How Can Everyday Users Efficiently Teach Robots by Demonstrations?

MARAM SAKR, University of British Columbia, Canada
ZHIKAI ZHANG, Carnegie Mellon University, USA
BENJAMIN LI, University of British Columbia, Canada
HAOMIAO ZHANG, University of British Columbia, Canada
H.F. MACHIEL VAN DER LOOS, University of British Columbia, Canada
DANA KULIĆ, Monash University, Australia
ELIZABETH CROFT, University of Victoria, Canada

Learning from Demonstration (LfD) is a framework that allows lay users to easily program robots. However, the efficiency of robot learning and the robot's ability to generalize to task variations hinge upon the quality and quantity of the provided demonstrations. Our objective is to guide human teachers to provide more effective demonstrations, thus facilitating efficient robot learning. To achieve this, we propose to use a measure of uncertainty, namely task-related *information entropy*, as a criterion for suggesting informative demonstration examples to human teachers to improve their teaching skills. This approach seeks to minimize the requisite number of demonstrations by enhancing their distribution throughout the workspace. In a conducted experiment (N=24), an augmented reality (AR)-based guidance system was employed to train novice users to produce additional demonstrations from areas with the highest entropy within the workspace. These novice users were trained for a few trials to teach the robot a generalizable task using a limited number of demonstrations. Subsequently, the users' performance after training was assessed first on the same task (retention) and then on a new task (transfer) without guidance. The results indicate a substantial improvement in robot learning efficiency from the teacher's demonstrations, with an improvement of up to 198% observed on the novel task. Furthermore, the proposed approach was compared to a state-of-the-art heuristic rule and found to improve robot learning efficiency by 210% compared to the heuristic rule.

ACM Reference Format:
Maram Sakr, Zhikai Zhang, Benjamin Li, Haomiao Zhang, H.F. Machiel Van der Loos, Dana Kulić, and Elizabeth Croft. 2024. How Can Everyday Users Efficiently Teach Robots by Demonstrations?. In *Proceedings of Transactions on Human-Robot Interaction (THRI)*. ACM, New York, NY, USA, 16 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In Learning from Demonstration (LfD), robots acquire skills through examples provided by human demonstrators, offering a promising avenue for efficient and effective robot learning. This method not only identifies optimal search spaces for learning but also allows users to convey task information naturally, contrasting sharply with direct task programming [1, 4]. Such natural interactions are particularly advantageous for novice users, who may lack the technical expertise required for programming through conventional interfaces or a deep understanding of the underlying systems [14].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Association for Computing Machinery.
Manuscript submitted to ACM

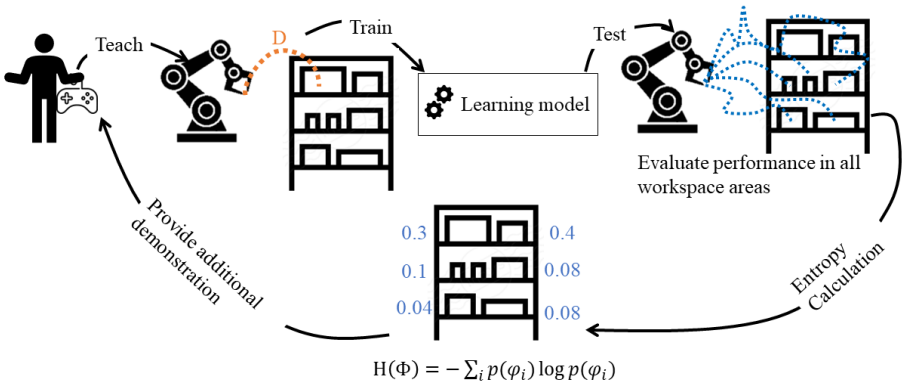


Fig. 1. System Overview: Using information entropy for guiding users to the highest uncertainty areas in LfD.

While LfD builds upon the successes of standard Machine Learning (ML) methods, which have been applied successfully across a diverse array of applications [14], teaching robots through human demonstrations introduces unique challenges. These include constraints such as limited human patience and the generation of low-quality and inefficient data [3]. The size and diversity of the training data set will determine the speed and accuracy of learning, including its generalization characteristics. Moreover, it has been shown that data modifications often impact generalization more profoundly than modifications to the learning algorithm itself [15]. Thus, this paper focuses on optimizing the input (i.e., demonstration data) to the learning algorithm, rather than altering the algorithm itself.

Human teaching strategies are typically optimized for human learners and may not align perfectly with the needs of machine learners [10]. Nevertheless, human teachers can adjust to a specific learner's requirements [22]. To facilitate this adaptation, we introduce the concept of "teaching guidance," which consists of instructions provided to human teachers with the goal of influencing their choice of examples towards those that are most informative for a particular learner (e.g., a robot).

