R2HandoverSim: A Simulation Framework and Benchmark for Robot-to-Human Object Handovers

Hanxin Zhang[1,2], Abdulqader Dhafer[1,2], Hongbiao Dong[3], Zhou Daniel Hao[1,2]

***Abstract*— We present R2HandoverSim, a simulation benchmark for robot-to-human (R2H) object handovers. Although R2H handover methods have advanced rapidly, the lack of standardized evaluation protocols impedes objective comparison. Our benchmark enables reproducible evaluation by systematically comparing four baselines on their predicted shared grasp poses. We conduct a user study with 30 participants, analyze baseline performance, and show that simulation results correlate with real-world evaluation outcomes. Crucially, five complementary metrics (planning feasibility, reachability, grasp stability, grasp affordance, and safety) better reflect user-perceived handover quality than overall success rate alone. Website and code: https://robot-future.github.io/r2handoversim/.**

I. Introduction

Robot-to-human (R2H) object handover, in which a robot transfers an object to a human receiver, is a fundamental capability of service, assistive, and collaborative robotic systems [1]. Recent methods focus on predicting a *shared grasp pose*, the configuration in which the robot presents the object so that the human receiver can grasp it. Approaches to this problem span heuristic optimization [2], contact prediction [3], and generative models [4]; yet they differ substantially in input modalities, object sets, and evaluation protocols. Because no standardized benchmark exists, comparisons across R2H methods ultimately rely on user studies whose results are sensitive to participant variability and difficult to reproduce.

In the human-to-robot (H2R) direction, simulation benchmarks such as HandoverSim [5] and GenH2R [6] have already established shared evaluation standards and accelerated progress. In parallel, datasets that capture interactions between hands and objects [7], [8], [9] provide reference hand configurations that can define handover goals. Unlike H2R handovers, the R2H direction still lacks an analogous simulation benchmark for standardized evaluation.

Constructing such a benchmark is nontrivial because R2H success is harder to operationalize than H2R success. In H2R handovers, success is typically measured by the robot's ability to grasp and retain an object, whereas R2H success depends on safe and comfortable reception by the human, whose adaptive behavior is difficult to model in simulation. However, most R2H methods converge on predicting a shared grasp pose, making simulation evaluation feasible: rather than modeling full human grasping behavior, one can assess whether each predicted pose satisfies geometric, kinematic, and safety constraints in a physics simulator.

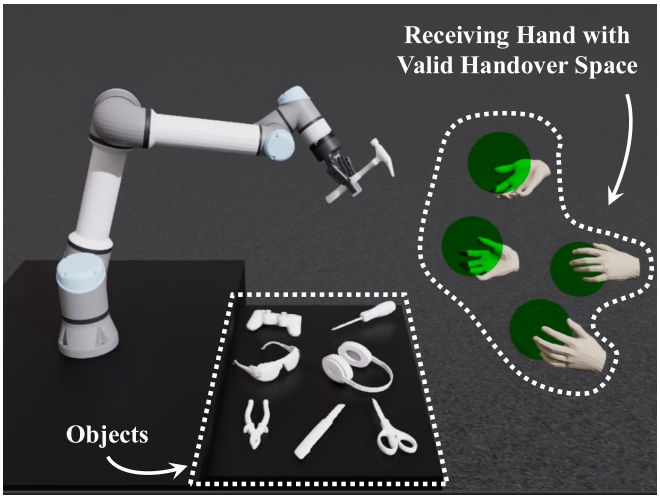


Fig. 1: Overview of the R2HandoverSim benchmark environment. A UR5e manipulator delivers an object to a static MANO hand. The green sphere denotes the valid handover space used for reachability evaluation.

In this paper, we make three contributions. **(1)** We introduce R2HandoverSim (Fig. 1), a simulation benchmark that enables standardized and reproducible evaluation of R2H handover methods across diverse objects and evaluation settings. **(2)** We define an evaluation protocol with five complementary criteria (planning feasibility, reachability, grasp stability, grasp affordance, and safety) and use it to systematically compare four baselines. **(3)** We conduct a user study with 30 participants, showing that simulation evaluation identifies the top method in real world deployment while subjective ratings reveal additional cross method differences, validating the benchmark's practical relevance.

II. Related Work

Robot-to-Human Handovers. The goal of R2H handover is to present an object so that the receiver can grasp it safely and use it immediately [10], [11]. Existing R2H studies can be divided into methods for general handover and methods for task-specific handover. General methods predict shared grasp poses through rule-based optimization [2], contact prediction [3], semantic and geometric reasoning [12], and human pose estimation [4]. Task-specific methods target scenarios such as assistive handover for older adults [13] and bimanual object transfer [14]. However, these works vary widely in object sets

[1]Authors are with DANiLab, University of Leicester, Leicester, UK {hz273,aamd2,d.hao}@leicester.ac.uk.

[2]Authors are with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, UK.

[3]The author is with the School of Metallurgy and Materials, University of Birmingham, Birmingham, UK. h.dong.1@bham.ac.uk.

