$, $x, y \in \mathbb{R}$. And any $g \in SE(3)$ together with H forms a coset[3] Hg [25]:

$$Hg = \{h \circ g : h \in H\} \subseteq SE(3) \tag{2}$$

The equivalence of functionality and perceptibility is basically saying any two poses belonging to the same coset are equivalent in terms of functionality and perceptibility. In other word, for any two poses $g_1 = (R_1, \mathbf{p}_1)$ and $g_2 = (R_2, \mathbf{p}_2)$, if there exists an $h \in H$ such that $g_2 = h g_1$, we say g_1 and g_2 are *functionally equivalent* and *perceptibly equivalent*. We further define two rotations R and R' are functionally and perceptibly equivalent if there exists $\alpha \in [0, 2\pi)$ such that $R' = R_Z(\alpha) R$. Otherwise, they are functionally and perceptibly unique. The relative rotation matrix R_{12} can be decomposed

---

[3] For brevity, we refer right cosets as cosets in this paper.

as $R_{12} = R_2 R_1^T = R_z(\alpha_{12}) R_{xy}(\phi_{12})$ where $R_{xy}(\cdot)$ is a rotation about an axis in the xy-plane and ϕ_{12} is the rotation angle. In practice, we consider two poses to be functionally and perceptibly equivalent if:

$$\phi_{12} < \Delta\phi_{es}, \quad |z_1 - z_2| < \Delta z_{es} \tag{3}$$

$\Delta\phi_{es}$, and Δz_{es} are two thresholds.

We further define *perceptible difference* d_{pcp} to describe the perceptibility variation when the object is placed in two poses:

$$d_{pcp}(g_1, g_2) = \phi_{12} = \arccos(r_{33}) \tag{4}$$

where $R_{12} = [r_{ij}]_{1\le i\le 3, 1\le j\le 3}$, and r_{33} is the $(3,3)$ element of R_{12}. For any $h_1 g_1 \in H g_1$ and $h_2 g_2 \in H g_2$, the perceptible difference are the same, *i.e.*, $d_{pcp}(h_1 g_1, h_2 g_2) = d_{pcp}(g_1, g_2)$. Proof can be found in the supplementary material on our project page.

B. Stable Pose Imagination

In stable pose imagination, our algorithm simulates dropping an object in different initial poses onto a flat plane to find a set of stable poses G_s as in [3]. For any $R \in SO(3)$, there exists a rotation $R' = R_{ZYX}(0, \beta, \gamma)$ that is functionally equivalent to R. This is because R can be decomposed as $R = R_{ZYX}(\alpha, \beta, \gamma) = R_Z(\alpha) R_{ZYX}(0, \beta, \gamma) = R_Z(\alpha) R'$. That is, R and R' belong to the same coset, and thus show functional equivalence. Therefore, for any stable pose $g_s = (R_s, \mathbf{p}_s)$, there exists a rotation $R_{ZYX}(0, \beta_s, \gamma_s)$ which is functionally and perceptibly equivalent to R_s. If the object is dropped from such a rotation in the simulation, it will very likely result in a pose that is functionally and perceptibly equivalent to g_s. Thus, the algorithm enumerates the initial rotations of the object before dropping by varying the roll γ and pitch β while keeping $\alpha = 0$. For each dropping, it adds the newly found stable pose to G_s if it is functionally and perceptibly unique to all the poses in G_s.

We note that the stability of a pose g is equivalent to all the poses in its coset Hg because planar motions do not change stability. That is, if g is a stable pose, hg is also a stable pose for any $h \in H$. And finding a stable pose means finding a coset of stable poses. If g is a stable pose, we define Hg as a stable pose coset. We claim that given a sufficiently fine enumeration of β and γ, the set of cosets formed by all the found stable poses $S = \{Hg_s | g_s \in G_s\}$ contains all the stable pose cosets of the object. Proof can be found in the supplementary material.