We aim to help novice teachers provide demonstrations efficiently by using information entropy as the criterion for selecting demonstrations for robot learning and generalization through the following contributions: i) introducing task-dependent information entropy into the LfD pipeline to define uncertainty across all task space areas, as shown in Fig. 1; ii) proposing the integration of Augmented Reality (AR) with the entropy-based guidance system as a training framework to help novice users better identify informative demonstrations; iii) evaluating the proposed training framework in a user study; and iv) comparing robot learning efficiency using our information entropy approach with a state-of-the-art heuristic rule [32].

2 RELATED WORK

Teaching robots through demonstration poses challenges for lay users as they may struggle to provide useful [22, 30] and sufficient demonstrations [32]. Robots can become more proactive in their learning by requesting additional demonstrations [34] or seeking clarification [9]. However, this shift to active learning can lead to users feeling frustrated and disengaged [6]. Typically, learning from an expert teacher who selects examples optimally is faster than an active learning process where learners choose their own examples [17]. Therefore, teachers require guidance in selecting optimal examples for robot learning and generalization and ultimately learn how to define these optimal examples.

Whilst there is a great deal of research on policy-learning methods for LfD, there is considerably less research on the quality of teaching and how this might be improved [1, 4]. In fact, effective teaching can be challenging, even when instructing humans, requiring an understanding of the learner's knowledge, suitable teaching strategies, and more. Teaching robots presents an even greater challenge. Cakmak and Takayama [8] proposed using various instructional materials to guide novice teachers on how to teach robots using kinesthetic teaching and a spoken dialogue interface. Mohseni-Kabir et al. [26] designed an interactive system for teaching users a hierarchical task structure. Study participants were given explicit instructions on how to teach the robot for each task. Such an approach is untenable for scaling up to a vision of ubiquitous robotics, as it is impractical for experts to teach every end user how to program robots for every desired task. Instead, we examine how people teach novel tasks to robots in the absence of a roboticist's explicit tutelage and focus on guiding users to optimally select a set of demonstrations for the robot to learn a task.

Interactive LfD is a technique that allows robots to incrementally learn and refine their skills from human demonstrators [21, 35]. Unlike traditional LfD, interactive LfD features an iterative process of demonstration, execution, and real-time feedback from the demonstrator, allowing the robot to incrementally refine its task performance. Calinon and Billard [12] and Weiss et al. [36] both described an incremental learning approach where teachers decide on the next demonstration by observing the robot attempt the taught skill in new locations. This overlooks the question of *how* the attempts should be selected to hasten robot learning progress. Sena et al. [32] proposed heuristic rules for deciding demonstration locations in the task space. These rules are as follows: (i) provide one demonstration, starting from anywhere in the task space, (ii) continue providing demonstrations within 4 cm of the first demonstration, until the first demonstration is surrounded by successful test points, (iii) provide further demonstrations within 4 cm of the successful test points, in the area with the greatest number of failed test points. These rules are limited to their experimental setup and are hard to generalize for a different task. To overcome this limitation and provide a more general framework, we propose using a task-independent rule (information entropy) to guide users to provide efficient demonstrations [32].

A closely related field of research that considers optimal teaching of machine learning systems is *machine teaching*, which aims to design the optimal training data to drive the learning algorithm to a target model [37]. Two relevant methods are: the "teaching dimension model," which represents the minimum number of instances a teacher must reveal to uniquely identify any target concept [17], and the "curriculum learning principle," which is a general approach in which training examples are presented to the learner in a sequence tailored to learning, progressing from simple to difficult concepts [2]. This opens the door for further research on which optimal teaching strategy should be used with robots. Khan et al. [22] compared the teaching dimension model and curriculum learning principle in teaching a simple 1D classification problem, Cakmak and Lopes [7] studied optimal teaching in sequential decision tasks. Other related approaches examine how a robot can give informative demonstrations to a human [20], or formalize optimal teaching as a cooperative two-player Markov game [18]; however, neither approach addresses the machine teaching problem of finding the minimum number, quality, and locations of demonstrations needed to teach a task.

3 OBJECTIVES

The objective of this research is to enhance human teaching skills to provide efficient demonstrations to robots. Therefore, we propose using information entropy as the criterion for suggesting to the human demonstrator the most informative demonstrations from which the robot can learn. The research questions we aim to address are as follows:

(1) Does information entropy improve robot learning efficiency by guiding demonstration selection?
(2) Does an entropy-based guidance system enhance users' performance in teaching robots through demonstration?