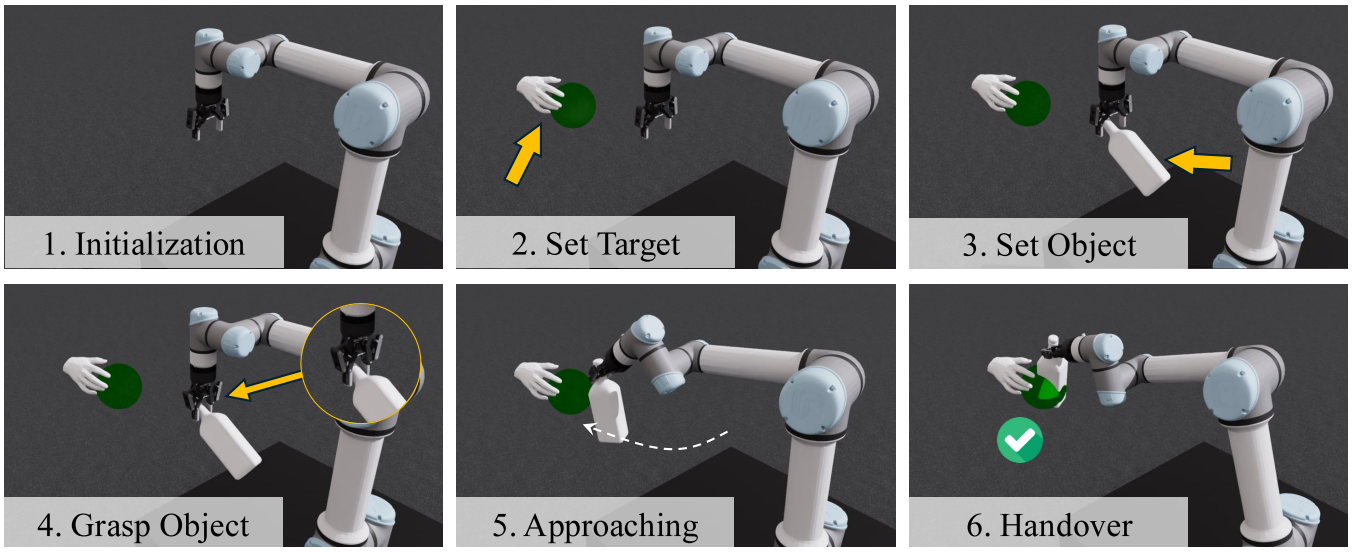


Fig. 2: Trial protocol of R2HandoverSim: initialization, target pose assignment, object placement, robotic grasping, approach motion, and handover evaluation.

(from a single item to over 30 household objects), evaluation protocols (simulation only, real robot trials, or user studies with different subjective scales), and metrics (from success rate to biomechanical comfort indices), making direct comparison difficult. R2HandoverSim addresses this gap with a unified simulation environment built on a UR5e robot with a Robotiq 2F-85 gripper and 16 objects, together with five evaluation criteria covering planning feasibility, reachability, grasp stability, grasp affordance, and safety (Section III-D), allowing R2H methods to be compared on equal footing for the first time.

Handover Benchmark. In the H2R direction, benchmarks have already driven rapid progress. Hand-object interaction datasets such as H2O [7], GRAB [8], ARCTIC [9], ContactPose [15], and DexYCB [16] provide reference grasps for evaluating grasp feasibility, while simulation benchmarks such as HandoverSim [5], GenH2R [6], MobileH2R [17], and DexH2R [18] offer standardized environments for H2R policies. These resources share two ingredients transferable to R2H: curated hand-object interaction data defining plausible handover goals, and physics simulation enabling reproducible evaluation. No benchmark yet combines them for R2H. R2HandoverSim fills this gap by integrating hand-object interaction sequences from established datasets with robotic grasp candidates in a physics simulation, providing the first standardized R2H *simulation* benchmark.

III. R2HandoverSim: The Benchmark

A. Task Formulation

We consider the robot-to-human (R2H) object handover task, in which a robotic manipulator transfers an object to a human receiver at a target handover pose. At the start of each trial, the robot is initialized at a fixed home configuration $q_0 \in \mathbb{R}^n$, where n denotes the number of actuated joints. The receiver hand pose $T_{\text{hand}} \in SE(3)$ is randomly sampled per trial from a reachable set and then fixed in the world frame for that trial; its position $\mathbf{p}_h \in \mathbb{R}^3$ and outward palm normal $\mathbf{n}_h \in \mathbb{R}^3$ are extracted for computing evaluation criteria.

Each method outputs two end-effector poses: the grasp pose $T^g_{ee} \in SE(3)$ at which the robot holds the object, and the handover pose $T^h_{ee} \in SE(3)$ at which the object is delivered to the receiver. The benchmark does not restrict how these poses are generated. Once T^g_{ee} is specified, the object is placed relative to the end-effector and the gripper closes. If the object width along the closing axis exceeds the maximum aperture, the trial is recorded as a drop failure; otherwise, the object is rigidly attached for subsequent planning and execution.

Starting from q_0, the robot executes an inverse kinematics and motion planning pipeline to reach T^h_{ee}. Inverse kinematics is solved with a numerical Jacobian iterative solver; motion planning uses RRT-Connect within MoveIt with collision checking enabled for the robot arm, end-effector, attached object, and MANO hand mesh. The complete trial protocol is illustrated in Fig. 2. Each trial is evaluated based on the following five criteria.

- **Plan:** A kinematically feasible, collision-free trajectory to the handover pose must exist.
- **Reach:** The object must be delivered near the receiver's hand.
- **Stability:** The gripper must physically close on the object without dropping it.
- **Affordance:** The robot's grasp must not occupy the receiver's intended grasp region.
- **Safe:** The delivery motion must complete without contacting the human hand.

B. Simulation Environment

All experiments are conducted in NVIDIA Isaac Sim 5.0.0 [19], which provides high-fidelity simulation quality and ease of use compared with engines such as MuJoCo and PyBullet. The scene comprises a UR5e manipulator with a

Robotiq 2F-85 parallel gripper, a static table, and a static MANO hand mesh representing the receiver. Each object mesh is loaded as a rigid body with convex hull collision geometry, aligned with its canonical coordinate frame.

In the preparation stage, the benchmark provides ground-truth object representations for each trial (mesh, point cloud, voxel grid, and receiver MANO hand pose); each method reads only the inputs it requires and estimates the two poses T_{ee}^{g} and T_{ee}^{h} defined in Sec. III-A. In the execution stage, given T_{ee}^{g}, the gripper closes on the object and checks width feasibility ($w_{\mathcal{O}}(T_{ee}^{g}) \leq w_{\max}$); if this check passes, the object is rigidly attached to the end-effector. Given T_{ee}^{h}, the robot solves $f(q_h) = T_{ee}^{h}$ for a collision-free joint configuration and executes the planned trajectory in simulation.