C. Sitting Imagination

To imagine the object sitting affordance, the robot simulates a passive agent sitting on the object with it in a stable pose. Here we revisit the imagination method introduced in [3], [4] briefly. The human agent is modeled with an articulated human body with the forearms and feet trimmed off [26], [27]. For each $g_s \in G_s$, the algorithm enumerates the orientations of the object by varying α in a discrete increment while fixing $\beta = \beta_s$, $\gamma = \gamma_s$. The enumerated object poses are functionally equivalent to g_s. It drops the agents onto the object with it in different enumerated orientations. Before each drop, it positions the agent above the object and sets it to a pre-sitting configuration C_{pre} as shown in Fig. 3(b). The dropping position is enumerated along the positive x-axis of the agent.

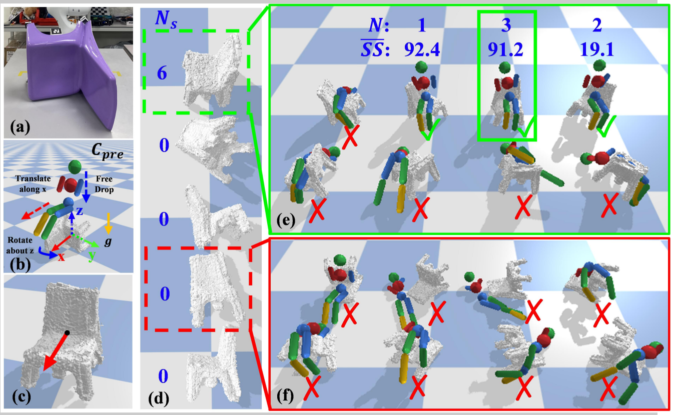

Fig. 3. Sitting Imagination and Applications. (a) A test chair. (b) Sitting imagination setting. It shows the agent's initial configuration C_{pre} before dropping. (c) The chair model is in a functional pose. The black dot shows the sitting position $\mathbf{p}_{sit}$; the red arrow shows the sitting direction α_{sit}. (d) A set of stable poses G_s. The total number of correct sittings N_s of each stable pose is shown on the left. (e) and (f) show the resultant configurations of sitting imagination of two stable poses. The correct sitting number N and average sitting configuration score $\overline{SS}$ for each orientation are also shown. The stable pose in the green frame is recognized as a functional pose. In (e), the three sitting configurations with a check are correct sittings, with the framed one being the best sitting. The stable pose in the red frame is not a functional pose. In (f), there is no correct sitting.

For each drop, the agent resultant configuration C_{res} is evaluated by the sitting affordance model (SAM) developed from [3], [4] with five criteria. The first two criteria are similar to that in [3], [4]. **1) Joint Angle.** The joint angle score J is the weighted L1 distance between the joint angle vector of C_{res} and a key configuration C_{key}, a pre-defined ideal sitting configuration of the agent. Lower is better. **2) Link Rotation.** Link rotation score L is the weighted angular distance between the z-axis of all the links in C_{res} and C_{key}. Lower is better. **3) Contact Points.** In IBD, the buttocks and back of the agent are supported by the object when it is sitting. Therefore, the number of contact points between the agent and chair is considered. When sitting, the agent usually rests on the back with the torso and/or head in contact with the chair. In other cases, the back of the chair supports the body on the shoulders. The algorithm counts the upper body contact point $P_{upper} = P_h + P_t$ if $P_h + P_t > 0$; $P_{upper} = (P_{ls} + P_{rs})/2$ if otherwise. P_h, P_t, and P_{ls}/P_{rs} are the contact points of the head, torso, and left/right shoulder links, respectively. The contact point number of the whole body is computed as $C = P_{upper} + P_p + P_{rt} + P_{lt}$ where P_p and P_{lt}/P_{rt} are the contact point of the pelvis and left/right thigh links, respectively. **4) Symmetry.** Human bodies are generally symmetric when sitting. Therefore, we add the symmetry score S to emphasize the body symmetry, *i.e.*, $S = |j_{ls} - j_{rs}| + |j_{lt} - j_{rt}| + |j_{lk} - j_{rk}|$. j_{ls}/j_{rs}, j_{lt}/j_{rt}, and j_{lk}/j_{rk} are the joint angles of the left/right shoulders, thighs and knees, respectively. Lower is better. **5) Sitting Height.** Sitting height is also an important factor in sitting. The thigh height H_t and pelvis height H_p are the average height of all the contact points of the thighs and pelvis, respectively. We use a **Sitting Configuration Score**

