Learning End-to-End 6DoF Grasp Choice of Human-to-Robot Handover using Affordance Prediction and Deep Reinforcement Learning

Po-Kai Chang, Jui-Te Huang, Yu-Yen Huang, Hsueh-Cheng Wang*

***Abstract*— Human-to-robot handover is a key capability of service robots and human-robot interaction. Recent work takes advantage of existing hand and object segmentation, and pose estimation algorithms to generate grasps. End-to-end grasping directly from sensor data without object models has made tremendous progress in logistic tasks, but has not been used for human-to-robot handover. However, both approaches aim for grasping without inducing harm, but neither consider which types of grasps may be intrusive to human users. We present our end-to-end 6DoF grasp choice for human-to-robot handover. We first leverage existing end-to-end grasping for the network backbone, and then finetune for preferred grasps using deep reinforcement learning. Comprehensive evaluations are carried out against various baselines using multi-stage hand and object prediction and subsequent planning. We show that the proposed approach was more robust to partial occlusions, and executed human preferred 6DoF grasps without hard-coding the correspondence of hand grasp classification. A dataset of end-to-end grasping and trajectories for human-to-robot handover and all pretrained models are available at `https://arg-nctu.github.io/projects/socially-aware-handover.html`.**

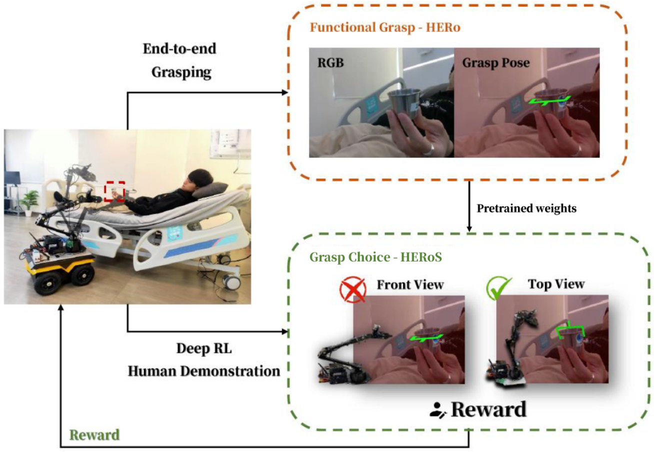

Fig. 1. Our approach allows human-to-robot handovers using end-to-end grasp choice, learning-based algorithms. When the program is executed, the agent will generate an action. Through the end-to-end approach we proposed, there is a 93.5% probability that the target object can be grasped. In addition, we also incorporated reinforcement learning rewards, so that the agent can learn human grasping preferences to make the robot more anthropomorphic.

I. Introduction

Human-to-robot handovers, involving the transfer of an object from a human giver to a robot receiver, are fundamental capabilities that allow for service robots in our daily lives. There has been increasing interest in studying human experiences [1], [2], predicting human hand poses or object affordance for grasping selections [3], [4], and expanding the types of handovers for a wider range of objects [4], [5] in order to improve human-to-robot handover.

Grasp choice may be influenced by several constraints, such as object shape, task requirements, gripper types, and social convention [2]. By observing and analyzing the human grasp, the robot grasp could be adapted accordingly. Therefore, previous works tend to utilize human grasp classification [4] or hand pose estimation, in order to generate a corresponding grasp pose.

Several deep learning-based techniques have been applied in handover pipelines, including object detection [6], and object pose estimation [7].In general, human-to-robot handovers are successful as long as the objects and hands are accurately tracked.

Recent advances in end-to-end grasping prediction have shown promising progress. [8] Used a ResNet-101 backbone end-to-end network to generate the affordance map for either suction or two-finger parallel grippers in heavily cluttered scenarios during the Amazon Robotics Competition. The series of DexNet models [9]–[12] used large amounts of simulation-only datasets with thousands of objects to train the end-to-end CNN networks. Although the above research was successful in overcoming occlusions in cluttered scenarios, this has not yet been used to tackle input images involving objects that are held in human hands, nor has their been consideration of what grasp may be intrusive for users during human-to-robot handovers.

Compared with other existing end-to-end grasping networks [8], [12]–[14], human-to-robot handovers requires the following adaptations. **1) Background:** In previous works, the objects are placed in tote or on tables and therefore the backgrounds are often relatively plain. It has been shown that background may largely affect deep network performance, such as replacing the background by a checkerboard [15]. [5] modified the grasp prediction network GG-CNN [16] by adding a planar surface background. **2) Occlusion:** Given that the target object is held by a human hand, there are a certain number of occlusions that may affect the success rate of segmentation and pose estimation of the object and the hand. In prior end-to-end grasping work, grasp selection networks are usually programmed to pick the objects from the top, and thus are not occluded by other objects in cluttered scenes. Each object is then removed from the tote one by one. In this context, the grasping behavior only has to deal with background clutter, but not occlusion challenges. In contrast, depending on how the object is held in the human's hand, occlusion may be higher and thus existing segmentation and pose estimation algorithms may fail. Therefore, although there are only two objects (a single target object and the human hand) in the input images, the occlusion challenge remains. **3) Planar vs. 6 DoF Grasping:** In prior research, end-to-end grasping is usually formulated as planar grasping [17], whereby a grasping point (x, y) and an angle θ of a two-finger gripper are generated from an input image. However, human-to-robot handovers involve grasping in 3D-space, meaning that there are several grasping and trajectory solutions. **4) Grasp Choice:** Human grasp types are known to associate with different robot grasp during handover [2]. Previous work [4] needs to classify human grasp so that the *hard-coded* robot grasp could be carried out. Learning for such associations in a data-driven manner is not yet studied in previous end-to-end grasping work.