C. Datasets

We source 16 benchmark objects from the OakInk object set [20], which collectively span ShapeNet [21], YCB [22], and ContactDB [23]. The selected objects are shown in Fig. 3. OakInk is widely used in robot manipulation and hand-object interaction, enabling direct comparability with prior work. Objects are selected to cover (1) whether the handover pose is functionally constrained (e.g., screwdrivers, drills) or admits flexible orientations (e.g., bowls, game controllers), and (2) a range of sizes from compact (e.g., cups, cans) to bulky (e.g., bottles, dispensers).

To support methods that predict hand-object interactions, we draw handover segments from H2O [7], GRAB [8], and ARCTIC [9], and annotate them with textual instructions (e.g., *hand over the screwdriver handle to human*). We use 200 left-hand sequences and 200 right-hand sequences, with GRAB contributing the largest share. For robot grasp supervision, we use Multi-GraspLLM [24] to generate affordance-aware grasp proposals for parallel grippers in the canonical object frame, together with functional region segmentation of the object surface. For each object, we retain the top 100 candidate grasps.

D. Evaluation Metrics

We evaluate each trial across five binary metrics: planning feasibility, reachability, grasp stability, grasp affordance, and safety. Throughout this subsection, let $g = T_{ee}^{g}$ and $h = T_{ee}^{h}$ for conciseness.

Planning feasibility checks whether the robot can reach handover pose h with a collision-free IK solution and a feasible motion plan under grasp g:

$$\text{Plan}(g, h) \;\Leftrightarrow\; h \in \mathcal{S}_{\text{plan}}(g), \tag{1}$$

where $\mathcal{S}_{\text{plan}}(g) \subset SE(3)$ denotes the set of handover poses that satisfy both conditions.

Reachability evaluates whether the object is delivered into a predefined region near the human hand. Let $\mathcal{B}_h$ denote a sphere centered at $\mathbf{c}_h = \mathbf{p}_h + d_h \mathbf{n}_h$ with radius r_h, where $\mathbf{p}_h$ and $\mathbf{n}_h$ are the palm position and outward normal for each trial; we set $d_h = 12\,\text{cm}$ and $r_h = 10\,\text{cm}$. Let $\mathcal{V}_{\mathcal{O}}(g, h)$ denote the object volume in the world frame when the end effector is at h under grasp g. Reachability is defined as

Fig. 3: The 16 benchmark objects from OakInk, including objects with a functional receiver region (e.g., screwdrivers, drills), those without (e.g., bowls, game controllers), and varied sizes (cups to dispensers).

$$\text{Reach}(g, h) \;\Leftrightarrow\; \mathcal{V}_{\mathcal{O}}(g, h) \cap \mathcal{B}_h \neq \emptyset. \tag{2}$$

Grasp stability evaluates whether the gripper can retain the object at the predicted grasp pose. Given g, let $w_{\mathcal{O}}(g)$ denote the object width along the gripper closing axis, and let $w_{\max} = 85\,\text{mm}$ denote the maximum aperture of the Robotiq 2F-85 gripper. Stability is defined as

$$\text{Stability}(g) \;\Leftrightarrow\; w_{\mathcal{O}}(g) \leq w_{\max}. \tag{3}$$

Grasp affordance evaluates whether the robot fingers occlude the receiver's intended grasp area. For each object $\mathcal{O}$, let $\mathcal{G}_{\mathcal{O}}^{H} \subset \mathbb{R}^3$ denote the grasp region intended for the receiver, and let $\mathcal{G}_{\mathcal{O}}^{H}(g)$ be its rigid transform in the world frame under g. Let $\mathcal{V}_{\text{grip}}(g)$ denote the volume occupied by the gripper fingers at pose g. Affordance is defined as

$$\text{Affordance}(g) \;\Leftrightarrow\; \mathcal{V}_{\text{grip}}(g) \cap \mathcal{G}_{\mathcal{O}}^{H}(g) = \emptyset. \tag{4}$$

Safety evaluates whether the human hand contacts the robot arm or end effector during execution. Let $\mathcal{M}_{\text{robot}}(g, h)$ denote the collision volume of the robot arm and end effector during trajectory execution under (g, h), and let $\mathcal{M}_{\text{hand}}$ denote the static MANO hand mesh. Safety is defined as

$$\text{Safe}(g, h) \;\Leftrightarrow\; \mathcal{M}_{\text{hand}} \cap \mathcal{M}_{\text{robot}}(g, h) = \emptyset. \tag{5}$$

We report SR (success rate), T_{plan} (planning time), T_{exec} (execution time), $T_{\text{tot}} = T_{\text{plan}} + T_{\text{exec}}$ (total time), and failure rates F_{plan}, F_{reach}, F_{safe}, F_{stab}, and F_{afford}. In S1 (see Sec. III-E), the five criteria are evaluated sequentially: Stability is checked first (gripper closure), then Plan (IK and motion planning), then Reach, Affordance, and Safe. In S0 (see Sec. III-E), the Affordance check is not evaluated. Therefore, F_{afford} is not reported for S0 and does not contribute to SR in S0. A trial is attributed to the *first* evaluated criterion that fails; consequently, the reported failure rates are mutually

TABLE I: Baseline comparison on R2HandoverSim under S0 and S1 (see Sec. III-E), and Avg. Metric definitions follow Sec. III-D.