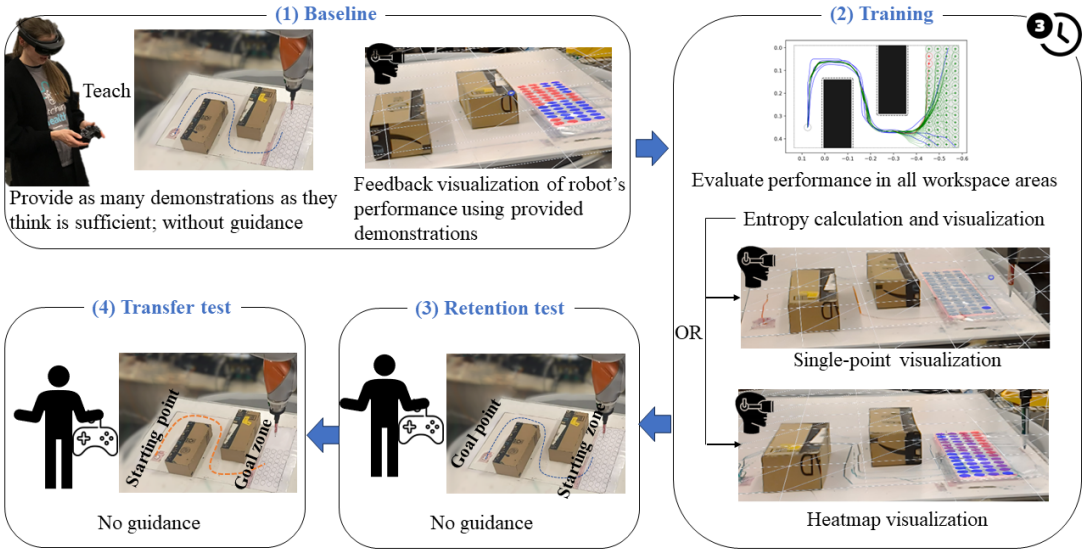


Fig. 2. Overview of the user study procedure. **(1) Baseline:** The user provides as many demonstrations as they think would be sufficient for the robot to learn the task. **(2) Training:** The user is guided to the uncertain areas for three trials. **(3) Retention test:** The user is evaluated on the same task without guidance. **(4) Transfer test:** The user teaches the robot to move from a specific starting point to a wider goal zone without guidance.

(3) How much visual information should the guidance system provide to improve the user's teaching efficiency?

To address the first question, we compared an entropy-based guidance system with a heuristic guidance system as proposed in [32] (described in Section 2). To tackle the second question, we incorporated the proposed entropy-based guidance system into a training framework for novice users to assess whether their teaching skills improved—specifically in terms of the number and distribution of provided demonstrations—after training. For the third question, we conducted a user study comparing two visual guidance systems: 1) single point suggestion and 2) heatmap representation of the task space's entropy values. This comparison aims to determine how the amount of visual information provided affects the users' teaching efficiency.

4 PROPOSED APPROACH

As described above, we propose using information entropy as the criterion for suggesting to the user the location of the next demonstration. In information theory, entropy is a measure of the uncertainty or randomness of a system or signal [33]. If the entropy is higher, more information is needed to represent an event. Mathematically, entropy is represented by the Shannon entropy formula, which is given by:

$$H(\Phi) = -\sum_{i} p(\phi) \log p(\phi) \tag{1}$$

where $p(\phi)$ is the probability of a particular event i. This probability can be calculated as a weighted sum of different task-dependent features [38].

The proposed approach applies to any task where the success criteria are defined over the entire workspace. Herein, we consider teaching a robot to navigate in a constrained workspace. This problem is generic to many robotic applications in domestic, industrial, and space exploration contexts. This category of tasks requires the robot to navigate its workspace while avoiding collisions to reach a goal position. The success criteria for these tasks can be defined as follows: a) the robot's end effector reaches the goal location, b) no collisions occur with obstacles in the environment, and c) all points on the trajectory are within the workspace (i.e., the robot should not deviate from its designated workspace).

Fig. 1 provides an overview of the teaching process using the proposed approach on an example task to illustrate its applicability. The figure shows an example of teaching a robot how to retrieve items from a shelf. The user aims to provide the minimum number of demonstrations —an efficient set of demonstrations— that enables generalization across the shelf (i.e., the workspace). The task requires the robot to pick up all items from a shelf without encountering collisions. To efficiently teach the robot, the workspace and the success criteria for the task are first defined. Next, the user provides initial demonstrations of picking up an item from the shelf. Using the provided demonstration(s), the robot learns a model of how to perform the task. The robot is then evaluated on all regions of the shelf, and the entropy of each region is calculated as shown in equation 1. The probability is defined based on the success criteria as follows:

$$p(\phi) = w_1 d_{goal} + w_2 n_{collision} + w_3 n_{outside} \tag{2}$$

where d_{goal} is the distance between the trajectory endpoint and the task goal point, $n_{collision}$ is the number of samples in collision with the obstacles, and $n_{outside}$ is the number of samples outside the task space. The coefficients w_1, w_2, w_3 are the weights for each feature of the probability and are experimentally calculated. Each feature is normalized by its maximum value before using it in entropy $H(\Phi)$ calculations. A 2D instantiation of this task was used in our evaluations to facilitate the comparison with the state-of-the-art work [32], as explained in Section. 5.1.