In this paper, we propose to address the problem of grasp poses in human-to-robot handovers through an end-to-end network. The contributions of our work are as follows:

- **An end-to-end grasping approach of affordance prediction for human-to-robot handovers.** Building upon the work in [8], we collected and manually labelled a handover dataset, and trained a model taking RGB-D images as inputs and generating an affordance map for grasping. In contrast to existing methods, this approach bypasses the separate stages of predicting the pose of the hand and the object. This is especially useful when more than 40% of the target object is occluded, and we systematically analyzed the grasping success rate across several baseline methods.
- **A data-driven approach for learning grasp choice without hand pose nor hard-coded human-robot grasp associations.** Given that a grasp choice is affected by object, task, gripper, or social convention constraints, we learn the implicit associations directly from image inputs. We trained grasping behaviors based on human feedback as rewards through deep RL.
- **An open dataset for human-to-robot handover.** The data and pretrained weights of the above end-to-end grasping follow widely used end-to-end grasping in [18]. We also provide the grasp choice of human annotations for deep RL. Both are available open access for reproducing this study.

II. Related Work

A. Hand and Object Pose Estimation

Recent state-of-the-art hand and object pose estimation methods have been developed using a variety of real or simulated datasets. MANO (hand Model with Articulated and Non-rigid defOrmations) [19] was trained on 1,000 high-resolution 3D scans of hands. DOPE (Deep Object Pose Estimation) [20] took advantage of synthetic data for training deep neural networks for YCB household objects. This demonstrated that the combined domain randomized and photorealistic synthetic data closed the reality gap in the context of 6-DoF pose estimation of known objects from a single RGB image. Moreover, ObMan [21] leveraged a large-scale synthetic dataset for joint reconstructions of hands and objects. With manipulation constraints, a model exploited plausible hand-object constellations. We will use these prior methods as our baselines, and systematically evaluate scenarios that include objects with high rates of occlusions.

B. Handovers by Estimating Hand and Object Poses

Prior research on human-to-robot handovers have used establishing hand and object detection and pose estimation algorithms. Yang et al. [3] trained a deep neural network using PointNet++ [7] to classify point clouds around human hands into one of seven pre-defined grasping categories. A subsequent motion plan was then carried out to complete the handovers. Consequently, [4] further tackled the challenges of unseen objects. Closed-loop designs refined the tracking of segmented hands and objects over time. A grasp selection model based on the 6-DoF GraspNet [16] was then performed. Similarly, Rosenberger and colleagues [5] proposed a method for grasping generic objects, using an YOLO V3 object detector [6] trained on 80 object categories from the COCO [22] dataset. They simultaneously predicted hand and body segmentation, which were then excluded from a modified GG-CNN [14] model to generate safe grasps. Our work is different from this prior work, incorporating a data-driven end-to-end grasping model that does not require hand and object detection and pose estimation algorithms.

C. End-to-end Grasping

End-to-end grasping has attracted much attention. [17] used a real robot to collect a dataset of 50k grasps in a self-supervised fashion, and trained a deep neural network classifier to predict grasp success. [23] further collected a grasping dataset with more variability by using images from individual's homes instead of only in lab settings using a low cost robotic platform. DexNet [9]–[12] is a series of end-to-end approaches that used a 6.7 million dataset which was generated entirely through simulation. The proposed GQ-CNN method evaluated the quality of each grasp configuration from the previous step [0,1], and outputs the highest quality grasp configuration. In contrast, [24] used human annotations for suction grasping and parallel grasping. They used a ResNet-101 backbone network to implement an end-to-end affordance prediction method, without any preprocessing of object segmentation and classification.

Among these methods, self-supervised learning may not be suitable for handovers which include humans. Although simulation offers an abundance of data, it has been known to fall short in certain application setups due to the domain gap between synthetic and real data. How to incorporate human knowledge into the grasping algorithms remains an open question. In our work, the affordance prediction method [24] for human-to-robot handovers was adapted using human annotations.

