

# Physically Consistent Preferential Bayesian Optimization for Food Arrangement

Yuhwan Kwon<sup>1</sup>, Yoshihisa Tsurumine<sup>1</sup>, Takeshi Shimmura<sup>2,3</sup>,  
Sadao Kawamura<sup>4</sup>, and Takamitsu Matsubara<sup>1</sup>

**Abstract**—This paper considers the problem of estimating a preferred food arrangement for users from interactive pairwise comparisons using Computer Graphics (CG)-based dish images. As a foodservice industry requirement, we need to utilize domain rules for the geometry of the arrangement (e.g., the food layout of some Japanese dishes is reminiscent of mountains). However, those rules are qualitative and ambiguous; the estimated result might be physically inconsistent (e.g., each food physically interferes, and the arrangement becomes infeasible). To cope with this problem, we propose Physically Consistent Preferential Bayesian Optimization (PCPBO) as a method that obtains physically feasible and preferred arrangements that satisfy domain rules. PCPBO employs a bi-level optimization that combines a physical simulation-based optimization and a Preference-based Bayesian Optimization (PbBO). Our experimental results demonstrated the effectiveness of PCPBO on simulated and actual human users.

## I. INTRODUCTION

In recent years, much study has concentrated on using robots in the food industry, especially research on manipulating ingredients and foods [1]–[4]. However, most previous studies have focused on robustness [2], [3] and accuracy [5] of the food manipulation; only a few have addressed eye appeal [4]. It may be of great value for robots in the foodservice industry to serve dishes with users’ preferred food arrangement by providing the estimated preferred arrangement to the grasping and manipulation planners (as in [5]) that explicitly require target arrangement. Therefore, we focus on the estimation problem of a preferred food arrangement in this paper.

Comparing pairs of alternatives is intuitively more appealing to humans in determining their preferred one than evaluating them individually, supported by evidence from cognitive psychology [6]. Preference-based Bayesian Optimization (PbBO) [7], [8] is one of the methods to estimate preferences efficiently while generating such a two-choice query interactively. Hence, this paper considers the estimation problem of users’ preferred food arrangement by PbBO with Computer Graphics (CG)-based dish images. The use of

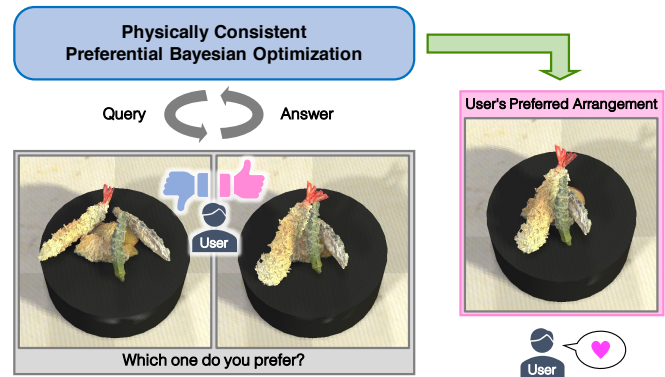


Fig. 1: Rough sketch of preferred food arrangement estimation by proposed Physically Consistent Preferential Bayesian Optimization, which estimates a preferred arrangement that satisfies both physical laws and domain rules through interactive pairwise comparisons.

CG is naturally motivated because it is difficult to generate those many images for comparison in a real environment [2].

As a foodservice industry requirement, we need to utilize domain rules about the geometry of the arrangement as used by chefs [9]. However, those rules are qualitative and ambiguous; the arrangements based on the rules can be physically inconsistent (e.g., physical interference of foods on CG makes the arrangement physically infeasible). For example, in Japanese dishes as tempura, the rule “place the foods to resemble mountains” is essential [9]. However, no specific instructions are given on how to place them without collapsing. Therefore, for robotic automation purposes, it becomes crucial to estimate the preferred arrangement that is consistent with both such rules and physical constraints.

Since the domain rules are geometric, they can be easily considered in the food image generation so that PbBO can estimate the food arrangement that satisfies the domain rules and user preferences. However, physical consistency cannot be considered simultaneously in the same manner due to its complexity. To overcome this issue, we utilize physics simulation-based optimization; by generating dish images that satisfy both domain rules and physical consistency through the optimization, we could estimate user-preferred food arrangements that meet both of them.

In this paper, we propose Physically Consistent Preferential Bayesian Optimization (PCPBO) as a method that estimates the preferred food arrangement satisfying both physical laws and domain rules. For its estimation, PCPBO employs a bi-level optimization that combines a physical simulation-based optimization and PbBO to generate two-

<sup>1</sup>YK, YT, and TM are with the Department of Science and Technology, Graduate School of Science and Technology, Nara Institute of Science and Technology, Nara, Japan.

<sup>2</sup>TS is with the Ganko Food Service Co., Ltd., Osaka, Japan.

<sup>3</sup>TS is also with the College of Gastronomy Management, Ritsumeikan University, Shiga, Japan.

<sup>4</sup>SK is with the Department of Robotics, Graduate School of Science and Engineering, Ritsumeikan University, Shiga, Japan.

This work was supported by the Cabinet Office (CAO), Cross-ministerial Strategic Innovation Promotion Program (SIP), “An intelligent knowledge processing infrastructure, integrating physical and virtual domains” (funding agency: NEDO).

choice arrangements consistent with them. Figure 1 shows how PCPBO estimates it through interactive pairwise comparisons with a user. We also developed an environment of physical simulation with CG on Japanese food arrangement tasks and experimentally evaluated our proposed method. Our experimental results on the tasks demonstrated the effectiveness of PCPBO for estimating physically consistent and users’ preferred arrangements to simulated and actual human users.

The following are this paper’s main contributions:

- To our knowledge, we proposed the first framework for estimating the preferred food arrangement that considers the geometric rules and physical constraints essential for robotic food arrangement in the foodservice industry.
- We formulated PCPBO as the solution method to the above problem with bi-level optimization.
- We conducted experiments on tasks that reflect the rules of Japanese cuisine for real and simulated users.

## II. RELATED WORK

### A. Studies for Manipulating Foods

Many studies have addressed food recognition [10] and handling mechanisms [1], although few previous studies focused on food-manipulation planning. Matsuoka *et al.* focused on the arrangement errors and environmental changes in the arrangement of tempura, a popular Japanese dish, and proposed a food arrangement planning method that is robust to such errors [2]. Takahashi *et al.* proposed a method that utilizes the uncertainty of deep learning models obtained through self-supervised learning to grasp granular ingredients from a tray as accurately as possible [3]. While those studies focused on robustness and accuracy, other studies focused on preference. Endo *et al.* developed a gripper that arranges noodles and chopped ingredients in a conical shape to make them more appealing and verified their performance under several conditions [4]. Junge *et al.* proposed a method using Bayesian optimization (BO) to optimize the taste and the preference of omelets cooked by a robot [11]. Our study differs because it considers how foods are arranged to match the preference of the appearance.