54 pt
0.75 in
19.1 mm

54 pt
0.75 in
19.1 mm

to measure the difference $SS = \frac{H_t}{JLH_{agent}}$ between C_{res} and C_{key}. H_{agent} is the agent size. Higher is better.

A resultant agent configuration is considered as a *correct sitting* if:

$$J < J_{max}, L < L_{max}, C > C_{min}, S < S_{max}, SS > SS_{min} \tag{5}$$

$$P_{upper} > 0, P_{lt} > 0, P_{rt} > 0 \tag{6}$$

$$H_t \in (H_{tmin}, H_{tmax}), H_p \in (H_{pmin}, H_{pmax}) \tag{7}$$

$$|H_p - H_t| < \Delta H_{min} \tag{8}$$

J_{max}, L_{max}, S_{max}, and SS_{min} are thresholds corresponding to different scores. C_{min} is a minimum contact point number threshold. We note that the object can support the agent body with fewer contact points when the agent is in a (nearly) symmetrical configuration. Thus, our algorithm loosens the contact point threshold if the symmetry score is smaller than a more restricted threshold. H_{tmin}, H_{tmax}, H_{pmin}, H_{pmax}, and ΔH_{min} are thresholds corresponding to the sitting height criteria. Eqn. 8 gives a more strict restriction of the contact between the thigh and the object. Considering a more comprehensive evaluation of contact points, sitting height, and symmetry, our method has a good performance when generalized to real data and the complicated application of chair preparation. More details can be found in the supplementary material.

D. Application of Imagination

The imagination method introduced in Sec. IV-B and IV-C can be used for 1) object reconstruction, 2) chair v.s. non-chair classification, 3) functional pose prediction, and 4) finding the sitting pose of the agent. Fig. 2 shows the pipeline.

Given an unseen object placed on a plane in a random pose g, the robot first scans the object and reconstructs an incomplete model as part of the object is occluded by the plane. In order to reconstruct a complete object model, it needs to find a stable pose that exposes the occluded part the most. This means finding the stable pose coset which has the largest perceptible difference from the current pose. Therefore, the robot performs stable pose imagination with this incomplete model to find a set of stable poses G_s. The perceptible difference d_{pcp} between each stable pose $g_s \in G_s$ and g is calculated to find the stable pose g_s^{pcp} with the largest d_{pcp}. The poses in the stable pose coset Hg_s^{pcp} offers the best perceptibility for the occluded part. The robot manipulates the object to a pose $hg_s^{pcp} \in Hg_s^{pcp}$ that is perceptibly equivalent to this pose and scans the object again. The complete object model is reconstructed by fusing this scan with the initial one. More details can be found in V-C.

Stable pose imagination is performed again on the complete object model to find the set of stable poses G_s. The algorithm then performs sitting imagination on each $g_s \in G_s$ to check if any g_s is a functional pose. For every g_s, it counts the number of correct sittings N for each orientation α. It accumulates N of all orientations as the correct sitting number N_s for g_s. The stable pose g_s that has the largest N_s is selected as the candidate functional pose. It classifies an object as a sittable chair if the largest N of the candidate functional pose satisfy:

$$\max N > N_{min} \tag{9}$$

Fig. 4. Experiment Details. (a) Robot Experiment Data. The right chair is used for parameter tuning. The rest 12 objects are previously unseen and are used for testing our method. (b) A test chair with handles attached under the seat. (c) The end effector is mounted with two RGB-D cameras for scanning and tracking respectively.