III. Problem Formulation

A. End-to-end Grasping using Affordance Prediction

We follow the affordance prediction [24] formulated as planar grasping. During pre-processing we filter backgrounds that have a depth greater than 75 cm. Given RGB-D images (I_{RGB} and depth I_{Depth}) of the scene, a fully convolutional network is trained to infer the affordances ($I_{Affordance}$) across a dense pixel-wise sampling of end-effector orientations and locations. Each pixel correlates to a different position by which to execute the grasping. By rotating the same input and running inference for N times, where N was set to 8 for 0, 30, 45, 60, 90, 120, 135, and 150 degrees, the largest area among the $224 \times 224 \times N$ affordance map is selected as the grasp by a two-finger parallel gripper. We also map the grasp pose ($^{O}X^{G_{Grasp}}$) from the camera frame to a world frame ($^{W}P^{P_i} = {}^{W}X^{C}\ {}^{C}P^{P_i}$). We then use MoveIt [25], open-loop motion planning in real-time, to execute the grasping.

B. Learning Grasp Choice using Deep RL

Through pixel-wise affordance prediction, we can predict the planar graspable positions and gripper angles. However, grasp choice may be influenced by object or human givers' different preferences when interacting with the robot. We modify the training process to include the human in the loop through deep reinforcement learning.

We formulate our task as a standard Markov Decision Process (MDP), which is defined as $M = \{S, A, R, P, \gamma\}$. We reuse the pre-trained weight obtained in our affordance prediction method, which share the RGB and depth image inputs as the state S. The action $A \in R^{224\times224\times N\times M}$ is a discrete space that maps to the 224×224 grasping position, N as different orientations of the end effector, and M as the number of 6DoF grasping. R is the reward signal, and P is the transition function to the next state, $p(s_{t+1}|s_t, a_t)$, finally the γ is a discount factor for future value estimation. The objective is to find a state-action value estimator using rewards as a guide. Here we transform our value estimator into $N \times M$ affordance maps, and the largest area is selected for the 6DoF grasping task.

IV. Proposed Methods

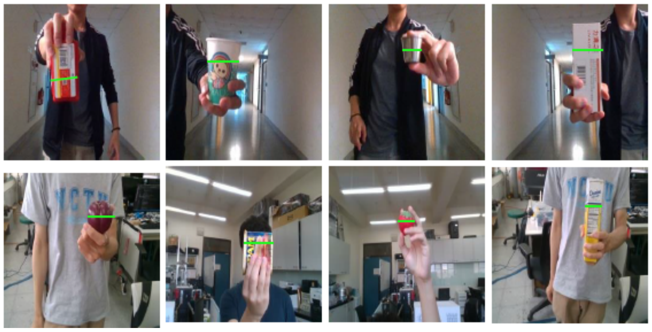

Fig. 2. Handover Dataset: line annotation example.

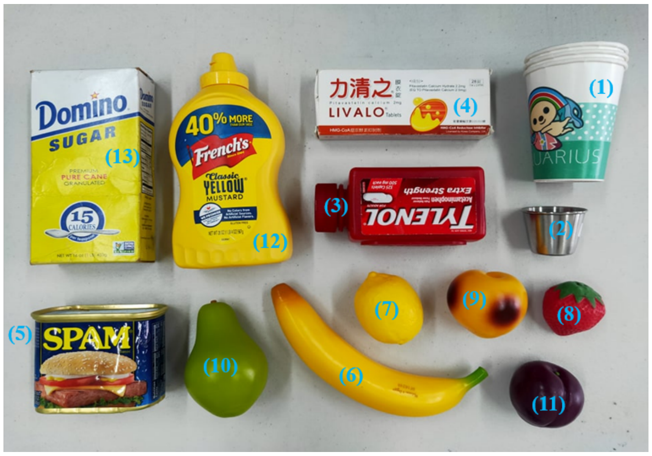

Fig. 3. Handover Dataset: (1) water cup, (2) small medicine cup, (3) medicine bottle, (4) medicine box, (5) SPAM, (6) banana, (7) lemon, (8) strawberry, (9) peach, (10) pear, (11) plum, (12) mustard, (13) sugar. Fruit objects were plastic.

TABLE I
Comparison table of three datasets of HERoS. HERoS-Af: subset for predicting affordance; HERoS-Ch: subset for learning grasp choice; HERoS-Tr: subset for studying trajectories of human demonstrations.

	HERoS-Af	HERoS-Ch	HERoS-Tr
Type	RGB-D	RGB-D	RGB-D, Joint-state
Frame	1,368	1,127	26,693
Label	Pixel-wise	Reward	Sparse Reward
Note	18,868 labels as grasps	1,127 feedback as reward	1000 trajectories from human demo

A. The HERoS Dataset

We collected a dataset including 9 YCB [26] objects and 4 additional objects — water cup, small medicine cup, medicine box, and a medicine bottle, shown in Fig. 3. We used the Intel D435 RGBD camera mounted at the end-effector of ViperX300s on Jackal UGV. We parsed the recorded ROS bag files into RGB images and depth images, which were 640×480 in resolutions. We collected data in different scenes and angles of light sources, and held objects in various postures with both the left and right hands.