### B. Preference Learning

Several studies have attempted to optimize CG and animation parameters based on human preference. Eric *et al.* proposed an Active Preference Learning method with BO to estimate the desired parameters for changing the appearance of materials [12]. Koyama *et al.* proposed an original User Interface and an optimization method to adjust image and CG parameters [13]. Several studies also leverage Preference Learning in the context of robotics. Sadigh *et al.* proposed an active learning method to learn human reward functions from pairwise comparisons using Gaussian Process (GP) [14]. Christiano *et al.* also proposed a method of learning reward functions in Deep Reinforcement Learning from pairwise comparisons [15]. Roveda *et al.* proposed a method to tune the parameters of a path-based velocity planner for a sealing task with preference-based optimization [16]. There are also several PbBO studies on a framework for machine learning

(not for robotics). Gonzalez *et al.* proposed a BO method for pairwise comparison and an improved acquisition function by focusing on the characteristics of the bandit problem [7]. Siivola *et al.* proposed a PbBO method under batched query settings and discussed some approximate inferences [8]. Our study differs because it can be applied to tasks where ambiguous rules and physical consistency need to be considered, such as food arrangement. In addition, considering the users’ burden and maintaining motivation for answering, we adopted a sample-efficient Bayesian optimization framework. In the experiments shown later, the number of responses required in our approach was several dozed. On the other hand, deep learning-based methods may require at least several hundred responses to find a preferred arrangement.

## III. PREFERENCE-BASED BAYESIAN OPTIMIZATION

By following [8], a Bayesian optimization (BO) [17] method finds  $\mathbf{w}^* = \arg \max_{\mathbf{w}} f(\mathbf{w})$  from as few queries as possible for black-box function  $f(\mathbf{w}) : \mathcal{W} \rightarrow \mathbb{R}$ , which is difficult to evaluate directly. Therefore, BO first constructs a surrogate model that is relatively easy to evaluate with a probabilistic model and then regresses  $f$  on it from set of queries  $\{\mathbf{w}_i\}_{i=0}^{K-1}$  and corresponding  $\{f(\mathbf{w}_i)\}_{i=0}^{K-1}$ . Then BO selects new query  $\mathbf{w}'$  by evaluating the surrogate with an acquisition function that considers the balance between exploration and exploitation.

On the other hand, in PbBO, although we cannot directly observe  $f(\mathbf{w})$ , instead, answer  $y$  from a user corresponding to two-choice query  $(\mathbf{w}^0, \mathbf{w}^1)$  is given by the magnitude relation between them:

$$y = \begin{cases} 0, & \text{if } f(\mathbf{w}^0) \geq f(\mathbf{w}^1) \\ 1, & \text{if } f(\mathbf{w}^0) < f(\mathbf{w}^1) \end{cases}. \quad (1)$$

Next we determine a candidate point for a new query using  $p(\mathbf{f}|Y)$ , which is the probability distribution of  $\mathbf{f} := f(W)$  when user’s answers  $Y := \{y_i\}_{i=0}^{K-1}$  are given to queries  $W := \{\mathbf{w}_i^0, \mathbf{w}_i^1\}_{i=0}^{K-1}$ . Here we simplify the notation by writing  $f(\mathbf{w}^0)$  and  $f(\mathbf{w}^1)$  as  $f^0$  and  $f^1$  and the preference relations as  $\mathbf{w}^0 \succ \mathbf{w}^1 := f^0 \geq f^1$  and  $\mathbf{w}^0 \prec \mathbf{w}^1 := f^0 < f^1$ .

Let the hyperparameter be  $\theta \in \Theta$ , and let prior distribution  $p(\mathbf{f}|\theta)$  be the Gaussian Process (GP) [18] with mean  $\mathbf{0}$  and covariance matrix  $\mathbf{K}$ . Furthermore, assuming that the preference relation for a given two-choice query is polluted by Gaussian noise, the likelihood of the magnitude relation [19] can be defined as  $p(\mathbf{w}^0 \prec \mathbf{w}^1 | f^0, f^1) = \int_{-\infty}^{\frac{f^1 - f^0}{\sqrt{2}\sigma}} \mathcal{N}(\gamma|0, 1) d\gamma$ . Also, for  $Y$ , it is defined as  $p(Y|\mathbf{f}) = \prod_{i=0}^{K-1} p(\mathbf{w}_i^0 \prec \mathbf{w}_i^1 | f_i^0, f_i^1)$ . From the Bayes’ theorem, posterior distribution is defined as  $p(\mathbf{f}|Y, \theta) \propto p(Y|\mathbf{f})p(\mathbf{f}|\theta)$ . Since we can only obtain the likelihood numerically, we cannot obtain the posterior analytically. Therefore, we approximate  $p(\mathbf{f}|Y, \theta)$  by the Variational Bayesian method [20]. If the variational distribution is  $q(\mathbf{f}) \approx p(\mathbf{f}|Y, \theta)$ , the inference of  $\mathbf{f}'$  corresponding to new candidate point  $\mathbf{w}'$  is given by  $p(\mathbf{f}'|Y, \mathbf{w}') = \int p(\mathbf{f}'|Y, \mathbf{f}, \mathbf{w}')q(\mathbf{f})d\mathbf{f}$ . Using this, we can construct the acquisition function for selecting a candidate point for a new query.

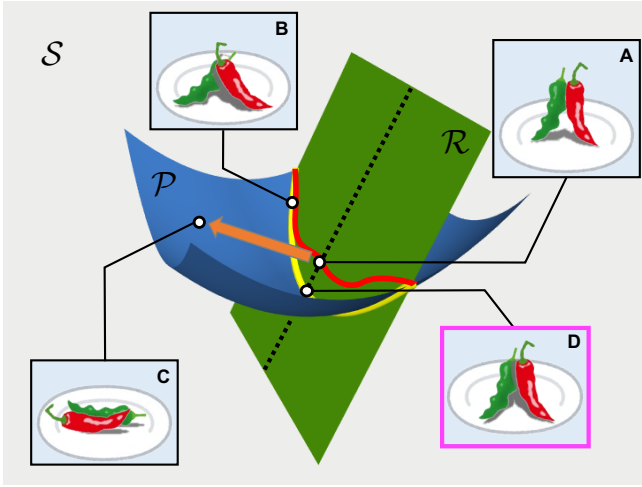


Fig. 2: Conceptual diagram of a space satisfying physical consistency (blue) and a space following domain rules (green) in food arrangement state space: Our proposed PCPBO generates arrangements in the yellow curve where both the rules and physical consistency are satisfied. Also, the image highlighted in pink is what we aim to estimate.