Split	Methods	Success† (%)	Time† (s)			Failure† (%)				
		SR	T_{plan}	T_{exec}	T_{tot}	F_{plan}	F_{reach}	F_{safe}	F_{stab}	F_{afford}
S0	FC-Handover	70.4	1.48	8.56	10.04	18.0	2.0	5.4	4.2	N/A
	Handover-VA	64.8	0.58	8.40	8.98	20.0	5.0	7.8	2.4	N/A
	Contact-Handover	77.0	0.92	8.48	9.40	13.0	3.6	2.2	4.2	N/A
	Intent-Handover	69.8	2.32	8.70	11.02	16.4	4.8	5.4	3.6	N/A
S1	FC-Handover	48.8	1.76	8.92	10.68	23.0	4.6	7.8	6.6	9.2
	Handover-VA	39.0	0.78	8.70	9.48	29.0	9.2	12.6	4.0	6.2
	Contact-Handover	55.4	1.16	8.76	9.92	19.6	8.4	5.0	7.4	4.2
	Intent-Handover	66.0	2.62	8.96	11.58	14.8	4.0	7.2	4.2	3.8
Avg	FC-Handover	59.6	1.62	8.74	10.36	20.5	**3.3**	6.6	5.4	4.6
	Handover-VA	51.9	**0.68**	**8.55**	**9.23**	24.5	7.1	10.2	**3.2**	3.1
	Contact-Handover	66.2	1.04	8.62	9.66	16.3	6.0	**3.6**	5.8	2.1
	Intent-Handover	**67.9**	2.47	8.83	11.30	**15.6**	4.4	6.3	3.9	**1.9**

†Averaged over all trials per object per method. For F_{afford}, S0 is not evaluated; in Avg, S0 is treated as 0 and averaged with S1.

Highlighting in Avg indicates best values per metric (higher for SR; lower for time and failure rates).

exclusive and sum to $1 - \text{SR}$ within each split. The overall outcome of a trial is defined as

$$\text{Success}(g,h) = \begin{cases} h \in \mathcal{S}_{\text{plan}}(g) \\ \mathcal{V}_{\mathcal{O}}(g,h) \cap \mathcal{B}_h \neq \emptyset \\ w_{\mathcal{O}}(g) \leq w_{\max} \\ \mathcal{V}_{\text{grip}}(g) \cap \mathcal{G}^H_{\mathcal{O}}(g) = \emptyset \\ \mathcal{M}_{\text{hand}} \cap \mathcal{M}_{\text{robot}}(g,h) = \emptyset \end{cases} \tag{6}$$

where Eq. (6) is the full criterion set used in S1, and S0 uses the same definition without the Affordance term.

E. Evaluation Settings

We evaluate all methods under two splits: S0 and S1. S0 contains objects with relatively unconstrained handover configurations. S1 contains objects with stronger functional constraints on grasp and handover orientation. Accordingly, affordance failure F_{afford} is evaluated in S1 but not in S0. As a result, SR in S0 is computed without the affordance criterion. These two splits correspond to the S0 and S1 rows in Table I, and Avg reports their average performance.

The robot does not perform a tabletop pick-up; instead, the gripper is initialized vertically downward and the object is loaded at the generated grasp pose relative to the end-effector (see Sec. III-A). This broadens the feasible grasp range for each object and isolates handover performance from artifacts of the simulated grasping process.

IV. Experiments

A. Baseline Implementation

FC-Handover [25]. This method uses GanHand [26] to predict the human receiving pose. GraspIt! [27] then generates robot grasp candidates opposite the human hand to avoid collision. The robot gripper is aligned 180° opposite the target hand orientation, calculated from the wrist and middle finger tip positions.

Handover-VA [2]. This method uses AffNet-DR [28] to partition the object into functional regions, within which ICP determines the grasp pose. While the original method targets the midpoint between human and robot, we adapt this by placing the gripper opposite the target hand's palm center. The orientation follows the original rule-based strategy, aligning the object's utility axis to present the handle to the human.

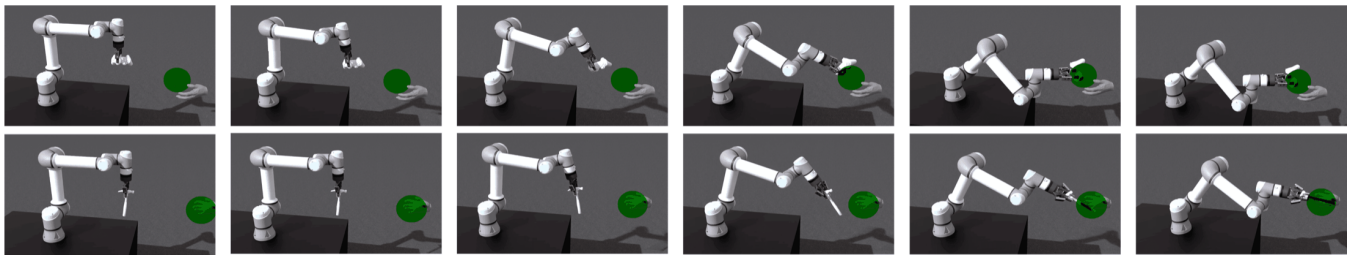
Contact-Handover [3]. This system uses 3D VoxNet [29] to predict contact heatmaps and Contact-GraspNet [30] for grasp candidates. DBSCAN clusters grasps and filters collisions with the MANO hand to rerank the optimal pose. To mitigate computational costs, we perform offline generation using *Open3D*'s *RaycastingScene* for efficient batch collision detection. The handover target is set 15 cm above the estimated palm center, with orientation minimizing collision cost.

Intent-Handover. This method uses Text2HOI [31] to generate hand-object interaction poses. Grasp candidates are evaluated by a joint affordance and safety score, and the highest-scoring candidate is selected as the optimal grasp. We adopt the same handover position strategy as Contact-Handover. The orientation follows the original method: a 15° inward rotation along the predicted palm plane.