1) Human Labelling for Affordance: We labelled the RGB images using the LabelMe [27] tool. The output size is also a 640×480 densely labeled pixel-wise map. Each pixel value of the pixel-wise map was normalized to between 0 and 1 in the form of a heat map. We followed the dataset in [8] and asked human annotators to label a line of where a two-finger gripper could grasp the object without touching the human hand. Similar to the labelling in [8], one object was densely labelled with the graspable positions and angles displayed in green, shown in Fig. 2. Each object included around 100 images, resulting in 1,368 RGB-D images; there were 1,368 annotations with 18,868 possible grasps (green lines).

2) Human Feedback for Grasp Choice: We collected an offline training replay buffer that contained medical items and 9 additional YCB [26] objects that were shown in Fig. 3

, a total of 13 objects with 1127 transitions, whereby a transition can be defined by a tuple $\{S_t,\ S_{t+1},\ a_t,\ r,\ E\}$ of the initial state, the state after an action (each state is a pair of RGB and Depth images), the action (position and orientation), reward and an indicator of success. We labelled the indicator of success as either True or False and the reward as either 5 or -5, depending on whether the transition is a successful grasp or not.

We hold the objects at four angles (90, 0, -45, 45 degrees) to the horizontal ground for our model to predict an action based on this. If the transition is successful, the states are one pair of images that contain the held object and another without. If the transition fails, the states are two pairs of images which both have the held object. There were 1127 labelled image/action pairs.

We also collected a subset of human demonstration trajectories and evaluate if grasp choice affects handover trajectories.

B. *End-to-end Grasping by Affordance Prediction*

In order to accomplish the handover task, we needed to obtain the 3D position of the grasp point, which is defined as the point where the object can be stably grasped without touching the human's fingers. Through the labelled dataset, the model learned whether each pixel of the RGB-D image is a point that can be grasped.

1) Model Architecture: In this paper, we used ResNet-101 [28] as the main network architecture of HERoS and used residual networks to solve the degradation of deep networks and reduce the amount of computation. Although this dataset is considered small for training a deep network, using ResNet pre-trained on ImageNet [29] is sufficient for finetuning our architecture. The architecture is a dual-stream network. The RGB image (RGB, 3 channels) and the depth image through cloning are normalized by subtracting the mean and dividing by the standard deviation (DDD, 3 channels) and are sent into the ResNet-101 network separately. The depth is cloned across channels to use ResNets pre-trained weights from ImageNet on 3-channel (RGB) color images to avoid non-convergence due to the small dataset. Finally, we concatenated the two ResNet-101 outputs (RGB and depth), followed by 3 additional spatial convolution layers to merge the features. We then spatially up-sampled the outputs bilinearly and soft- maxed to output two pixel-wise layers (non-grasping and graspable layers) to represent the inferred affordances.

2) Model Training: We post-processed the captured depth information to make the background such that depths greater than 75cm have the same depth value and become a flat surface to reduce noise and increase prediction accuracy.

We flipped the RGB images, depth images, and labelled images horizontally and vertically to increase diversity. We used cross-entropy for the loss function, a batch size of 10, a fixed learning rate of 10^{-3}, and a momentum of 0.99 to train HERoS through stochastic gradient descent. Our model was trained in PyTorch using NVIDIA RTX2080 on Intel Core i5-9400F. Total training time took approximately one hour.

3) Motion Planning: This module translates the output of the target pose into joint and gripper actions by running the Moveit [25] path planning node. This process is mainly based on the best grasp pose generated by HERoS. When the manipulator VX300s reaches the 5cm in front of the target grasp point, the velocity of the manipulator will become $\frac{1}{5}$ the normal speed, and it will slowly approach the target point. This action is to protect the safety of the human interactors during handover and to ensure that s/he will not change the position of the original object or change the grasp pose due to fear of being touched by the manipulator. After reaching the target point, the gripper will close and detect whether the object has been grasped correctly through the gripper position.

C. *Learning Grasp Choice using Deep RL*

However, there is no available simulator for human-robot hand-over tasks to boot-strap the training of deep RL. Also, directly training the deep RL agent from scratch in the real world to interact with humans is inefficient and has the potential to cause physical injury. Therefore, we used the collected human demonstration replay buffer with an offline reinforcement learning algorithm to train another affordance prediction model.

1) Deep RL Value Prediction: The algorithm we chose was Double Q-learning (DDQN) [30], which is an off-policy learning algorithm that can work with our pre-stored experience replay buffer. The algorithm was designed to train a Q-network with parameter θ to predict the future value of each action through the reward signal. DDQN leverages a separate target network with parameter $\theta^{'}$ that gradually update from θ to stabilize training process. We also use a prioritized experience replay buffer to increase efficiency by sampling important transitions to update the gradient. The objective of the value network, Q, was to minimize the Bellman error

$$R_t + \gamma Q_{\theta'}(s_{t+1}, \underset{a}{\operatorname{argmax}} Q_\theta(s_{t+1}, a)) - Q_\theta(s_t, a_t)$$

Our Q-network was based on a previous affordance prediction network, with only one additional convolution layer to output the Q-value. We also rotated the input images according to every possible end-effector orientation to obtain the Q-values for each possible action in the action space. The full value map can be used as an affordance map to select the best grasping point and rotation. The training process was on a NVIDIA RTX2070 GPU for 5000 steps which took 3 hours.