N_{min} is thresholds. Otherwise, the object is classified as a non-chair.

If an object is recognized as a chair, the candidate functional pose is a functional pose denoted as g_f. Each orientation α of g_f generates an imagined sitting pose if the number of correct sittings $N > 0$. The sitting position $\mathbf{p}_{sit}$ and direction α_{sit} of this orientation is the weighted average of the agent base link position and yaw angle of the correct sittings. The weight for each sitting is its corresponding SS. And for each orientation, the algorithm computes the average sitting configuration score $\overline{SS}$ of all the correct sittings. It ranks the agent sitting poses by N. If more than one sitting pose has the same N, it ranks the one with the larger $\overline{SS}$ higher. See Fig. 3 for an example. More details on preparing the chair for sitting can be found in Sec. V-D.

V. EXPERIMENTS

Fig. 1(a) shows the experimental setting. The object is randomly placed on a table in front of the robot arm. A Franka Emika robot arm is used to manipulate the object. Two RGB-D cameras are mounted on the end effector to scan the object and detect the grasping position, respectively (Fig. 4(c)). A NAO humanoid robot is used to carry the teddy bear and seat it on the chair. An ArUro tag is placed on top of the NAO for tracking.

A. Data

The data of the experiment contains 7 real chairs and 6 non-chair objects that have different sizes, shapes, and appearances (Fig. 4(a)). The chairs are all designed for 0-3-year-old children. We choose kiddie chairs because the size and weight of the chair are restricted by the workspace and payload of robots. If the chair is too tall, the NAO is not able to reach the seat; if the chair is too heavy or large, the robot arm cannot manipulate it.

The imagination model introduced in Sec. IV is developed from that in [3]. To decrease the sim-to-real gap, we tune the parameters of our model using the 30 synthetic chairs of the synthetic dataset in [3] and one real chair shown in Fig. 4(a). The real chair is also used for determining the parameters in object reconstruction, motion planning, and control modules. The rest chairs and all the non-chairs, which are unseen by the robot, are used as the test set.

B. Robot Arm Manipulation

We attach handles to all the objects to simplify the grasping problem in this paper. An Alvar tag is attached to each handle for the robot to track the handle pose. The robot detects handles autonomously, without any prior knowledge of the number of handles and handle poses. In particular, for each chair, we attach four handles underneath the seat. We make sure the design and arrangement 1) do not change the object bounding box (OBB) of the chair, 2) do not affect the seat and back of the chair, 3) do not change the stable poses of the chair, and 4) guarantee that there is always at least one handle that is reachable by the robot arm for any orientation of the chair. After each manipulation, the grasped handle pose is tracked to provide an estimation of the object transformation in the manipulation. IKFast [28] is used to solve the inverse kinematics of the robot arm. MoveIt [29] is used to plan the motion. When putting the object down, the robot uses a simple force controller, which terminates the downward motion of the arm when the vertical force exceeds a threshold.

C. Object Reconstruction

Part of the object is occluded when it is placed on the table. In order to reconstruct a complete model, the robot manipulates the object to multiple perceptibly unique stable poses and scans it. For each perceptibly unique stable pose, the robot arm moves to a set of pre-defined poses to capture depth images of the scene. We call it a scan. The object point cloud is segmented from the point cloud of the scene which is reconstructed with TSDF fusion [30]. The object transformation by manipulation is first estimated from the pose change of the handle and then refined by registering the object point clouds of both scans using iterative closest points (ICP) [21]. The complete object model is reconstructed by integrating depth images captured from different scans. Specifically, the algorithm transforms all the captured depth images into the same object frame with the estimated object transformations and uses TSDF fusion to reconstruct the complete model. To improve the quality of the complete model, it filters the depth image by removing the pixels corresponding to the table and the occluded part of the object as the depth values of the occluded part are usually noisy. During each object reconstruction, it attaches the object frame to the geometric center of the reconstructed model and set its rotation according to the world frame and the OBB.