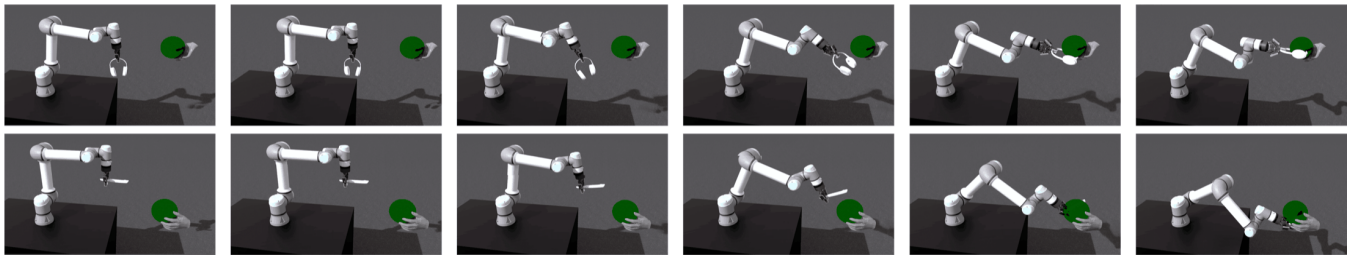
B. Results

Table I reports baseline performance under S0 (unconstrained objects), S1 (functionally constrained objects), and their average. Qualitative examples are shown in Fig. 4.

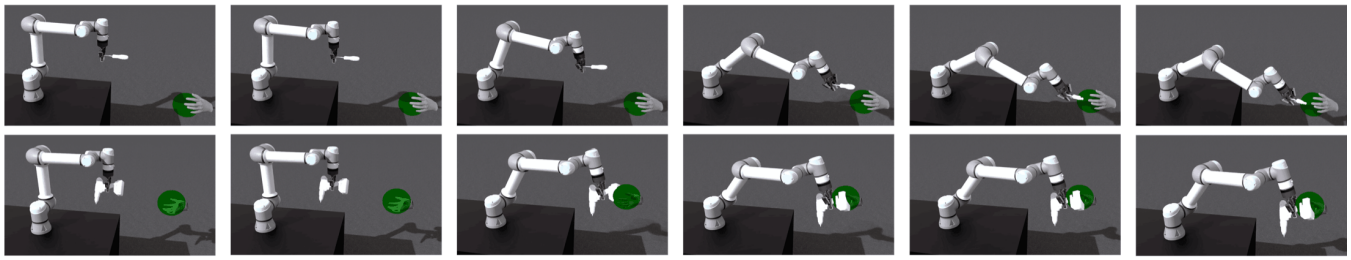
Contact-Handover achieves the highest S0 success rate



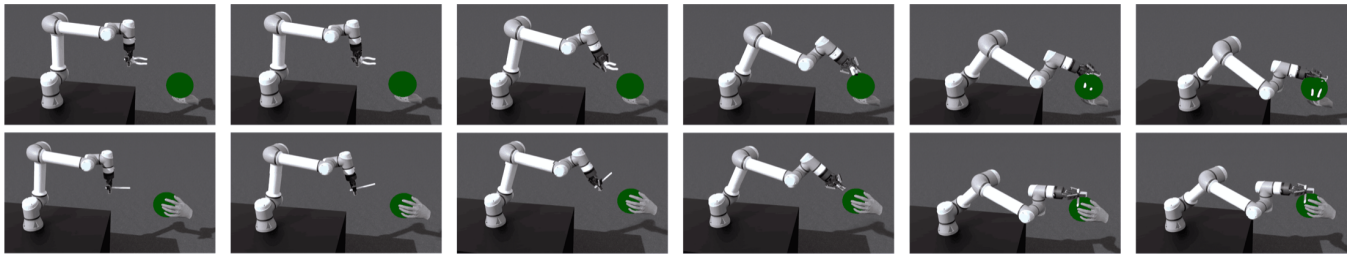
(a) FC-Handover



(b) Handover-VA



(c) Contact-Handover



(d) Intent-Handover

Fig. 4: Qualitative comparison of predicted handover configurations across four baselines. Each row shows the grasp pose, approach trajectory, and final handover pose for the same object–hand pair.

(77.0%) and the lowest average F_{safe} (3.6%). Under S1, all kinematic failure modes worsen: F_{plan} rises from 13.0% to 19.6%, F_{reach} from 3.6% to 8.4%, F_{safe} from 2.2% to 5.0%, and F_{stab} from 4.2% to 7.4%, indicating that the collision-cost objective does not transfer to functional constraints. In S1, F_{afford} is 4.2% (second lowest), suggesting that contact priors from human interaction data implicitly encode functional grasp preferences.

Intent-Handover achieves the highest average success rate (67.9%) and the highest S1 success rate (66.0%). The lowest average F_{plan} (15.6%) and S1 F_{afford} (3.8%) indicate that language-conditioned pose generation steers candidates toward configurations that satisfy functional constraints. T_{plan} averages 2.47 s (2.4× Contact-Handover), and F_{safe} reaches 5.4% in S0, as collision avoidance is applied against a generated hand pose rather than a directly estimated one.

FC-Handover (Avg SR: 59.6%) records the lowest average F_{reach} (3.3%), with F_{reach} rising only from 2.0% to 4.6% across splits, indicating that the placement opposite the predicted hand produces poses within the robot kinematic workspace. In S1, F_{afford} reaches 9.2% (highest) and F_{plan} rises to 23.0%, showing the orientation strategy does not encode which object region should face the receiver.

Handover-VA records the lowest average SR (51.9%) but the shortest T_{plan} (0.68 s) and lowest F_{stab} (3.2%). The highest average F_{plan} (24.5%) and F_{safe} (10.2%), together with the largest S0-to-S1 drop (−25.8 pp), indicate that the orientation heuristic does not account for robot kinematics or collision geometry.