V. Human-to-robot Handover Experiment

In this section, we demonstrate the effectiveness of our approach with experiments. We compare the concepts and effects of various human-to-robot handovers, including the methods we propose.

A. *Experimental Setup*

We used a mobile manipulator platform, and put the manipulator system in a fixed area, prohibiting moving vehicles

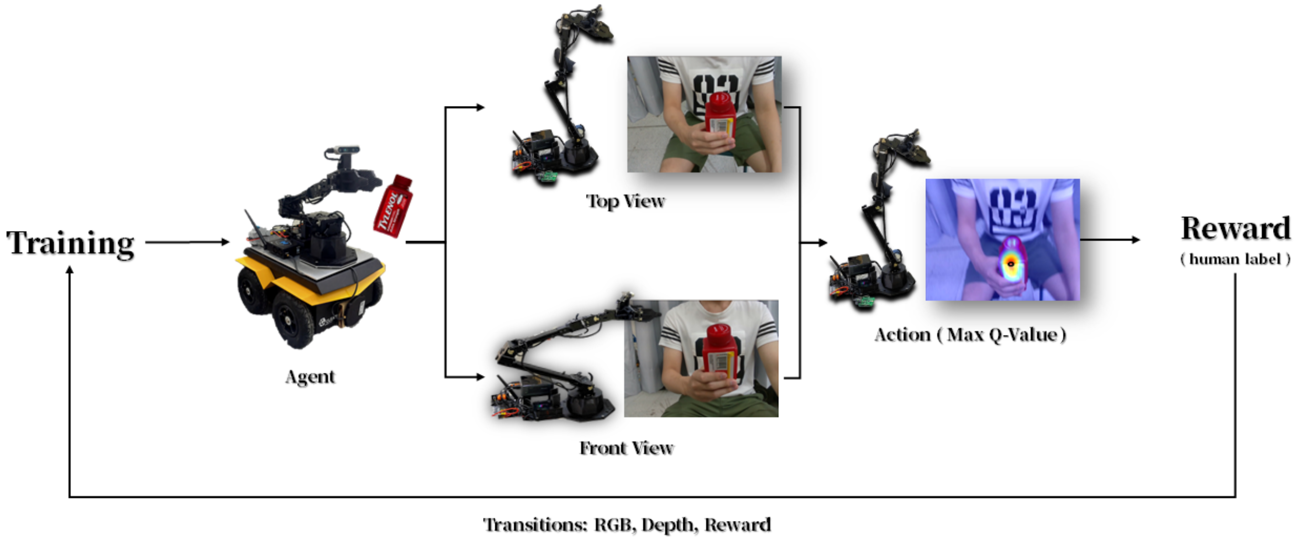

Fig. 4. Grasp choice system architecture diagram. When the image is received, it will control the manipulator from two different perspectives for prediction, selecting the perspective with a larger Q-value to perform the grasping, and receives the reward through the human label after the grasping is completed, so that the agent can learn human preferences.

to increase the fairness of the experiment. Testers arbitrarily handed over objects to the robot. During the experiment, we tested the following 5 approaches, conducting 20 trials for each object, and recorded the number of successes, and the reason for failures.

Baseline I. Hand [19] + Object Poses [20] (DOPE & MANO): This approach represented the intuitive method of handovers, which predicted the pose of the object and the human hand separately, where the algorithm excluded the hand to grasp the target. The object pose prediction component used the state-of-the-art DOPE model [20], and the hand prediction component used the MANO model [19]. We combined the two and perform post-processing, and finally selecting the grasp pose for the handover task.

Baseline II. Joint Reconstruction [21] (ObMan): As prediction of the object and hand separately led to occlusion problems, we tried to use the ObMan [21] model to directly generate the pose of the object and the hand through joint reconstruction in order to obtain their relative positions. The reconstructed result was mapped to real-world coordinates by coherent point drift methods [31], and the grasp pose was generated through the algorithm to complete the handover task.

Baseline III. Modified GG-CNN [5], [14] (Rosenberger et al.): The approach described in Section. II was similar to the goal we want to achieve. We used this method as the baseline for this experiment and applied it to the D435 Camera and ViperX300s manipulator.

HERoS (Current approach): The proposed system called **HERoS** was described in Section. IV, used RGB-D images to predict the position and orientation that can be grasped.

Teleoperation: In order to compare the success rate of the manual control and the autonomously grasping methods, we reported the results using a joystick to control the manipulator to complete this task.

All approaches used the same process for evaluation. During the experimental handover task, we performed 20 trials for each object, divided into four orientations: vertical, horizontal, diagonally to the right, and diagonally to the left. We performed 5 trials in each orientation using random poses.

B. Evaluation Metrics

We used a set of indicators to evaluate the performance of the system and analyze it by recording the number of successes, failures and their reasons.