D. Chair Preparation

If the object is recognized as a chair, it can be non-upright or upright but inaccessible. In this case, the robot arm needs to manipulate the chair to a functional *and* accessible pose to prepare it for the NAO to seat the teddy bear. In our setting, an upright chair is accessible if it is facing the NAO robot and inaccessible if otherwise (*i.e.*, the chair faces the robot arm or the longer edge of the table). The NAO uses the same method in [4] to plan the $SE(2)$ trajectory to walk to the chair and the whole-body motion to seat the bear. In the experiments, for a functional pose, our approach considers it inaccessible if the planning fails. To find a functional and accessible pose, the robot first tries to plan the motion to seat the bear with the highest-ranked imagined sitting pose. If the planning is not successful, the robot tries the next-ranked imagined sitting pose. If the planning is successful, the robot further plans the walking trajectories with the chair placed in different functional poses $hg_f \in Hg_f$. The functional pose which results in successful planning is considered accessible. The robot manipulates the chair to this pose. The NAO executes the motions to walk to the chair and seat the bear on the chair.

E. Baseline

We compared our method with five baseline methods on functional pose prediction of chairs. **1) ICP Canonical.** Given a canonical chair and its functional pose, denoted as (R_{cano}, p_{cano}). This baseline uses ICP to register the point cloud of unseen chairs to the point cloud of the canonical chair. The relative pose given by the registration is R_{reg}. The rotation $R_{cano}R_{reg}$ is applied to the unseen chair. **2) OBB Random.** The algorithm randomly selects one face of the object OBB and drops it with the selected face facing vertically downwards. The idea comes from the observation that a chair often has one of the OBB faces in contact with the ground when it is in a functional pose. **3) OBB Select.** The algorithm casts a ray from the center of each OBB face to the center of its opposite one. The face of which the ray has the longest collision-free section before intersecting with the object is selected to face vertically upwards. We note that the chair seat faces upwards and has no obstruction above when the chair is upright. This algorithm uses raycasting to check the existence of the seat and the free space above it. For the first three baselines, the prior knowledge that the object is a chair is given. The object is dropped from the selected orientation using the same setting in stable pose imagination. The resultant stable pose is the predicted functional pose. **4) OBB Stable + Imagination.** The algorithm drops the object from the pose with each OBB face facing down to find a set of stable poses. Sitting imagination is then performed on these stable poses. In this baseline, we replace stable pose imagination with OBB Stable to explore the effectiveness of stable pose imagination. **5) Incomplete Reconstruction + Imagination.** To study the effectiveness of complete object reconstruction, we perform imagination on incomplete object models reconstructed from the initial scanning. For these two baseline methods, the prior knowledge that the object is a chair is not given.

TABLE I
RESULTS OF FUNCTIONAL POSE PREDICTION

Method	Prior	Success Rate (%)
ICP Canonical	✓	23.3
OBB Random	✓	13.3
OBB Select	✓	26.7
OBB Stable + Imagination	✗	76.7
Incomplete Object Model + Imagination	✗	80.0
Imagination (Ours)	✗	**100.0**