Cross-method analysis. Intent-Handover and Contact-Handover rank first and second in S1 F_{afford} (3.8% and 4.2%), while FC-Handover and Handover-VA reach 9.2% and 6.2%,

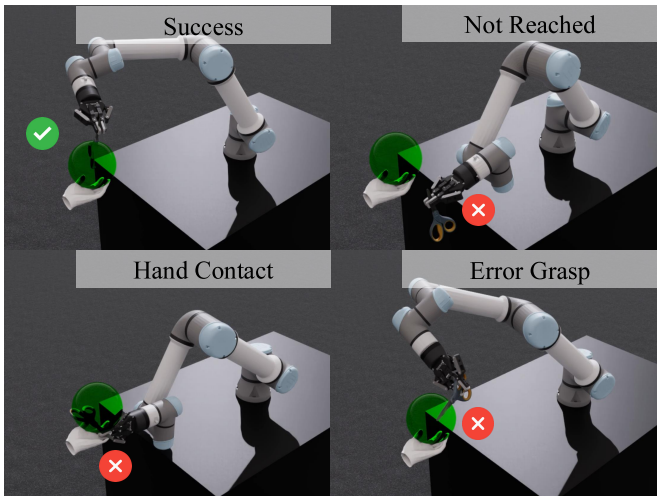


Fig. 5: Representative success and failure cases in simulation. Failures are categorized by planning infeasibility, reachability violation, hand–gripper collision, object drop, and functional-region occlusion.

Fig. 6: Real hardware setup for the sim-to-real experiment. A UR5e robot with a Robotiq 2F-85 gripper delivers objects to a seated human receiver.

confirming that functional awareness (via language or contact priors) reduces affordance failure under constrained objects. Each remaining method leads in one metric: FC-Handover in F_{reach} (3.3%), Contact-Handover in F_{safe} (3.6%), and Handover-VA in F_{stab} (3.2%), while Intent-Handover incurs the highest T_{plan} (2.47 s). These complementary strengths suggest that combining collision-cost ranking with language-guided affordance generation could address the main failure modes in both settings. Fig. 5 illustrates representative success and failure cases.

C. Sim-to-Real Experiment

We deploy all four baselines on the physical platform (Fig. 6) and evaluate each method with 30 participants (5 trials per method per participant, 600 total trials). For each participant, five objects are randomly drawn: 2 from S0 and 3 from S1. The same five objects are used across all four methods (one trial per object), and method order is randomized to support within-subject comparison. For objects whose geometry prevents stable tabletop resting (e.g., knife, game controller), a 3D-printed fixture supports the object and exposes the designated grasp region, allowing the robot arm to approach from a wider range of angles.

Table II reports success rates and total times. Intent-Handover achieves the highest real-world success rate (73.3%, ΔSR = +5.4 pp), followed by FC-Handover (63.3%, ΔSR = +3.7 pp). Handover-VA shows the largest success rate gain (60.0%, ΔSR = +8.1 pp). Contact-Handover is the only method with negative transfer (ΔSR = −9.5 pp, from 66.2% to 56.7%). We attribute this drop to the sensitivity of its voxel-based contact predictions to real-world sensor noise and extrinsic calibration errors, which can partially occlude the receiver grasp region. Contact-Handover was also validated exclusively in simulation, without real-world adaptation.

All methods exhibit increased total time in the real world (ΔT = 3.12–3.48 s), because T_{tot} in simulation measures only robot planning and execution, whereas the real world additionally includes human reception time (visual assessment, wrist reorientation, and grasp closure).

TABLE II: Sim-to-real transfer results. Δ is computed as real-world result minus simulation Avg in Table I.

Methods	**SR**† **(%)**	**Δ**	$T_{\text{tot}}{}^{\dagger}$ **(s)**	**Δ**
FC-Handover	63.3	+3.7	13.52	+3.16
Handover-VA	60.0	**+8.1**	**12.35**	**+3.12**
Contact-Handover	56.7	−9.5	13.14	+3.48
Intent-Handover	**73.3**	+5.4	14.62	+3.32

†Averaged over all trials.

Highlighting: Yellow indicates best real-world performance (higher SR, lower Time). Orange indicates largest SR increase and smallest Time increase.

The other three baselines, each validated or calibrated on physical hardware, all achieve positive transfer. We attribute these gains partly to active receiver adaptation: participants adjusted their wrist orientation and approach angle to accommodate the presented object, a compensatory behavior absent in the static simulation receiver. We additionally observe that object weight in the real world increases object drop failures relative to simulation; even with the support fixture, instability during delivery remains an open challenge.

While modeling the adaptive behavior of human receivers in simulation remains an open challenge, R2HandoverSim nonetheless discriminates effectively among methods: it reveals distinct failure mode profiles across baselines (Table I), correctly identifies the top-performing method in real-world deployment, and exposes divergent sim-to-real transfer characteristics (e.g., Contact-Handover's negative transfer versus positive transfer for the other three methods).

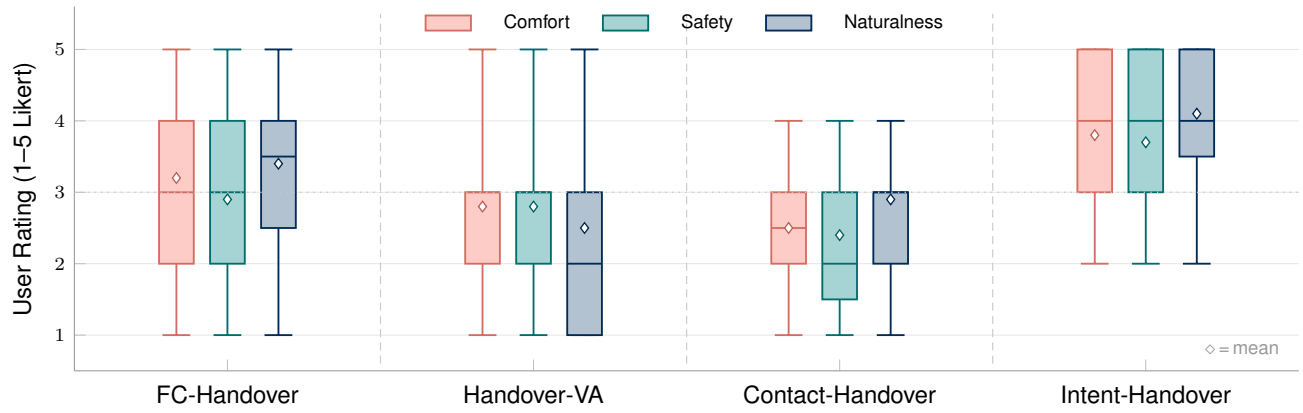


Fig. 7: Subjective ratings for handover comfort, safety, and naturalness ($n = 30$; Likert scale from 1 to 5). The boxes span the interquartile range (Q1 to Q3), the horizontal lines indicate medians, and the diamonds mark means. The whiskers extend to the minimum and maximum ratings. The dotted horizontal line marks the scale midpoint (3).