Hand Occlusion:

To ensure fairness of the comparison of various approaches and quantitative data analysis, we captured the current frame of the grasp execution and human label to obtain the occlusion relationship between the hand and the object. We categorized hand occlusion into either $< 40\%$ or $> 40\%$.

$$occlusion = \frac{(object\ area\ \cap\ hand\ area)}{object\ area}$$

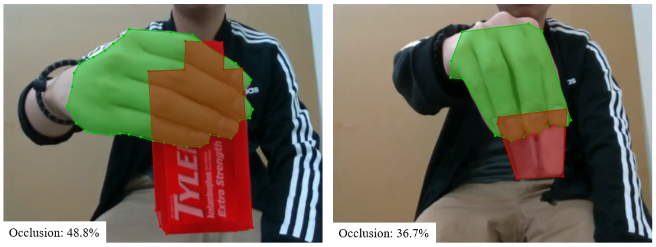

Fig. 5. Example capture of the current frame of the grasp execution for labelling hand/object segmentation and estimating the occlusion ratio.

Success: How often the robot was able to successfully take the object from the human's hand safely.

Planning Fail: The robot failed to predict the object in the human hand, resulting in an inability to grasp the object but the human's fingers were not touched.

Touched Fingers: Regardless of whether an object is grasped or not, if the robot grasps the human fingers, it will be considered a failure and recorded.

C. Results

We first compared with Baseline III. (Rosenberger et al. [5]), which used a modified GG-CNN for grasping. 9 YCB objects with $< 40\%$ occlusion were evaluated, and we replicated the success rate reported in [5] using our manipulation platform. Our proposed affordance prediction method achieved higher success rate, shown in Table. II.

In a separate experiment, Table. III further showed the results of each of the main metrics in our system evaluation process. We first reported that our hardware equipment is sufficient to complete the handover task by teleoperation. The success rate of other three benchmark methods (DOPE & MANO, ObMan, and Rosenberger et al. [5]) were found decreased due to hand occlusions, but our method consistently outperformed the success rate, suggesting the efficacy and reliability of our method to severe hand occlusion ($> 40\%$).

The overall success rate of our method for all objects was 94%. It is noted that the success rate in Baseline III decreased to 63%. This may be caused by the model for detecting objects uses a YOLO V3 object detector [6], which was prone to false positives with a complex background and occlusion problems.

In the Baseline I. (DOPE & MANO) experiment, this method was used to predict the positions of the objects and fingers separately and then performed post-processing calculations. Common problems were that the hand occludes the objects and results in pose estimation errors. As the object cannot be detected if there was an occlusion, the DOPE model was only able to successfully predict the pose of the object during the experiment when the fingers are placed on the sides of the object.

Compared with the Baseline II. (ObMan) approach, the method of using hands and objects for joint reconstruction deals with the problem of not being able to detect objects. However, during post-processing, the reconstructed 3D-point cloud needs to be projected onto objects in real world coordinates to obtain the grasp points, which often causes errors such as skew.

VI. Grasp Choice Experiment

In the previous experiment V, we showed that the target object can be successfully grasped, but some successful grasp may violate some grasp choice constraints (such as social convention). In this experiment, we further used deep RL to learn the associations of raw image inputs and grasp choice of human preferences. We set up our robot arm using the model described in Section.IV-C and continuously updated the model through the guiding of rewards given by users.

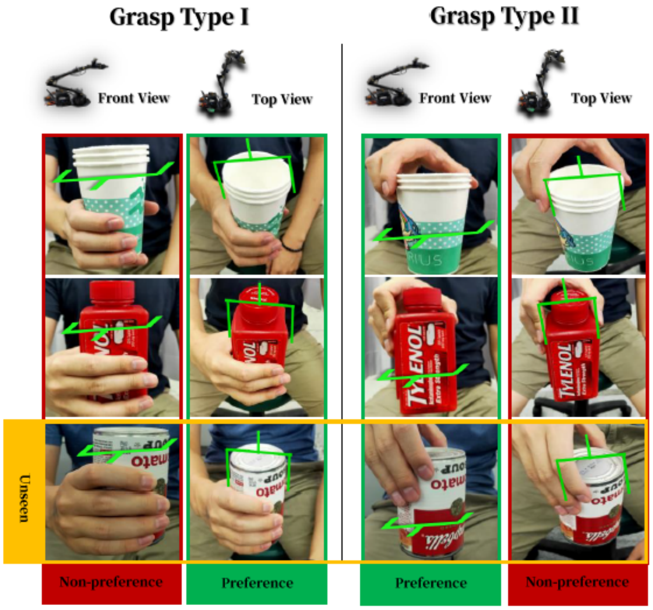

Fig. 6. There are Grasp Type I and Grasp Type II preferences for picking objects. Grasp Type I is picking the middle of objects, and Grasp Type II is picking the top of objects, using different perspectives to predict objects, and finally selecting the grasping perspective preferred by humans.