Areas with high entropy are highlighted to guide the user in providing additional demonstrations. This iterative process continues until the entropy for all regions becomes sufficiently low, indicating that the robot is confident in generalizing across all regions in the workspace. Importantly, our information entropy approach is not influenced by the user's expertise level, the learning algorithm, or the task complexity. If low-quality demonstrations are provided, the guidance system may recommend more demonstrations to facilitate effective robot learning and generalization.

To guide the users to the uncertain areas in the task space, a Microsoft HoloLens AR head-mounted display [23] is used to visualize the entropy values in task space areas. Providing the user with more insight into the learning process may lead to more effective demonstrations. However, providing too much information may overwhelm and hinder the user, especially one that is untrained. To find an adequate amount of information to be presented in the visual interface about the robot's learning progress, we proposed two visualization schemes. *Heatmap visualization* shows the entropy values across the task space as a heatmap, while *single-point visualization* only highlights the most uncertain area in the task space to provide an additional demonstration as shown in Fig. 2-2. Using AR in guidance makes the training system widely applicable to both 2D and 3D tasks and increases the user's situational awareness, as the feedback is directly overlaid on the task, in contrast to the computer screen used in [32].

For task learning, we used the Task Parameterized Gaussian Mixture Model (TP-GMM) [11]. TP-GMMs have been extensively used for LfD [27, 28], and they show good generalization using a limited set of demonstrations. TP-GMMs model a task with parameters that are defined by a sequence of coordinate frames. In a D dimensional space, each task parameter/coordinate frame is given by an $A \in \mathbb{R}^{D \times D}$ matrix indicating its orientation and a $b \in \mathbb{R}^D$ vector indicating its origin, relative to the global frame. A K-component mixture model is fitted to the data in each local frame of reference.

Each GMM is described by $\left(\pi_k, \mu_k^{(j)}, \Sigma_k^{(j)}\right)$, referring to the prior probabilities, mean, and covariance matrices for each component k in frame j, respectively.

To use the local models for trajectory generation, they must be projected back into the global frame of reference and then combined into one global model. Continuous trajectories can then be generated from the global mixture model using Gaussian Mixture Regression (GMR); Calinon [11] provides further details. A Bayesian Information Criterion (BIC) was used to define the optimal number of K-Gaussian components to fit the demonstrations.

5 EVALUATION

To evaluate the proposed entropy system, a preliminary experiment in simulation comparing the proposed approach and one of the state-of-the-art heuristic guidance rules is presented. It aims to study the impact of the proposed approach on robot learning efficiency. After that, a user study is presented to explore the contribution of the proposed approach to improving human teaching skills.

5.1 Experimental Task

Fig. 3 represents the experimental task, in which a Kuka IIWA LBR14 robot is taught to navigate a 2D maze by drawing a trajectory on the whiteboard from a starting zone to a target point. Similar success criteria as in Section 4 are used. The robot should avoid collisions with the obstacles in the workspace and it should not go outside the workspace specified by a two-dimensional bounding rectangle (45 cm by 72 cm) drawn on the whiteboard. The starting zone is defined by a 45 cm by 15 cm rectangle. In the experiment, the starting zone is discretized to a 4 x 14 test grid with 56 circles separated by 0.5 cm.

Participants use the joystick to teach the robot to navigate through the maze. The teaching process is finished if the robot can generate a "successful" trajectory for 90% of the points in the starting zone to the goal position. The success criteria of the trajectories are similar to what is defined in the example task in Section 4. Microsoft Hololens is used to overlay a geometrically accurate starting zone on the real environment, offering user feedback on the robot's progress and guiding efficient demonstration, as shown in Fig. 2.

A modified version of this task was used in the transfer test for evaluating the *user's* performance in new tasks. The training task was flipped; the goal point became a starting point and the starting zone became the goal zone. This goal zone was shortened to a 4 x 8 test grid, as shown in Fig. 2.

5.2 Performance Metrics

5.2.1 *Quantitative measures.*

(1) **Task Completion Time:** We measured the total time, t, required by the user to complete the experiment task.
(2) **Number of Demonstrations:** We measured the number of demonstrations $|\mathcal{D}|$ that achieve at least 90% coverage of the starting zone.
(3) **Teaching Efficacy:** Efficacy refers to the ability to perform a task to a satisfactory or expected degree. Here, we used the teaching efficacy defined in [32], which is the ratio between the successful points,$|\hat{I}|$, and the total number of tested points, $|I|$, ($\epsilon = \frac{|\hat{I}|}{|I|}$).
(4) **Teaching Efficiency:** Efficiency in any given application is often context-dependent; however, typically it is desirable to minimize the total number of demonstrations required because this is correlated with both the time spent teaching and the space needed to store data. Here, we used the teaching efficiency defined in [32], which is