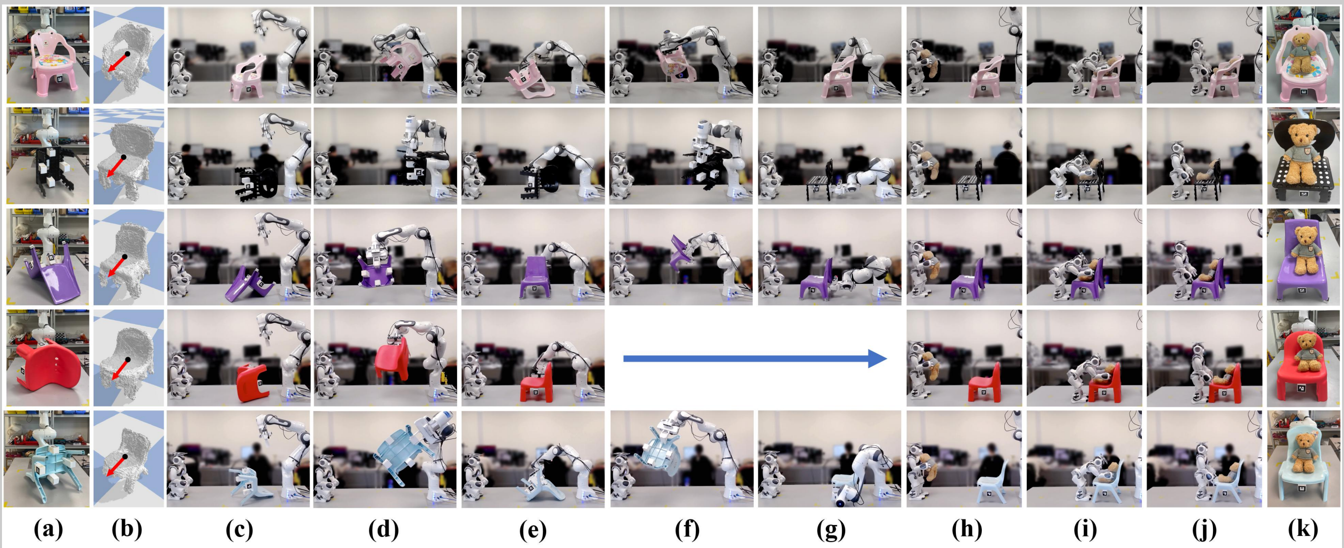

Fig. 5. Real Robot Experiment Results. (a) Snapshot of the chair. (b) Imagined sitting pose. (c) Initial. (d)-(e) Rotating the chair to a perceptibly different pose to view the occluded part. (f)-(g) Preparing the chair to a functional and accessible pose. (h)-(j) Seating the bear on the chair. (k) Results.

F. Annotation and Evaluation

We recruited 5 volunteers to annotate the experiment result. For each object, we asked: “Do you think the object is a chair that is able to afford sitting?” If it was annotated as a chair, we further asked the volunteers to annotate the functional pose. For each experimental trial, we showed the experiment video and images, then asked: “Do you think the object is in an upright pose that is accessible for the NAO robot?” If the answer was positive, we asked: “Do you think the robot has been successful in seating the bear on the chair?” For each question, we considered the answer positive if more than 3 out of 5 volunteers gave a positive answer.

To evaluate the result of functional pose prediction, we compared the predicted pose with the annotated pose. If they were functionally equivalent, it was considered correct. If a chair object is classified as a non-chair, we considered it incorrect.

VI. RESULTS

We tested each object by placing it in random poses that are functionally unique to each other. In particular, the six unseen chairs were placed in five functionally unique poses. In total, we performed 45 trials, including 30 trials of chairs and 15 trials of non-chairs, on the 12 objects in the test set.

A. Functional Pose Prediction

The results are shown in Tab. I. The performance of ICP canonical is relatively low. The large shape and size variation and the random orientation of the chairs bring challenges to registration. The success rate of OBB Random is slightly lower than 1/6. This verifies our heuristic that a chair has one of its OBB faces contacting the ground when it is in a functional pose. But there are cases in which the heuristic is not correct, leading to a lower success rate. OBB Select performs better by considering more heuristics, but the success rate is still not good. OBB Stable + Imagination performs much better. Failure occurs when the set of stable poses found does not include a functional pose. Compared with our method, it emphasizes the effectiveness of stable pose imagination. The first five baseline methods are tested on the complete model of the chairs while Incomplete Object Model + Imagination is not. It fails on 6 trials when the chair is overturned or lying on the side, resulting in poorly reconstructed seats. Our full method achieves a 100% success rate, outperforming all the comparing baseline methods.