A. Experimental Setup

We designed fined two grasp types that were suitable for one of the two 6DoF grasp choices (front and top views), as shown in Fig. 6. The robot arm moved to the initial positions of the two grasp choices positions, and collected two input RGB-D images. The inputs were then pre-processed with 4 planar rotations, resulting in $224 \times 224 \times 4 \times 2$ before feeding into the deep model. The model predicted the Q-values of the outputed 2×4 6DoF affordance map, and the best grasping choice and orientation were selected. It is often to have a successful grasp from both grasp choices. However, we have set preferences for the robot to learn. We programmed the model to learn that When the user was holding the bottom of object, the robot should approach from the top to stay away from the user's finger's. Noted that no human hand detector nor human grasp classifier were used.

TABLE II

INDIVIDUAL OBJECT GRASPING RESULTS USING YCB DATASET IN COMPARISON WITH MODIFED GG-CNN [5], [14] BASELINE.

	Hand Occlusion <40%									Avg.
Method \Object	SPAM	Banana	Lemon	Strawberry	Peach	Pear	Plum	Mustard	Sugar	Success Rate
Baseline III. Modified GG-CNN	80%	75%	75%	90%	70%	85%	80%	85%	80%	80%
HERoS (Current approach)	**95%**	**85%**	**80%**	**95%**	**90%**	**90%**	**90%**	**95%**	80%	**88.89%**

TABLE III

COMPARISON OF THE SUCCESS RATE OF GRASPING FOUR OBJECTS ACROSS APPROACHES.

Method \Hand Occlusion	<40%	≥ 40%
Baseline I. Hand [19] + Object Poses [20]	62%	45%
Baseline II. Joint Reconstruction [21]	75%	61%
Baseline III. Modified GG-CNN [5], [14]	63%	50%
HERoS (Current approach)	**94%**	**94%**
Teleop	96%	91%

B. Training-stage

The deep RL model was trained while the robot interacted with human users, and a total number of 100 handovers was carried out. During each handover the reward given by the user. For every trial where the handover task was completed, it sampled from the prioritized reply buffer for training. 6-10 points were given when the grasping pose matched the human preference, 1-5 points were given when the pose grasped the object but did not match the human preference, and -5 points was given if the grasping fails or collision with a finger.

C. Results for Known and Novel Objects

We carried out a pre-test and a post-test to evaluate the grasp choice of the pre-trained DRL model before and after the training with 100 handovers. We had 40 handovers interacting with the robot arm, but without updating the weights. All trials usded known objects shown in the training stage. As mentioned the model generated 4 orientations × 2 grasp choice, and the best grasp position and orientation among the 8 affordance maps were selected. According to the experimental results, we found that the overall success grasp achieve 82.5%, and the probability for the preference was 47.5%.

The same procedure was performed for the model trained with deep RL. We tested 40 handovers for known objects (shown in the training stage) and 20 handovers for novel objects, which were not included in any of our training dataset HERoS. After obtaining rewards from humans and training for 100 trials, the average probability of grasping using the human preference was 82.5%. This means that the model can be updated with human-preferred grasping through learning to achieve a more preferred grasp choice, but a trade-off was found that the success rate (based on predicted positions and orientations from the affordance map) slightly decreased to 73.33%.

VII. CONCLUSIONS

Human-to-robot handover has many challenges. Considering the issues of complex backgrounds, occlusion and trajectory, we presented an end-to-end approach for handover. We filtered out depth noise to increase the robustness of the model. Without using separate predictions of the hands and objects, we used RGB-D images as input to generate affordance maps for selecting the grasp point, thus solving the occlusion problem. Then we used deep reinforcement learning to train our prediction model. After each grasping trial, the user gives a point reward, allowing the robot to learn the grasp choice preferred by humans.

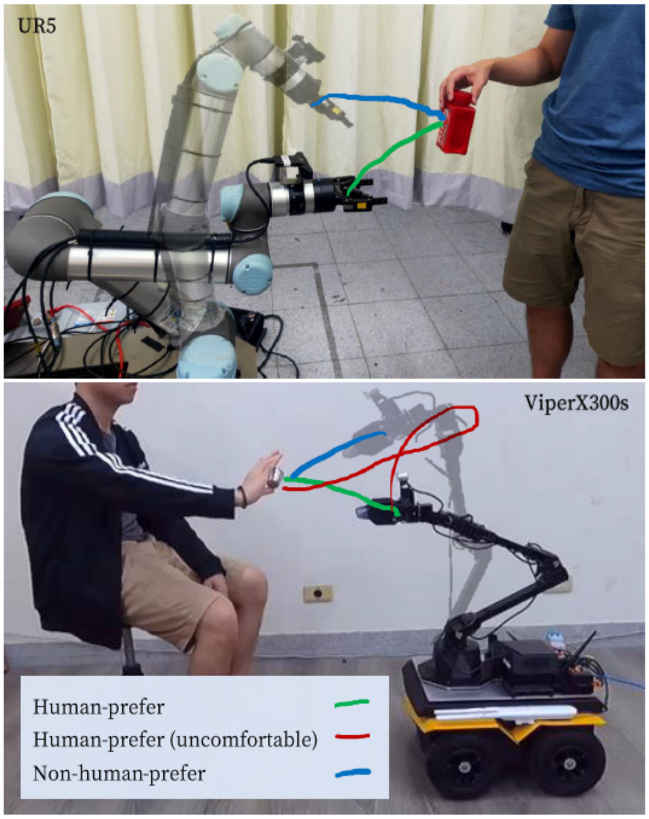

Fig. 7. The trajectory of the manipulator during the handover task. The green line is the human-prefer trajectory, the red line is the human-prefer but not comfortable trajectory, and the blue line is the non-human-prefer trajectory.