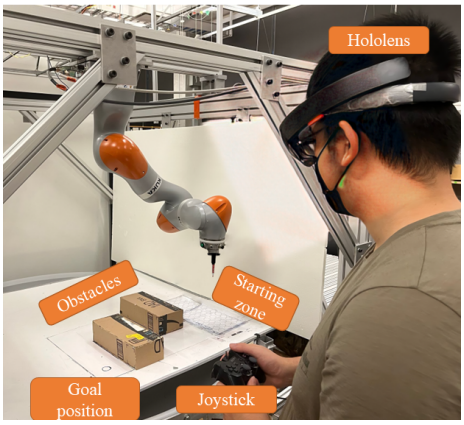


Fig. 3. Experimental setup used in the user study.

the efficacy normalized by the number of the provided demonstrations, $|\mathcal{D}|$,

$$\eta = \frac{\epsilon}{|\mathcal{D}|}. \tag{3}$$

5.2.2 Qualitative measures.

(1) **NASA Task Load Index (NASA-TLX):** This questionnaire [19] is composed of six questions asking the participants to rate their perceived task load in six different aspects on a 21-point scale.
(2) **System Usability Scale (SUS).** This scale evaluates the usability of the proposed training system. The SUS [5] questionnaire includes ten questions asking the participant to rate different aspects of the system's usability on a 10-point scale. An overall score is then calculated.

5.3 Preliminary Experiment: Guidance Rule in Simulation

In this experiment our goal is to compare the entropy-based guidance rule with the heuristic guidance rule proposed in [32] (described in Section 2). Various factors influence robot learning and generalization, including the quality, quantity, and distribution of provided demonstrations. Here, we specifically focus on the guidance rule that governs the "number" and "distribution" of provided demonstrations, while keeping all other factors constant. To ensure consistent demonstration quality, we employed the motion planner RRT-Connect [24] to generate demonstrations for both guidance rules. Trajectories were generated from all points within the starting zone to the goal point and saved for the experiment. Additionally, to simulate a real-case scenario involving human demonstrations, we collected expert demonstrations, performed by one of the paper's authors, for all points within the starting zone of the experimental task shown in Fig. 3.

For the comparison study, for each guidance rule, we developed a Python script to guide the robot's learning process. The experimental setup was simulated with RViz, the standard tool for visualization and interaction with robot applications implemented in ROS. The script initially proposes a point in the starting zone, and its corresponding saved trajectory (either from the motion planner or expert demonstrations) is retrieved. Then, the TP-GMM model is trained by this trajectory and tested with all points in the starting zone to identify the successful and failed trajectories. Using the guidance rule, the script selects the next point in the starting zone for the next demonstration. These steps are repeated until the learning model generates successful trajectories for 90% or more of the points in the starting zone. The same procedure is repeated 56 times for both guidance rules until every point in the task space has been tested as

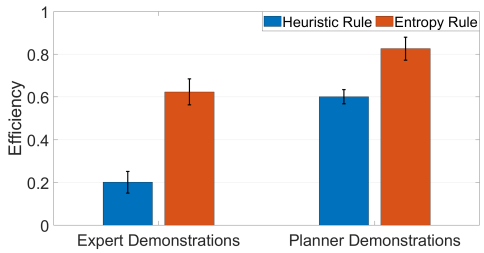


Fig. 4. The average robot learning efficiency for both the entropy rule and the heuristic rule driving the selection of planner and expert demonstrations.