## IV. APPROACH

### A. Problem Formulation

Let  $M$  be the number of foods to be arranged, let  $n_s$  be the number of dimensions of each food's state, let  $\mathbf{s}_m \in \mathcal{S} \subseteq \mathbb{R}^{n_s \times M}$  be the continuous state of all the foods when the  $m < M$ th food is set, and let  $\tau := (\mathbf{s}_0, \dots, \mathbf{s}_M)$  be the state sequence. Next we assume that we can evaluate the human preference of the food arrangement from image  $\mathbf{x}_M \in \mathcal{X} \subset \mathbb{R}^{n_x}$  corresponding to final dish state  $\mathbf{s}_M$ , where  $n_x := \dim \mathbf{x}$ . We simplify the notation by referring to  $\mathbf{s}_M$  and  $\mathbf{x}_M$  as  $\mathbf{s}$  and  $\mathbf{x}$ . Hence, using maps  $k: \mathcal{S} \rightarrow \mathcal{X}$  and  $g: \mathcal{X} \rightarrow \mathbb{R}$  and composite map  $h := g \circ k$ , the degree of human preference for  $\mathbf{s}$  is given by  $c = h(\mathbf{s})$ . Here  $g$  is interpreted as user's latent preference function, but we define  $h$  as the preference function for convenience.

Let  $\mathcal{R} \subset \mathcal{S}$  be set of final dish states that satisfy the domain rules used by chefs; we estimate user's preferred arrangement  $\mathbf{s}^* = \arg \max_{\mathbf{s} \in \mathcal{R}} h(\mathbf{s})$ . However, since  $\mathbf{s}$  can be high-dimensional and it is challenging to perform optimization directly in that space, we introduce a low-dimensional parameter  $\mathbf{w} \in \mathcal{W} \subseteq \mathbb{R}^{n_w}$  (e.g., a part of a particular food item), where  $n_w := \dim \mathbf{w}$ . In addition, we design an injective function  $d: \mathcal{W} \rightarrow \mathcal{R}$  to obtain a variety of rule-abiding arrangements that change with  $\mathbf{w}$ . Finally, we replace the problem of estimating  $\mathbf{s}^*$  with that of estimating preferred weight  $\mathbf{w}^*$ :

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} h(d(\mathbf{w})). \quad (2)$$

### B. Idea for Solution

We use a CG-based approach to generate  $\mathbf{x}$  corresponding to  $\mathbf{w}$  and aim to solve (2) with PbBO. However,  $\mathbf{s}^*$  estimated from comparisons that consider only the rules can be physically infeasible. Now we explain this problem using Fig. 2, which shows state space  $\mathcal{S}$  regarding the arrangement of red

and green peppers, and the green and blue regions represent  $\mathcal{R}$  and  $\mathcal{P}$ , the set of physically consistent arrangements. In this example, we set the following rule: "Place the foods to resemble mountains." Furthermore, the red curve represents  $d$ , and  $\mathbf{w}$  corresponds to the angle between the two peppers.

In Fig. 2, A and B have different values of  $\mathbf{w}$ , so the angles are different from each other. If we consider A corresponds to  $\mathbf{w}^*$ , we can represent it as  $d(\mathbf{w}^*)$  because it lies on the red curve. Here the black dashed line represents the set of arrangements corresponding to  $\mathbf{w}^*$ . Since we are interested in physically arranging the foods, we place them with A as a reference. Unfortunately, because of  $\mathcal{R} \ni \mathbf{B} \notin \mathcal{P}$ , we get C where collapses and does not follow the rule, shown at the tip of the orange arrow. This means that physically executing such an estimated arrangement would result in an unexpected one, which might not satisfy the rule. Thus, just using the rules fails because they are qualitative and ambiguous, and this tendency is especially pronounced in food arrangements where multiple objects come into contact.

Our idea to tackle this difficulty is to generate the arrangements simultaneously considering the physical consistency and the domain rules in PbBO for solving (2). In Fig. 2, the yellow curve represents the space where both the consistency and rules are satisfied. Hence, B and D are physically feasible, and D at the intersection with the dashed line is the arrangement we aim to estimate. Note that we describe the arrangement on the yellow curve as  $\hat{\mathbf{s}}$ , and when we want to emphasize that it corresponds to  $\mathbf{w}$ , we write it as  $\hat{\mathbf{s}}_{\mathbf{w}}$ .

### C. Physically Consistent Preferential Bayesian Optimization

We propose Physically Consistent Preferential Bayesian Optimization (PCPBO) to implement our idea. PCPBO performs alternately with a model-based optimization to optimize the arrangements against cost function  $\|\mathbf{s} - d(\mathbf{w})\|$  and PbBO to estimate  $\mathbf{w}^*$  in the preference function based on the optimized arrangement. Let  $\mathbf{a}_m \in \mathcal{A} \subseteq \mathbb{R}^{n_a}$  be the continuous action of arranging an  $m$ th food on a dish, and let  $\mathcal{U} := (\mathbf{a}_0, \dots, \mathbf{a}_{M-1})$  be the action sequence. Then  $\mathbf{s}$  is obtained by recursively applying  $\mathcal{U}$  to the state transition model with physical consistency  $F: \mathcal{P} \times \mathcal{A} \rightarrow \mathcal{P}$ . Therefore, we use  $\mathbf{s}$  and  $\|\mathbf{s} - d(\mathbf{w})\|$  in the model-based optimization to obtain optimal action sequence  $\hat{\mathcal{U}}$ . We use notations  $\hat{\mathbf{s}}_{\mathbf{w}}$  and  $\hat{\mathcal{U}}_{\mathbf{w}}$  to emphasize their dependence on  $\mathbf{w}$ .

To summarize, our PCPBO is formulated as the following bi-level optimization:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} h(\xi(\hat{\mathcal{U}}_{\mathbf{w}})) \quad \text{with} \quad (3)$$

$$\hat{\mathcal{U}}_{\mathbf{w}} = \arg \min_{\mathcal{U}} \|\mathbf{s}_M - d(\mathbf{w})\|, \quad (4)$$

$$\text{s.t. } \mathbf{s}_{m+1} = F(\mathbf{s}_m, \mathbf{a}_m), \mathbf{s}_0 = \boldsymbol{\alpha},$$

where  $\xi$  is a map that applies  $F$  recursively to give corresponding  $\hat{\mathbf{s}}_{\mathbf{w}}$  from  $\hat{\mathcal{U}}_{\mathbf{w}}$  and  $\boldsymbol{\alpha}$  is the initial state.