B. Real Robot Experiments

We further evaluated our method on real robot experiments which includes four subtasks: 1) chair classification, 2) functional pose prediction, 3) chair preparation, and 4) seating the bear. Qualitative results are shown in Fig. 5. For chair classification, our algorithm achieves a 100% success rate, correctly classifying the object in all 45 trials. For chair preparation, the robot successfully prepares the chair to a functional and accessible pose in all 30 trials with chairs. In the 30 trials of chairs, the robot eventually succeeds in seating the bear on the chair in 29 trials, achieving a success rate of 96.7%. A complete trial costs about 800s, in which object reconstruction, chair preparation, and robot seating the bear accounts for 358s, 279s, and 166s, respectively. Our algorithm can be accelerated by parallel computing, multi-camera setups, and higher robot execution speed. In real-world deployment, the robot could conduct offline perception and imagination for novel objects during the night, and perform real-time manipulation in the daytime. Direct comparison with learning-based methods would require their training time to be considered. The failure case in our method is a small chair without an armrest. The highest-ranked imagined sitting pose is a pose in which the agent does not face the direction perpendicular to the back of the chair. The control error and the restricted workspace of the NAO robot make it difficult to

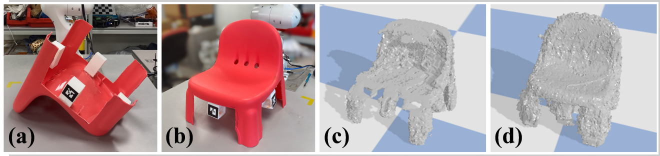

Fig. 6. Reconstruction Results. (a) Initial pose of a test chair. (b) The chair in an upright pose. (c) Incomplete model reconstructed from data captured when the chair is in the initial pose. (d) Complete model reconstructed by manipulation.

put the teddy bear precisely according to the imagined sitting poses. Thus, the bear has its head somewhat supported by the back, which is not considered a success by 3 out of the 5 annotators.

In Fig. 6, we show an example of the incomplete object model and the complete model reconstructed using our method. The complete model provides a more adequate representation of the seat and the back, allowing the robot to more accurately imagine the functionality and interact with the object physically.

VII. CONCLUSIONS & FUTURE WOWK

In this letter, we proposed a novel method based on real2sim2real transfer in which robots imagine the "sittability" (affordance of chairs) of a previously unseen object and use this affordance-based reasoning to guide robot-object interaction. We develop a robot manipulation system to actively perceive the object, prepare the chair for sitting and seat a teddy bear on the chair autonomously. Results show that our method enables the robot to manipulate 6 novel chairs in 30 trials to a functional and accessible pose and seat the bear on them with a very high success rate. Our method can be applied to various real-world tasks including tidying up the chairs, preparing chairs for a meeting, and can be possibly extended to help elderly people with sitting. In the future, we plan to extend our imagination method to other objects that involve human whole-body interactions and explore more functionality-related physical properties through active interaction.