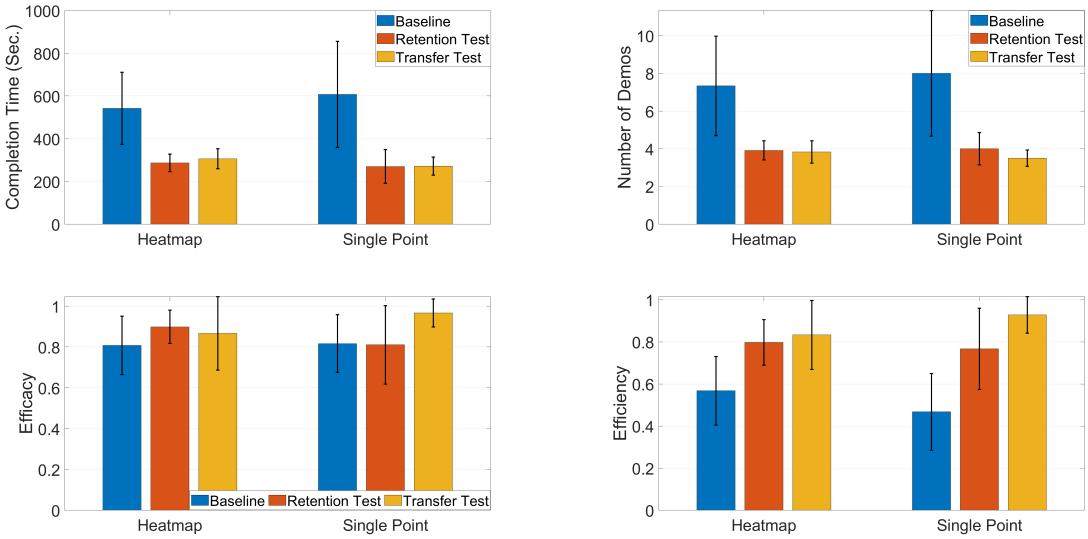


Fig. 5. Performance metrics for comparing heatmap and single-point groups in baseline, retention, and transfer tests.

the initial point. Finally, the two guidance rules are compared with respect to the number of demonstrations needed to generate successful trajectories for 90% or more of the starting zone.

5.3.1 *Hypotheses*.

The experimental hypotheses are chosen to test whether robot learning efficiency, Equation 3, improves using the proposed approach compared to the heuristic rule. The alternative hypothesis is formally defined as:

- **H1(a)**: For robot learning using planner demonstrations, learning efficiency using the entropy guidance rule is significantly higher than learning efficiency using the heuristic rule.
- **H1(b)**: For robot learning using human expert-generated demonstrations, learning efficiency using the entropy guidance rule is significantly higher than learning efficiency using the heuristic rule.

5.3.2 *Results*.

Fig. 4 shows the mean and 95% confidence interval of the robot learning efficiency in 56 trials using the heuristic rule and entropy rule with both expert and planner demonstrations. Overall, using the entropy rule achieves a higher robot learning efficiency than the heuristic rule, for both planner and expert demonstrations. A 2x2 Bayesian ANOVA was

conducted to investigate the impact of a) guidance rule, and b) demonstration type on robot learning efficiency. Both the guidance rule ($BF_{10} = 7.43 \times 10^{15}$) and demonstration type ($BF_{10} = 2.82 \times 10^{13}$) show a large effect on robot learning efficiency. In addition, robot learning efficiency using either guidance rule strongly depends on the demonstration type (the quality of the provided demonstrations) ($BF_{10} = 184.11$). These results support **H1(a)** and **H1(b)**.

Interestingly, a slight decrease in the demonstration's quality (i.e., from the planner to the expert demonstrations) causes a drastic decrease in the robot learning efficiency by 200% using the heuristic rule, while using the proposed entropy rule the efficiency decreased by only 33.9%. This shows the importance of the guidance rule in robot learning from demonstrations and its contribution to the robot learning efficiency which cannot be compensated only by improving the demonstration quality. Thus, it is important to guide the users to provide an efficient set of demonstrations for the robot to learn from.

5.4 User Study Experiment

From the simulation experiment results, we found that the entropy guidance rule significantly improves robot learning efficiency. Therefore, we utilized this rule in a training framework for users to guide them to provide an efficient set of demonstrations for the robot. Fig. 2 shows an overview of the user study procedure as follows:

(1) **Baseline:** The user provides as many demonstrations as they think would be sufficient for the robot to generate successful trajectories from any point in the starting zone to the goal position. The provided demonstrations are used to train a TP-GMM model and evaluate the performance which is provided back to the user via the Hololens as visualized feedback. The blue circles represent successfully learnt trajectories while the red circles represent failed learnt trajectories.
(2) **Training:** The user provides an initial demonstration, and then the learning model is evaluated across the starting zone. The entropy is calculated in each area and visualized through the HoloLens as a heatmap or single-point, according to the training group. The user provides an additional demonstration in the highest entropy point and the entropy visualization is updated, and so on. This continues until the robot learns successful trajectories for 90% or more of the starting zone. Each user was trained for three trials.
(3) **Retention Test:** This step is similar to the baseline to evaluate the performance after training.
(4) **Transfer Test:** The user teaches the robot to move from a specific starting point to a goal zone without guidance, described in Section 5.1.

The users were asked to fill in the NASA-TLX questionnaire [19] after their baseline and retention tests to see the effect of the training on their perceived task load. They were also asked to fill in the SUS questionnaire [5] after the training to get their subjective evaluation of the proposed training system. Lastly, a semi-structured interview was conducted before and after training to inquire about users' reasoning behind their selection of the number of provided demonstrations and their respective placements. These interviews were conducted with the intention of delving into users' mental models of robot learning before and after training.

*5.4.1 **Hypotheses**.*

The experimental hypotheses are chosen to test whether the user's teaching skills measured through robot learning efficiency are significantly improved using the proposed approach. They are defined as:

- **H2**: Robot learning efficiency will be significantly improved after training the human teacher using the proposed approach (over both groups).