Since this bi-level optimization problem cannot be solved analytically, we solve it by the following sequential procedure with a query generation that simultaneously considers the physical laws and domain rules:

- 1) select  $(\mathbf{w}^0, \mathbf{w}^1)$  by an acquisition function (IV-D);

- 2) generate  $(\hat{s}^0, \hat{s}^1)$  corresponding to  $(w^0, w^1)$  by physical simulation-based optimization (IV-E);
- 3) generate  $(x^0, x^1)$  corresponding to  $(\hat{s}^0, \hat{s}^1)$  by a renderer;
- 4) present  $(x^0, x^1)$  as a query to a user;
- 5) do a query synthesis if the user skipped (IV-F);
- 6) do learning preference function using answer  $y$  and  $(w^0, w^1)$  (III),

where  $\hat{s}_{w^k}$  and  $x_{w^k}$  are denoted as  $\hat{s}^k$  and  $x^k$  for simplicity of notation. Here steps 1) through 5) are repeated until the estimation converges. We illustrate the flow of this procedure in Fig. 3. Also, the code is available on our project page<sup>1</sup>. From the following, we describe each procedure.

#### D. Weight Candidate Selection

As in the previous study [7], the space of  $w$  is discretized, and Thompson sampling is used to determine candidate points  $(w_0, w_1)$  for a new query. Specifically, after  $f^0, f^1 \sim q(\mathbf{f})$ ,  $w_0$  and  $w_1$  are searched for  $w^0 = \arg \max_w f^0$  and  $w^1 = \arg \max_w f^1$ , respectively. However, we found a tendency to present candidates with close distances on  $w$  when  $\dim w$  is low. Therefore, we finitely reselect the candidates when their distances are close. In addition, they are randomly selected for  $N_{\text{init}}$  times in the early stages. Furthermore, we design the  $w$  space so that each dimension is within  $[0, 1]$ .

#### E. Physical Simulation-based Optimization

Designing  $d$  so that  $d(w)$  satisfies the physical constraints over  $\mathcal{W}$  is difficult since  $d$  is designed based on chefs' qualitative and ambiguous advice, and there are complex contacts between objects in the food arrangements. Hence, we seek a physically consistent  $s_M$  that is sufficiently similar to  $d(w)$  that meets the rules but does not consider physical feasibility. PCPBO finds the optimal action sequence  $\hat{U}$  that generates the desired arrangement by solving an optimal control problem with  $\|s - d(w)\|$  as the cost function in (4) using physical simulation-based optimization. Since the state transitions become nonlinear and complex when multiple foods come into contact, (4) is optimized by the Cross Entropy Method (CEM) [21], a sampling-based optimal control method. Also, we constructed a parallelized food arrangement simulation environment on Isaac Gym [22] and directly used it as the state transition model for CEM.

Since a physical simulator is used directly as a state transition model, PCPBO does not require learning state transitions in advance to handle food items with different colors and shapes but only preparing their meshes. Specifically, we prepared meshes for rendering, used for user queries, and meshes for detecting contact. The former was obtained by 3D scanning food samples, and the latter was designed with a simplified shape (e.g., rectangular) based on the former. In addition, the design of  $d$  and  $w$  for new dishes is performed on CG without a physics engine using a mesh for rendering [23]. Note that for tasks that do not require consideration of physical constraints, much performance difference would

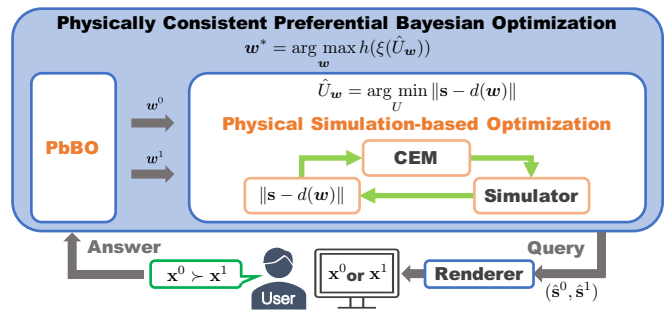


Fig. 3: Framework of proposed Physically Consistent Preferential Bayesian Optimization (PCPBO)

not occur between PCPBO and PbBO since there is no need to perform physical simulation-based optimization.

#### F. Query Skipping and Synthesis

We believe that if both  $\hat{s}^0$  and  $\hat{s}^1$  are highly undesirable, the users find it difficult to answer and might make a wrong selection. Therefore, we asked them to skip the query in this case. On the other hand, to effectively use the skipped queries, we synthesized new queries using them and the arrangement selected in the past and added them to the dataset. More precisely, the two arrangements in the skipped query are combined with the most recently selected arrangement to make two queries. We call this method query synthesis. If the selected arrangement does not yet exist, we accumulate the queries until they are no longer skipped.

However, since  $s$  can be high-dimensional, we design map  $d : \mathcal{W} \rightarrow \mathcal{R}$  by introducing lower-dimensional weight  $w \in \mathcal{W} \subseteq \mathbb{R}^{n_w}$  that satisfies the rules, although it can also change the arrangement, where  $n_w := \dim w$ .

### V. SIMULATION EXPERIMENT

#### A. Experimental Setup

We conducted a simulation experiment with simulated users that automatically answers queries according to designed human preference models. Note that we assume that no access to their preference models is given; instead, we estimate  $w$  in the models from the answers according to those models. This experiment has two purposes. First, we investigated the estimation performance of the preferred weights with and without considering physical consistency using simulated users that answers accurately. Second, we investigated the effectiveness of the query synthesis using the users that returned uncertain answers.

1) *Arrangement Tasks:* We chose the arrangement of simmered taro and deep-fried shrimp as the tasks for our simulation experiment. Those were designed from rules based on advice from the professional chefs of a Japanese foodservice company. These rules are qualitative and ambiguous (Table I). Figure 4 shows the physical contact environment with Isaac Gym [22]. In Fig. 4a, the white and green object represents taro and snap pea. In Fig. 4b, the yellow and orange cuboids represent the fried shrimp, and the green and gray rigid bodies represent shredded cabbage and a small dipping bowl, each of which is fixed on a white plate. Note

<sup>1</sup><https://y-kwon.github.io/pcpbo/>

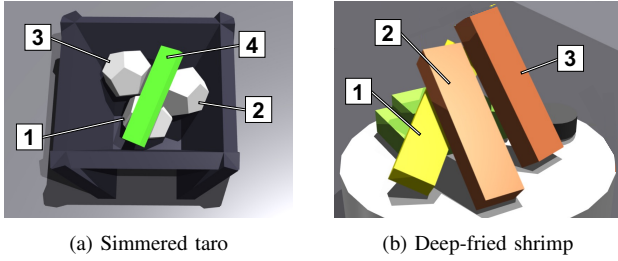


Fig. 4: Physical contact environment in the simulation experiments: We approximate foods and the dish with rigid bodies of simple shapes. The numbers in the figure indicate the order of arrangement. Also, each rigid body is colored for debugging purposes only.