D. User Study

Simulation success rate captures task completion but not user experience: a technically successful handover may still feel uncomfortable, unsafe, or unnatural to the receiver. We test the hypothesis that *higher simulation success rate tends to predict higher perceived handover quality across comfort, perceived safety, and naturalness.*

Beyond task success, we collect subjective assessments from the same 30 participants. After completing all trials of a given method, each participant rates the handover on three independent 5-point Likert scales: comfort, perceived safety, and naturalness (1 = very poor, 5 = very good). Fig. 7 summarizes the distributions. A Friedman test reveals significant differences among the four methods for all three dimensions: comfort ($\chi^2(3) = 28.4$, $p < 0.001$), perceived safety ($\chi^2(3) = 24.1$, $p < 0.001$), and naturalness ($\chi^2(3) = 31.7$, $p < 0.001$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction ($\alpha = 0.05/6$) confirm that Intent-Handover is rated significantly higher than every other method across all three dimensions ($p < 0.005$ in all pairwise comparisons).

Intent-Handover achieves the highest ratings across all three dimensions. Participants rated it highest in naturalness (median = 4, Q1 = 3.5, mean = 4.1), followed by comfort (mean = 3.8) and perceived safety (mean = 3.7). We attribute this to the language-guided diffusion model consistently presenting objects at orientations that match the expected grasp, reducing the wrist adjustment required before closing the hand. This effect is most pronounced for functionally constrained objects (e.g., scissors, hammers), where task-semantic encoding ensures the handle faces the receiver.

FC-Handover and Handover-VA receive comparable comfort and safety scores. FC-Handover yields mean scores of 3.2 for comfort and 2.9 for safety, whereas Handover-VA averages 2.8 in both metrics. However, they differ substantially in naturalness (FC: median = 3.5, mean = 3.4; VA: median = 2, Q1 = 1, mean = 2.5). The fixed orientation heuristic in Handover-VA frequently presents objects at angles that require wrist rotation to complete the grasp, which we identify as a key factor limiting naturalness. FC-Handover avoids this by predicting a full hand mesh proxy. This approach yields more varied orientations and higher naturalness, despite the method ranking third in simulation success rate (59.6%).

Contact-Handover scores lowest in both comfort (median = 2.5, mean = 2.5) and perceived safety (median = 2, Q3 = 3, mean = 2.4). When the predicted contact map misaligns under sensor noise, the method selects grasps that partially occlude the receiver's intended grasp region, forcing awkward repositioning or causing the object to slip; we note that such repeated failures significantly reduce perceived safety. Even in successful trials, contact-optimized grasps occasionally orient the functional region away from the receiver. Interestingly, naturalness (mean = 2.9) was rated higher than comfort and safety, confirming that grasp pose occlusion rather than object orientation is the primary driver of low ratings.

These results partially refute the stated hypothesis. Intent-Handover ranks first in both simulation SR (67.9%) and user ratings across all three dimensions, consistent with the hypothesis. However, Contact-Handover ranks second in simulation SR (66.2%) yet receives the lowest comfort and safety ratings. Notably, no participant rated its safety above 4. Conversely, FC-Handover ranks third in simulation (59.6%) but second in user preference. This rank inversion demonstrates that simulation success rate alone is not a sufficient predictor of perceived handover quality. Crucially, we observe a divergence between geometric and perceived safety: Contact-Handover achieved the lowest simulation safety failure rate ($F_{\text{safe}} = 3.6\%$), yet users rated it least safe due to intrusive grasping behaviors. In contrast, Intent-Handover's low S1 affordance failure rate ($F_{\text{afford}} = 3.8\%$) strongly correlates with its superior naturalness ratings. This confirms that while standard metrics like SR and F_{safe} capture physical feasibility, higher-level metrics like F_{afford} are better indicators of human-centric handover quality. This validates the effectiveness of R2HandoverSim. By decomposing per-

formance into granular metrics, it successfully identifies the specific factors that drive user preference, offering a more predictive evaluation than success rates alone.

V. Conclusion

We presented R2HandoverSim, a standardized simulation benchmark for robot-to-human handovers. Our evaluation of four baselines demonstrates the trade-offs between geometric stability and human-centric metrics. Real-world experiments with 30 participants confirm that the method achieving the highest simulation success rate (Intent-Handover) also leads in real-world success, and that the benchmark reveals distinct sim-to-real transfer characteristics across methods. Furthermore, our analysis of multiple metrics shows that detailed metrics such as F_{afford} predict perceived handover quality better than success rate alone. Future work will incorporate dynamic hand models and closed-loop receiver policies to further bridge the sim-to-real gap.

References

[1] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulić, "Object handovers: A review for robotics," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1855–1873, 2021.

[2] D. Lehotsky, A. Christensen, and D. Chrysostomou, "Optimizing robot-to-human object handovers using vision-based affordance information," in *2023 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2023, pp. 1–6.

[3] Z. Wang, Z. Liu, N. Ouporov, and S. Song, "Contacthandover: Contact-guided robot-to-human object handover," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 9916–9923.

[4] J. Laplaza, A. Pumarola, F. Moreno-Noguer, and A. Sanfeliu, "Attention deep learning based model for predicting the 3D Human Body Pose using the Robot Human Handover Phases," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 2021, pp. 161–166.