- **H3**: The heatmap visualization group will achieve a significantly higher improvement in robot learning efficiency than the single-point visualization group on the retention test (i.e., same training task).
- **H4**: The heatmap visualization group will achieve a robot learning efficiency that is significantly higher than the single-point visualization group on the transfer test (i.e., on a novel task).

The user study is a between-participants design, so each participant is assigned to one group: either the heatmap group or the single-point group. A priori power analysis was conducted using G*Power version 3.1.9.7 [16] to determine the minimum sample size required to test the study hypotheses. Results indicated the required sample size to achieve 80% power for detecting a medium effect size (Cohen's $f = 0.3$), and a type I error rate $\alpha = 0.05$, is $N = 24$, or 12 participants per condition. Accordingly, the results reported below are for a population of 24 participants (14 male, 10 female; ages $\mu = 24, \sigma = 4$). We recruited participants for our user study through advertisements posted on the University of British Columbia (UBC) campus and social media. Prior to conducting the study, we obtained research ethics approval from UBC's Behavioural Research Ethics Board (application ID H20-03740-A001). We obtained informed consent from each participant before commencing the experiment.

5.4.2 ***Results***.

Quantitative Measures

Fig. 5 depicts the mean and 95% confidence interval for all performance metrics across baseline, retention, and transfer tests for both the heatmap and single-point groups. Overall, there is a noticeable improvement in all performance metrics during the retention and transfer tests when compared to the baseline performance. Efficiency improved by 140% and 164% from the baseline to the retention for both heatmap and single-point groups, respectively, and by 146% and 198% from the baseline to the transfer test for both heatmap and single-point groups, respectively. Interestingly, users' performance improved more in the transfer test than in the retention test compared to their baseline performance. Given that the efficiency metric encapsulates both the number of demonstrations and efficacy, the statistical analysis was centred on efficiency.

Bayesian mixed analysis of variance (ANOVA) showed a strong impact of the training on efficiency when comparing the baseline to the retention test ($BF_{10} = 1387.16$); supporting **H2**. The same analysis showed inconclusive results on the impact of the visualization type on efficiency ($BF_{10} = 0.4$); no support for **H3** and **H4**. Furthermore, the Bayesian independent samples t-test showed inconclusive results on the comparison between the transfer test efficiency in both heatmap and single point groups ($BF_{10} = 0.60$). This result highlights the effectiveness of the proposed entropy-based guidance system, irrespective of how the entropy information is visually presented to the user.

We also investigated the distribution of the demonstrations over the task space across the baseline, retention and transfer tests. This shows the different strategies users adopted to distribute their demonstrations over the task space. Fig. 6 shows a heatmap representing the number and locations of the provided demonstrations across different stages of the experiment for both heatmap and single-point groups. The number of the provided demonstrations notably decreased from the baseline to the two tests. The median number of baseline demonstrations was 6 and 7, respectively, for the heatmap and single-point groups, decreasing to a median of 4 in the retention and transfer tests for both groups. Compared to the baseline which has a higher number of demonstrations from points in the interior of the start zone, the retention and transfer test demonstrations are concentrated at the corners. This highlights the change in users' strategy for selecting the number and locations of demonstrations.

The proposed approach implicitly addresses the issue of low-quality demonstrations from novice users. If a user provides low-quality demonstrations, the system will request additional demonstrations to improve task learning and