TABLE I: Examples of qualitative and ambiguous rules in Japanese cuisine [9]: The chefs utilize several of these rules depending on the dish.

No.	Description
1	View a dish from an angle.
2	The colors of the dish and the foods are complementary.
3	Place the foods to resemble mountains.
4	Place the foods low in the front and high in the back.
5	Pay attention to the placement of vivid color foods.

that we omit the environment for visualization and use only that for contact detection since simulated users (described below) can directly observe the food item poses.

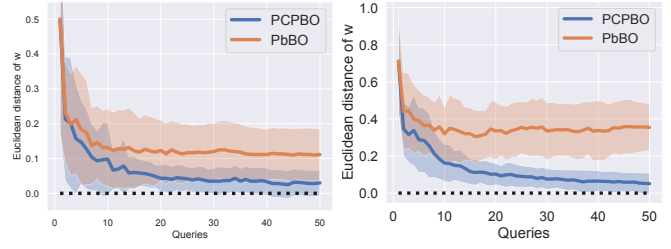
In both tasks, we defined  $\mathbf{s}_m$  and  $\mathbf{a}_m$  as

$$\mathbf{s}_m = (x_0, y_0, z_0, \theta_0, \phi_0, \psi_0, \dots, \psi_M), \quad (5)$$

$$\mathbf{a}_m = (x_m^{\text{ref}}, y_m^{\text{ref}}, \psi_m^{\text{ref}}), \quad (6)$$

where  $(x, y, z)$  and  $(\theta, \phi, \psi)$  are the position and posture (roll, pitch, yaw angles) of the food, and ref stands for reference. Note that we set  $\theta_m^{\text{ref}}$  and  $\phi_m^{\text{ref}}$  to 0, and  $z_m^{\text{ref}}$  is determined from a heuristic algorithm that drops the food from as low a height as possible. In simmered taro environment, for the four food items to appear, state sequence  $\tau = (s_0, s_1, s_2, s_3)$ . Since we also give a specific action to the taros, action sequence  $U = (a_3)$ , which is restricted by the rules. In deep-fried shrimp environment, we fix all the foods except the fried shrimp, so  $\tau = (s_0, s_1, s_2)$ , but since we give a specific action to the fried shrimp, represented by the yellow cuboid,  $U = (a_1, a_2)$ , which is also restricted by the rules.

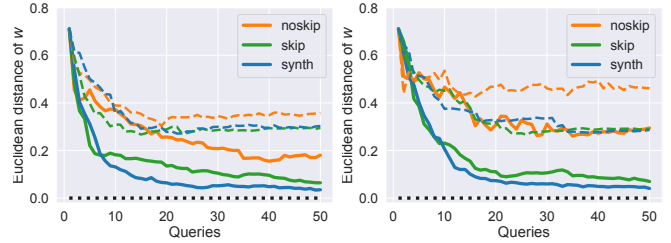
2) *Preferred Weight Estimation Methods*: To achieve the first purpose, we prepared PbBO as a comparison method, which does not consider physical consistency but only the rules. This PbBO physically places the foods with  $d(\mathbf{w})$  as a reference arrangement. In simmered taro environment, we assume that one-dimensional  $\mathbf{w}$  space corresponds to  $\psi_3$  and that  $x$  and  $y$  are roughly determined from  $\psi$  by the rules. In deep-fried shrimp environment, both methods search for two-dimensional  $\mathbf{w}$  space corresponding to  $\psi_1$  and  $\psi_2$ , assuming that the rules roughly determine  $x$  and  $y$  from  $\psi$  as well. Each dimension of the  $\mathbf{w}$  space is discretized at 21 points. We also performed an ablation study in the deep-fried shrimp environment to investigate the effectiveness of query synthesis, which is the second purpose of this experiment. There are three cases: query synthesis is enabled, it is disabled, and only query skipping is enabled (without



(a) Simmered taro

(b) Deep-fried shrimp

Fig. 5: Transition of Euclidean distance between preferred weights and estimated weights according to number of queries for case of ideal users: Each solid line and shaded region are mean and standard deviation of ten trials of three simulated users.



(a) Less uncertain users

(b) More uncertain users

Fig. 6: Comparison of preferred weight estimation results with and without query synthesis under uncertain users setting. Here *noskip* indicates a setting w/o query synthesis, *skip* denotes a setting w/ query skip (w/o synthesis), and *synth* indicates a setting w/ query synthesis. Additionally, solid and dashed lines represent estimation results by PCPBO and PbBO.

synthesis). The total number of queries  $N$  and the number of randomly generated queries  $N_{\text{init}}$  are 50 and 1.

3) *Simulated Users*: We designed several human preference models and prepared simulated users that answer queries according to these models. The simulated users have corresponding  $\mathbf{w}_{(n)}^*$  and answer whether  $\hat{s}^0$  or  $\hat{s}^1$  is preferable by directly calculating the degree of preference  $c_{(n)}$  from  $\hat{\mathbf{s}}$  rather than  $\mathbf{x}$ , where the subscript  $(n)$  indicates the  $n$ th user. Hence, they give low  $c_{(n)}$  to the arrangements that do not satisfy the rules. To investigate how the uncertainty of answering affects the estimation, we prepared the following two types of simulated users:

- **Ideal users**: return accurate answers based on (1).
- **Uncertain users**: return uncertain answers based on (7) in Appendix A. We prepared more uncertain users and less uncertain users that differ in the degree of uncertainty of their answers.

We also assume multiple users and prepared  $\mathbf{w}_{(n)}^*$  for each type of simulated user: 0.1, 0.5, and 0.9 for the simmered taro task and (0.2, 0.6), (0.3, 0.3), and (1.0, 0.4) for the deep-fried shrimp task.

We ran ten trials under each condition. The parameters used in this experiments are available on our project page.