In the future, we believe the same approach could also be applied to many applications of human-robot collaboration. In future work, we will 1) use off-line RL training to jointly train affordance map with preference, and 2) use end-to-end methods to modify the deep RL network architecture and generate a close-loop 6D pose using RGB images, which can make the manipulator trajectory more legible and friendly.

Acknowledgments

The research was supported by Taiwan's Ministry of Science and Technology (grants 109-2321-B-009-006, 109-2224-E-007-004, and 109-2221-E-009-074). This work was funded in part by Qualcomm through the Taiwan University Research Collaboration Project.

References

[1] C.-M. Huang, M. Cakmak, and B. Mutlu, "Adaptive coordination strategies for human-robot handovers." in *Robotics: science and systems*, vol. 11. Rome, Italy, 2015.

[2] F. Cini, V. Ortenzi, P. Corke, and M. Controzzi, "On the choice of grasp type and location when handing over an object," *Science Robotics*, vol. 4, 2019.

[3] W. Yang, C. Paxton, M. Cakmak, and D. Fox, "Human grasp classification for reactive human-to-robot handovers," *arXiv preprint arXiv:2003.06000*, 2020.

[4] W. Yang, C. Paxton, A. Mousavian, Y.-W. Chao, M. Cakmak, and D. Fox, "Reactive human-to-robot handovers of arbitrary objects," *arXiv preprint arXiv:2011.08961*, 2020.

[5] P. Rosenberger, A. Cosgun, R. Newbury, J. Kwan, V. Ortenzi, P. Corke, and M. Grafinger, "Object-independent human-to-robot handovers using real time robotic vision," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 17–23, 2020.

[6] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[7] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *arXiv preprint arXiv:1706.02413*, 2017.

[8] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, *et al.*, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 3750–3757.

[9] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 1957–1964.

[10] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *arXiv preprint arXiv:1703.09312*, 2017.

[11] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg, "Dex-net 3.0: Computing robust robot suction grasp targets in point clouds using a new analytic model and deep learning," *arXiv preprint arXiv:1709.06670*, 2017.

[12] J. Mahler, M. Matl, V. Satish, M. Danielczuk, B. DeRose, S. McKinley, and K. Goldberg, "Learning ambidextrous robot grasping policies," *Science Robotics*, vol. 4, no. 26, p. eaau4984, 2019.

[13] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.

[14] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," *arXiv preprint arXiv:1804.05172*, 2018.

[15] R. Julian, B. Swanson, G. S. Sukhatme, S. Levine, C. Finn, and K. Hausman, "Never stop learning: The effectiveness of fine-tuning in robotic reinforcement learning," *arXiv preprint arXiv:2004.10190*, 2020.

[16] A. Mousavian, C. Eppner, and D. Fox, "6-dof graspnet: Variational grasp generation for object manipulation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2901–2910.

[17] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 3406–3413.

[18] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[19] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: Modeling and capturing hands and bodies together," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 6, pp. 1–17, 2017.

[20] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," in *Conference on Robot Learning (CoRL)*, 2018. [Online]. Available: https://arxiv.org/abs/1809.10790

[21] Y. Hasson, G. Varol, D. Tzionas, I. Kalevatykh, M. J. Black, I. Laptev, and C. Schmid, "Learning joint reconstruction of hands and manipulated objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 807–11 816.

[22] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[23] A. Gupta, A. Murali, D. P. Gandhi, and L. Pinto, "Robot learning in homes: Improving generalization and reducing dataset bias," in *Advances in Neural Information Processing Systems*, 2018, pp. 9112–9122.

[24] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, N. Fazeli, F. Alet, N. C. Dafle, R. Holladay, I. Morona, P. Q. Nair, D. Green, I. Taylor, W. Liu, T. Funkhouser, and A. Rodriguez, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," *The International Journal of Robotics Research*, 2019. [Online]. Available: https://doi.org/10.1177/0278364919868017

[25] S. Chitta, I. Sucan, and S. Cousins, "Moveit! [ros topics]," *IEEE Robotics Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.

[26] B. Calli, A. Singh, J. Bruce, A. Walsman, K. Konolige, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Yale-cmu-berkeley dataset for robotic manipulation research," *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 261–268, 2017.

[27] K. Wada, "labelme: Image Polygonal Annotation with Python," https://github.com/wkentaro/labelme, 2016.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[30] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1, 2016.

[31] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.