[5] Y.-W. Chao, C. Paxton, Y. Xiang, W. Yang, B. Sundaralingam, T. Chen, A. Murali, M. Cakmak, and D. Fox, "HandoverSim: A simulation framework and benchmark for human-to-robot object handovers," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 6941–6947.

[6] Z. Wang, J. Chen, Z. Chen, P. Xie, R. Chen, and L. Yi, "GenH2R: Learning Generalizable Human-to-Robot Handover via Scalable Simulation Demonstration and Imitation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 362–16 372.

[7] R. Ye, W. Xu, Z. Xue, T. Tang, Y. Wang, and C. Lu, "H2O: A benchmark for visual human-human object handover analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 762–15 771.

[8] O. Taheri, N. Ghorbani, M. J. Black, and D. Tzionas, "GRAB: A dataset of whole-body human grasping of objects," in *European Conference on Computer Vision*. Springer, 2020, pp. 581–600.

[9] Z. Fan, O. Taheri, D. Tzionas, M. Kocabas, M. Kaufmann, M. J. Black, and O. Hilliges, "ARCTIC: A dataset for dexterous bimanual hand-object manipulation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 12 943–12 954.

[10] H. Duan, Y. Yang, D. Li, and P. Wang, "Human–robot object handover: Recent progress and future direction," *Biomimetic Intelligence and Robotics*, vol. 4, no. 1, p. 100145, 2024.

[11] V. Ortenzi, F. Cini, T. Pardi, N. Marturi, R. Stolkin, P. Corke, and M. Controzzi, "The grasp strategy of a robot passer influences performance and quality of the robot-human object handover," *Frontiers in Robotics and AI*, vol. 7, p. 542406, 2020.

[12] J. Liu, W. Dong, J. Wang, and M. Q.-H. Meng, "Leveraging semantic and geometric information for zero-shot robot-to-human handover," *Arxiv Preprint Arxiv:2409.17621*, 2024. [Online]. Available: https://arxiv.org/abs/2409.17621

[13] J. Nowak, P. Fraisse, A. Cherubini, and J.-P. Daures, "Assistance to older adults with comfortable robot-to-human handovers," in *2022 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*, 2022, pp. 1–6.

[14] S. E. Ovur and Y. Demiris, "Naturalistic robot-to-human bimanual handover in complex environments through multi-sensor fusion," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 3730–3741, 2023.

[15] S. Brahmbhatt, C. Tang, C. D. Twigg, C. C. Kemp, and J. Hays, "ContactPose: A dataset of grasps with object contact and hand pose," in *European Conference on Computer Vision*. Springer, 2020, pp. 361–378.

[16] Y.-W. Chao, W. Yang, Y. Xiang, P. Molchanov, A. Handa, J. Tremblay, Y. S. Narang, K. Van Wyk, U. Iqbal, S. Birchfield, and Others, "DexYCB: A benchmark for capturing hand grasping of objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9044–9053.

[17] Z. Wang, Z. Chen, J. Chen, J. Wang, Y. Yang, Y. Liu, X. Liu, H. Wang, and L. Yi, "MobileH2R: Learning generalizable human to mobile robot handover exclusively from scalable and diverse synthetic data," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 17 315–17 325.

[18] Y. Wang, J. Ye, C. Xiao, Y. Zhong, H. Tao, H. Yu, Y. Liu, J. Yu, and Y. Ma, "Dexh2r: A benchmark for dynamic dexterous grasping in human-to-robot handover," *arXiv preprint arXiv:2506.23152*, 2025.

[19] NVIDIA, "NVIDIA Isaac Sim: Robotics simulation and synthetic data," https://developer.nvidia.com/isaac-sim, 2023.

[20] L. Yang, K. Li, X. Zhan, F. Wu, A. Xu, L. Liu, and C. Lu, "Oakink: A large-scale knowledge repository for understanding hand-object interaction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20 953–20 962.

[21] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, "ShapeNet: An information-rich 3d model repository," *arXiv preprint arXiv:1512.03012*, 2015.

[22] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The YCB Object and Model Set: Towards common benchmarks for manipulation research," in *2015 International Conference on Advanced Robotics (ICAR)*. IEEE, 2015, pp. 510–517.

[23] S. Brahmbhatt, C. Ham, C. C. Kemp, and J. Hays, "ContactDB: Analyzing and predicting grasp contact via thermal imaging," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8709–8719.

[24] H. Li, W. Mao, W. Deng, C. Meng, H. Fan, T. Wang, P. Tan, H. Wang, and X. Deng, "Multi-GraspLLM: A multimodal LLM for multi-hand semantic guided grasp generation," Dec. 2024. [Online]. Available: https://arxiv.org/abs/2412.08468

[25] C. Meng, T. Zhang, and T. lun Lam, "Fast and comfortable interactive robot-to-human object handover," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3701–3706.

[26] E. Corona, A. Pumarola, G. Alenya, F. Moreno-Noguer, and G. Rogez, "GanHand: Predicting human grasp affordances in multi-object scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5031–5041.

[27] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics and Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[28] A. D. Christensen, D. Lehotský, M. W. Jørgensen, and D. Chrysostomou, "Learning to segment object affordances on synthetic data for task-oriented robotic handovers," in *The 33rd British Machine Vision Conference*. British Machine Vision Association, 2022.

[29] D. Maturana and S. Scherer, "VoxNet: A 3d convolutional neural network for real-time object recognition," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Ieee, 2015, pp. 922–928.

[30] M. Sundermeyer, A. Mousavian, R. Triebel, and D. Fox, "Contact-Graspnet: Efficient 6-dof grasp generation in cluttered scenes," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 438–13 444.

[31] J. Cha, J. Kim, J. S. Yoon, and S. Baek, "Text2HOI: Text-guided 3d motion generation for hand-object interaction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1577–1585.