## B. Experimental Results

1) *Preferred Weight Estimation*: Figure 5 shows the estimation results of  $\mathbf{w}$  by PCPBO and PbBO under the ideal users setting. The horizontal axis represents the number of queries, and the vertical axis represents the Euclidean distance between  $\mathbf{w}_{(n)}^*$  and estimated  $\mathbf{w}$ . Each solid line and

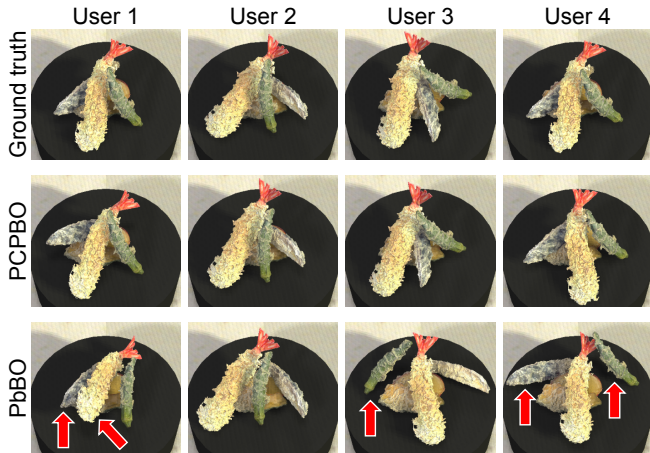


Fig. 7: Comparison of pre-selected (Ground truth) and estimated arrangements (PCPBO and PbBO) for each user: These figures are enlarged and clipped near center of original images. Red arrows indicate foods that fell over unexpectedly.

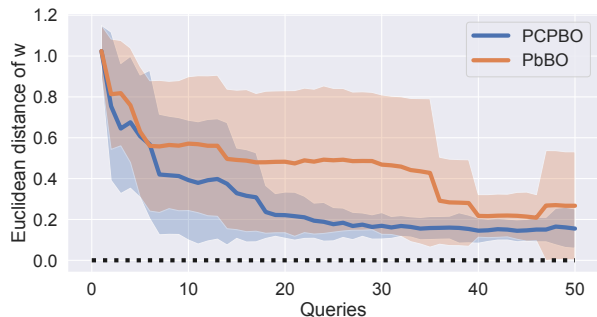


Fig. 8: Transition of Euclidean distance between preferred weights and estimated weights according to number of queries for human users: Each solid line and shaded region are mean and standard deviation of three trials of four users.

the shaded region are the mean and standard deviation of ten trials for all the users. PCPBO, which considers physical consistency, estimates the preferred weights more accurately than physically inconsistent PbBO.

2) *Analysis of Query Synthesis*: Figure 6 shows the estimation results of  $w$  under the uncertain users setting. *synth*, *noskip*, and *skip* respectively represent the cases where query synthesis is enabled, disabled, and only query skipping is enabled (i.e., skipped queries are discarded). In *synth* and *skip*, the users skipped the query in the third and fourth conditions on the right side of (7) in Appendix A. Each solid and dashed line represents PCPBO and PbBO estimates, corresponding to the mean of ten trials for all the users. Forcing uncertain users to answer without providing the query skip mechanism lowers accuracy, which becomes more pronounced when uncertainty increases. The results of *synth* are slightly better than that of *skip*. Finally, the skip rate’s mean and standard deviation were  $0.45 \pm 0.17$  for *skip* and  $0.25 \pm 0.12$  for *synth* when answered by less uncertain users and  $0.19 \pm 0.06$  for *skip* and  $0.12 \pm 0.04$  for *synth* when answered by more uncertain users.

In summary, our proposed PCPBO estimates preferred weights with higher accuracy than the physically inconsistent PbBO when simulated users return accurate answers.

TABLE II: Subjective likeability of estimated arrangements by a 7-point Likert scale: Each method uses 50 times queries to perform estimation.

		Trial 1	Trial 2	Trial 3	Average score
User 1	PCPBO	6	6	4	<b><math>5.33 \pm 0.94</math></b>
	PbBO	2	5	4	$3.67 \pm 1.25$
User 2	PCPBO	5	5	6	$5.33 \pm 0.47$
	PbBO	5	6	5	$5.33 \pm 0.47$
User 3	PCPBO	5	6	5	<b><math>5.33 \pm 0.47</math></b>
	PbBO	2	4	2	$2.67 \pm 0.94$
User 4	PCPBO	6	7	3	<b><math>5.33 \pm 1.70</math></b>
	PbBO	2	4	5	$3.67 \pm 1.25$

TABLE III: Query skips per trial and average skip rate for each user

		Trial 1	Trial 2	Trial 3	Average rate
User 1	PCPBO	11 / 50	23 / 50	11 / 50	<b><math>0.30 \pm 0.11</math></b>
	PbBO	38 / 50	21 / 50	43 / 50	$0.68 \pm 0.19$
User 2	PCPBO	5 / 50	14 / 50	7 / 50	$0.17 \pm 0.08$
	PbBO	8 / 50	5 / 50	5 / 50	<b><math>0.12 \pm 0.03</math></b>
User 3	PCPBO	8 / 50	11 / 50	4 / 50	<b><math>0.15 \pm 0.06</math></b>
	PbBO	21 / 50	23 / 50	15 / 50	$0.39 \pm 0.07$
User 4	PCPBO	0 / 50	2 / 50	0 / 50	<b><math>0.01 \pm 0.02</math></b>
	PbBO	8 / 50	10 / 50	3 / 50	$0.14 \pm 0.06$

Even when they return uncertain answers, PCPBO estimates precisely by combining the query synthesis.

## VI. USER STUDY

### A. Experimental Setup

Following our simulation experiment, we conducted an evaluation experiment with actual human users. This user study has three purposes. First, we compared the estimation performance of preferred weights between PCPBO and physically inconsistent PbBO. Second, we also compared the subjective preference of the results between them. Finally, we analyzed which queries were skipped by the users.

The experimental procedure is as follows:

- confirm user’s preferred arrangement (VI-A.2)
- answer  $N$  times two-choice queries (VI-A.3)
- answer likeability by a 7-point Likert scale (VI-A.4)

where b) and c) were repeated six times for statistical evaluation. Here the first two times were performed in the order of PbBO and PCPBO, and the order of the remaining four was randomly determined.

1) *Tempura Arrangement Task*: For this user study, we designed the arrangement task of tempura, which is a dish where ingredients are deep-fried. We chose sweet potato, pumpkin, eggplant, shrimp, and Japanese shishito pepper as ingredients. The task was designed from rules based on a Japanese chef’s advice (Table I). Based on the rules, first we fixed the food’s arrangement order and then set the sweet potato and pumpkin with specific states. We also restricted the states of the remaining foods and defined the states and actions as (5) and (6) and  $\tau = (s_2, s_3, s_4)$  and  $U = (a_2, a_3, a_4)$ . The meshes for contact detection are approximated by rectangles.