54 pt
0.75 in
19.1 mm

54 pt
0.75 in
19.1 mm

REFERENCES

[1] J. J. Gibson, *The ecological approach to visual perception*. Boston, MA, USA: Houghton Mifflin, 1979.
[2] L. Bogoni and R. Bajcsy, "Interactive recognition and representation of functionality," *Computer Vision and Image Understanding*, vol. 62, no. 2, pp. 194–214, 1995.
[3] H. Wu, D. Misra, and G. S. Chirikjian, "Is that a chair? imagining affordances using simulations of an articulated human body," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7240–7246.
[4] H. Wu, X. Meng, S. Ruan, and G. S. Chirikjian, "Put the bear on the chair! intelligent robot interaction with previously unseen chairs via robot imagination," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 6276–6282.
[5] S.-B. Ho, "Representing and using functional definitions for visual recognition," Ph.D. dissertation, The University of Wisconsin-Madison, 1987.
[6] L. Piyathilaka and S. Kodagoda, "Understanding human context in 3D scenes by learning spatial affordances with virtual skeleton models," in *arXiv preprint arXiv:1906.05498*, 2019.
[7] T.-T. Do, A. Nguyen, and I. Reid, "AffordanceNet: An End-to-End Deep Learning Approach for Object Affordance Detection," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 5882–5889.
[8] F.-J. Chu, R. Xu, and P. A. Vela, "Learning Affordance Segmentation for real-World Robotic Manipulation via Synthetic Images," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1140–1147, 2019.
[9] H. Wu and G. S. Chirikjian, "Can I pour into it? Robot Imagining Open Containability Affordance of Previously Unseen Objects via Physical Simulations," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 271–278, 2020.
[10] P. Mandikal and K. Grauman, "Learning dexterous grasping with object-centric visual affordances," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 6169–6176.
[11] R. Xu, F.-J. Chu, C. Tang, W. Liu, and P. A. Vela, "An affordance keypoint detection network for robot manipulation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2870–2877, 2021.
[12] W. Liang, Y. Zhao, Y. Zhu, and S.-C. Zhu, "Evaluating human cognition of containing relations with physical simulation."
[13] H. Liu, C. Zhang, Y. Zhu, C. Jiang, and S.-C. Zhu, "Mirroring without overimitation: Learning functionally equivalent manipulation actions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 8025–8033.
[14] E. Heiden, M. Macklin, Y. Narang, D. Fox, A. Garg, and F. Ramos, "Disect: A differentiable simulation engine for autonomous robotic cutting," in *Proceedings of Robotics: Science and Systems*, 2021.
[15] Y. Zhu, Y. Zhao, and S.-C. Zhu, "Understanding Tools: Task-Oriented Object Modeling, Learning and Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
[16] S. Boschert and R. Rosen, "Digital Twin—The Simulation Aspect," in *Mechatronic Futures*, 2016, pp. 59–74.
[17] L. Hinkle and E. Olson, "Predicting object functionality using physical simulations," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 2784–2790.
[18] H. Grabner, J. Gall, and L. Van Gool, "What makes a chair a chair?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1529–1536.
[19] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," in *Robotics: Science and Systems (RSS)*, 2018.
[20] C. Mitash, A. Boularias, and K. Bekris, "Physics-based scene-level reasoning for object pose estimation in clutter," *The International Journal of Robotics Research*, vol. 41, no. 6, pp. 615–636, 2022.
[21] P. J. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
[22] M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, and J. Sivic, "Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 3762–3769.
[23] P. Kadam, Q. Zhou, S. Liu, and C.-C. J. Kuo, "Pcrp: Unsupervised point cloud object retrieval and pose estimation," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 1596–1600.
[24] X. Li, Y. Weng, L. Yi, L. J. Guibas, A. Abbott, S. Song, and H. Wang, "Leveraging se (3) equivariance for self-supervised category-level object pose estimation from point clouds," *Advances in Neural Information Processing Systems*, vol. 34, pp. 15 370–15 381, 2021.
[25] G. S. Chirikjian, *Stochastic models, information theory, and Lie groups, volume 2: Analytic methods and modern applications*. Springer Science & Business Media, 2011, vol. 2.
[26] H. Forssberg and H. Hirschfeld, "Postural adjustments in sitting humans following external perturbations: muscle activity and kinematics," *Experimental Brain Research*, vol. 97, no. 3, pp. 515–527, 1994.
[27] K. Kerr, J. White, D. Barr, and R. Mollan, "Analysis of the sit-stand-sit movement cycle in normal subjects," *Clinical Biomechanics*, vol. 12, no. 4, pp. 236–245, 1997.
[28] R. Diankov, "Automated construction of robotic manipulation programs," Ph.D. dissertation, Carnegie Mellon University, 2010.
[29] S. Chitta, I. Sucan, and S. Cousins, "Moveit!" *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.
[30] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, pp. 303–312.