2) *Preferred Arrangement Confirmation*: The users selected the preferred arrangement in advance to quantitatively evaluate the distance between estimated  $w$  and  $w_{(n)}^*$ . To shorten the experiment time, we ran CEM ten times for each  $w$  in advance to collect candidate images and presented the lowest cost image to the users.

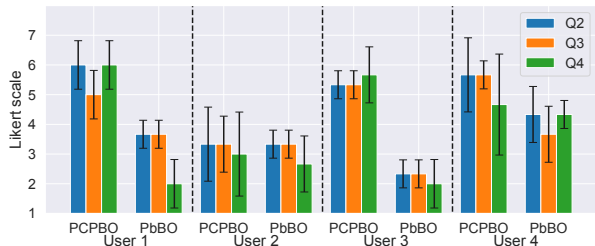


Fig. 9: Results on subjective evaluation of quality of generated queries by a 7-point Likert scale: Each colored and error bar represents mean and standard deviation.

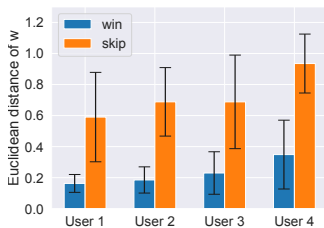


Fig. 10: Comparison of Euclidean distances between skipped (*skip*) and unskipped (*win*) arrangements generated by PCPBO: *skip* consists of the left and right side arrangements of skipped queries, and *win* consists of the arrangement of chosen side in unskipped queries. Each colored and error bar represents mean and standard deviation.

3) *Answering Two-Choice Queries*: The users answered  $N$ -time queries with “I prefer the left one,” “I prefer the right one,” and “I don’t like either.” The pre-selected preferred arrangement was displayed near the user to allow quantitative evaluation. Although the arrangement must meet domain rules as the requirement of the foodservice industry, physically inconsistent PbBO often generates arrangements that violate them. Therefore, we taught the users the rules and instructed them to select “I don’t like either” when neither arrangements were preferable or failed to satisfy the rules. To shorten the experiment time, CEM was not run in real-time in PCPBO, and the presenting images were randomly chosen from the candidate images collected in VI-A.2.

4) *Evaluation by Likert Scale*: The users judged the likeability of the estimated arrangement images on the 7-point Likert scale where higher is more preferable (Q1). In addition, they answered the following three items:

- Q2: quality of presented arrangements (the higher the better);
- Q3: ease of answering (the higher the easier);
- Q4: how quickly the preferred arrangement appeared (the higher the quicker).

At the end of the experiment, the users were allowed to revise their choices of the Likert scale for their consistency.

PCPBO and PbBO searched for three-dimensional  $w$  space corresponding to  $\psi_2$ ,  $\psi_3$ , and  $\psi_4$ , assuming the rules roughly determine  $x$  and  $y$  from  $\psi$ . Four men (age: 23–28) who are familiar with tempura participated in this user study. All the experiments were conducted under the approval of the Ethics Committee of Nara Institute of Science and Technology. The parameters used in this experiments are available on our project page.

## B. Experimental Results

1) *Preferred Weight Estimation*: Figure 7 shows each user’s pre-selected preferred arrangement (upper) and examples of the arrangement estimated by PCPBO (mid-

dle) and PbBO (lower). Here  $w_{(n)}^*$  is  $(0.1, 0.4, 0.8)$ ,  $(0.85, 0.15, 0.55)$ ,  $(0.65, 0.4, 1.0)$ , and  $(0.1, 0.4, 0.85)$  in order from user 1. PCPBO’s arrangements are less collapsed than PbBO’s. Focusing on user 3, his pre-selected one and PbBO’s are entirely different. On the other hand, there is little difference between the two estimated arrangements for user 2. This is because his pre-selected one is physically stable, and it is sufficient to use the map  $d$ , which only considers the rules. Therefore, we note that the results for user 2 tend to differ from those of the other users. Next Fig. 8 shows the estimation results of  $w$  by PCPBO and PbBO. The horizontal axis represents the number of queries, and the vertical axis represents the Euclidean distance between  $w_{(n)}^*$  and estimated  $w$ . Each solid line and the shaded region are the mean and standard deviation of three trials for four users. These results indicate that PCPBO estimated  $w_{(n)}^*$  more accurately and stably than PbBO.

2) *Subjective Preference of Queries*: Table II shows the results of asking about the likeability for the estimated arrangement images on the 7-point Likert scale. PCPBO’s arrangement tended to be preferred over PbBO’s, except by user 2, and the paired t-test results indicate a significant difference between them ( $p = 0.02$ ). Next Fig. 9 shows the Likert scale results for the three items described in VI-A.4. PCPBO’s scores are equal to or higher than PbBO’s in the subjective ratings, and the paired t-test results for each item are ( $p = 0.009$ ) for Q2, ( $p = 0.002$ ) for Q3, and ( $p = 0.013$ ) for Q4, indicating a significant difference between them. These results show that PCPBO’s arrangement is also more preferred in subjective evaluations.

3) *Analysis of Query Synthesis*: Table III compares PCPBO and PbBO in terms of query skipping. The average skip rate for PCPBO is lower except for user 2, which suggests that it generates higher quality queries than PbBO. Figure 10 compares the distances of  $w$  that corresponds to the skipped and unskipped arrangements in the queries generated by PCPBO. In Fig. 10, *win* corresponds to the set of selected arrangements by the user in the unskipped queries, and *skip* corresponds to the set of the left and right arrangements in the skipped queries. The users properly skipped the unpreferable ones from the *skip* results that are relatively far apart; the query synthesis properly worked well even for actual human users.

In summary, even with human users, our proposed PCPBO qualitatively and quantitatively outperformed the physically inconsistent PbBO. Moreover, these results show that the query synthesis framework worked as intended.

## VII. DISCUSSION

Although this paper focused on Japanese cuisine, PCPBO is likely applicable to various cuisines from other countries. However, our proposed method has some limitations. First, our experiments approximated the foods as rigid bodies with simple shapes. This simplification will be an obstacle when considering a robotic food arrangement. We might need a better solution, e.g., a better physical simulator and an approximation method of foods [24]. Next, a cost function and map corresponding to  $w$  must be implemented by

interpreting the rules, which is tedious. It would be valuable to address this difficulty with machine learning approaches. As another future direction, we will consider integrating two-choice queries and demonstrations [2] to estimate the preferred arrangement with fewer interactions. In addition, since the proposed method estimates the preferred object arrangement while simultaneously considering domain rules and physical constraints, it may be applicable to palletizing tasks [25] in the last mile, where driver preferences need to be considered, and preference-based cleanup tasks [26]. To those end, like our tempura case, 3D scanning of the environment and objects in the task would be needed to prepare a physical simulator with CG. Furthermore, in the former task, it is difficult for the users to estimate the weight of cardboard boxes from visual information alone, so we may need to handle weight data explicitly. Also, the latter task may require extending our method for handling many objects, including unknown ones.

### VIII. CONCLUSION

We proposed Physically Consistent Preferential Bayesian Optimization (PCPBO) for estimating the preferred food arrangement that satisfies domain rules and physical laws. The results of simulator-based tasks with simulated and human users show that our method can estimate user-preferred arrangements in which both are taken into account.

### APPENDIX

#### A. Uncertain User Model

Let two thresholds be  $t_0$  and  $t_1$  ( $t_0 < t_1$ ), and if  $c_{\min} = \min(c_{(n)}^0, c_{(n)}^1)$ , uncertain users return an uncertain answer as follows:

$$y = \begin{cases} H(c_{(n)}^1 - c_{(n)}^0), & \text{if } c_{\min} < t_0 \\ H(c_{(n)}^1 - c_{(n)}^0), & \text{if } t_0 \leq c_{\min} \leq t_1 \\ & \text{and } \frac{c_{\min} - t_0}{t_1 - t_0} \leq \text{rand}() \\ H(\text{rand}() - 0.5), & \text{if } t_0 \leq c_{\min} \leq t_1 \\ & \text{and } \frac{c_{\min} - t_0}{t_1 - t_0} > \text{rand}() \\ H(\text{rand}() - 0.5), & \text{if } c_{\min} > t_1 \end{cases} \quad (7)$$

where  $H(\cdot)$  is a unit step function and  $\text{rand}()$  returns a uniform random number within  $[0, 1]$ . To change the uncertainty of answering, we prepared two  $t_1$ s, 50 (less uncertain users) and 100 (more uncertain users). Also,  $t_0$  was set to 20.

### REFERENCES

- [1] Z. Wang, S. Hirai, and S. Kawamura, "Challenges and opportunities in robotic food handling: A review," *Frontiers in Robotics and AI*, vol. 8, 2022.
- [2] J. Matsuoka, Y. Tsurumine, Y. Kwon, T. Matsubara, T. Shimmura, and S. Kawamura, "Learning food-arrangement policies from raw images with generative adversarial imitation learning," in *17th International Conference on Ubiquitous Robots (UR)*, 2020, pp. 93–98.
- [3] K. Takahashi, W. K. H. Ko, A. Ummadisingu, and S.-i. Maeda, "Uncertainty-aware self-supervised target-mass grasping of granular foods," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 2620–2626.
- [4] G. Endo and N. Otomo, "Development of a food handling gripper considering an appetizing presentation," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4901–4906.

- [5] Z. Wang, Y. Torigoe, and S. Hirai, "A prestressed soft gripper: Design, modeling, fabrication, and tests for food handling," *IEEE Robotics and Automation Letters (RA-L)*, vol. 2, no. 4, pp. 1909–1916, 2017.
- [6] J. Furnkranz and E. Hullermeier, *Preference Learning and Ranking by Pairwise Comparison*. Springer Berlin Heidelberg, 2011, pp. 65–82.
- [7] J. Gonzlez, Z. Dai, A. Damianou, and N. D. Lawrence, "Preferential bayesian optimization," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, vol. 70, 2017, pp. 1282–1291.
- [8] E. Siivola, A. K. Dhaka, M. R. Andersen, J. Gonzlez, P. G. Moreno, and A. Vehtari, "Preferential batch bayesian optimization," in *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, 2021, pp. 1–6.
- [9] N. Miyazawa, *Delicious Making Basics—Transform your daily table with style (in Japanese)*. Ikeda Publishing, 2011.
- [10] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, "A survey on food computing," *ACM Comput. Surv.*, vol. 52, no. 5, 2019.
- [11] K. Junge, J. Hughes, T. G. Thuruthel, and F. Iida, "Improving robotic cooking using batch bayesian optimization," *IEEE Robotics and Automation Letters (RA-L)*, vol. 5, no. 2, pp. 760–765, 2020.
- [12] B. Eric, N. Freitas, and A. Ghosh, "Active preference learning with discrete choice data," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 20, 2007, p. 409–416.
- [13] Y. Koyama, I. Sato, and M. Goto, "Sequential gallery for interactive visual design optimization," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, p. 103–112, 2020.
- [14] D. Sadigh, A. Dragan, S. Sastry, and S. Seshia, "Active preference-based learning of reward functions," in *Proceedings of Robotics: Science and Systems (RSS)*, 2017.
- [15] P. F. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 4302–4310.
- [16] L. Roveda, B. Maggioni, E. Marescotti, A. A. Shahid, A. Maria Zanchettin, A. Bemporad, and D. Piga, "Pairwise preferences-based optimization of a path-based velocity planner in robotic sealing tasks," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 4, pp. 6632–6639, 2021.
- [17] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS)*, vol. 2, 2012, pp. 2951–2959.
- [18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006.
- [19] W. Chu and Z. Ghahramani, "Preference learning with gaussian processes," in *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 2005, pp. 137–144.
- [20] M. Opper and C. Archambeau, "The variational gaussian approximation revisited," *Neural Computation*, vol. 21, no. 3, pp. 786–792, 2009.
- [21] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Information-theoretic model predictive control: Theory and applications to autonomous driving," *IEEE Transactions on Robotics (T-RO)*, vol. 34, no. 6, pp. 1603–1622, 2018.
- [22] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State, "Isaac gym: High performance GPU based physics simulation for robot learning," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1 pre-proceedings (NeurIPS Datasets and Benchmarks 2021)*, 2021.
- [23] E. Rohmer, S. P. N. Singh, and M. Freese, "V-rep: A versatile and scalable robot simulation framework," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 1321–1326.
- [24] J. Gruenstein, T. Chen, N. Doshi, and P. Agrawal, "Residual model learning for microrobot control," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 7219–7226.
- [25] E. Lamon, M. Leonori, W. Kim, and A. Ajoudani, "Towards an intelligent collaborative robotic system for mixed case palletizing," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 9128–9134.
- [26] N. Abdo, C. Stachniss, L. Spinello, and W. Burgard, "Robot, organize my shelves! tidying up objects by predicting user preferences," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1557–1564.