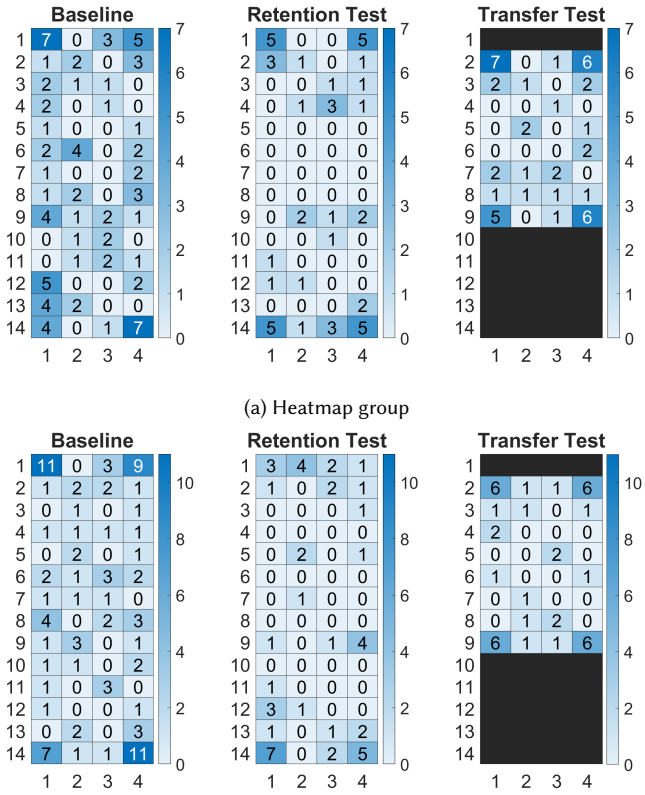
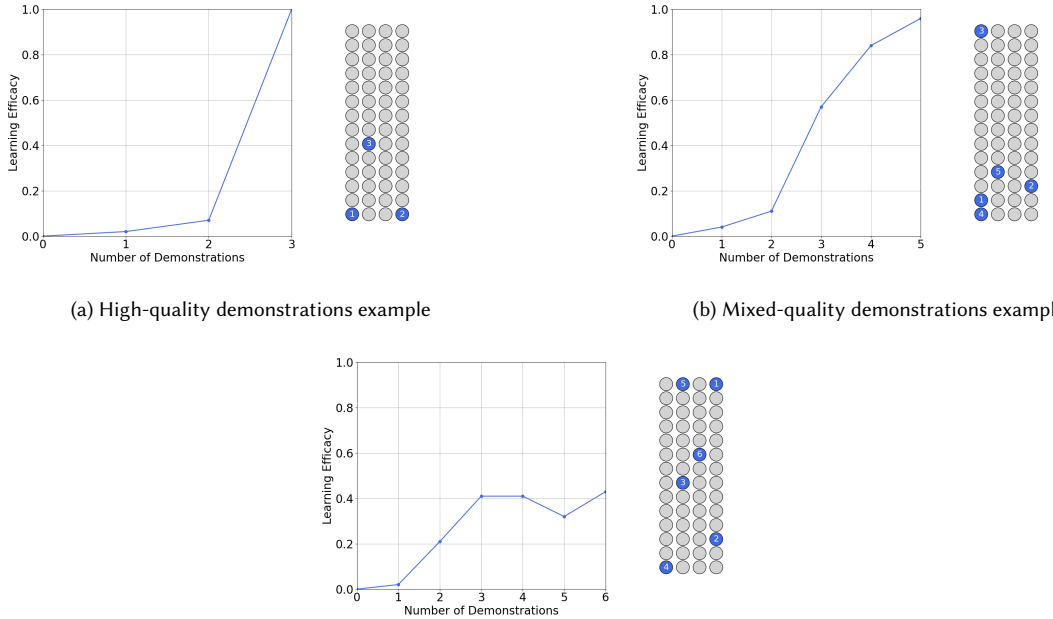


Fig. 6. The number and distribution of the demonstrations in the starting zone for the baseline, retention, and transfer tests for both the (a) heatmap group and (b) single-point group. Users provided a high number of demonstrations and distributed them across the starting zone in the baseline. After training, they provided a lower number of demonstrations focused on the corners.

generalization. For instance, Fig. 7 illustrates the pattern of robot learning efficacy for each provided demonstration from three different users along with the distribution of the demonstrations over the starting zone. The first user provided high-quality demonstrations, so only three demonstrations were sufficient for the learning model to generalize across the entire task space. The second user provided mixed-quality demonstrations, hence a higher number of demonstrations was required to achieve generalization across the task space, with an efficacy of 96%. The third user provided low-quality demonstrations, thus the learning model only generalized across 43% of the task space with six demonstrations. Here, we used task learning performance as a measure of quality, as proposed in our previous paper [30].

Qualitative Measures

Perceived Task Load: Fig. 8 illustrates the box plots for both the heatmap and single-point groups following the baseline and retention tests. It shows that both effort and frustration have decreased after training in the heatmap group while only the effort decreased in the single-point group. In contrast to the quantitative measures, users' perceived performance after training was lower than before training in both groups. In general, the task load metrics statistically



(a) High-quality demonstrations example

(b) Mixed-quality demonstrations example

(c) Low-quality demonstrations example

Fig. 7. An example of high-, mixed-, and low-quality demonstrations, illustrating their relationship to learning efficacy and their distribution across the task space.

indicate no significant difference before and after the training. This suggests that the training did not have an adverse impact on users' perceived task load or lead to overwhelming experiences.

System Usability: The usability score for the heatmap group is $(M = 75.21; 95\%CI, 64.99 to 85.42)$, while in the single point group $(M = 62.71; 95\%CI, 55.89 to 69.53)$. According to the global benchmark for SUS created by Sauro and Lewis through surveying 446 studies spanning different types of systems, the mean given score is 68 ± 12.5 [31]. When comparing this global benchmark mean score with the scores obtained from the two training groups, the observed differences were inconclusive (Bayesian one-sample t-test $BF_{10} = 0.75, BF_{10} = 0.89$, respectively). This suggests that both training approaches are usable from the user's perspective.

Interviews: Participants employed diverse strategies when determining the number and placements of demonstrations during the baseline phase. Among the 24 participants, 19 opted to establish a grid of points, strategically encompassing both the corners and intermediary positions. As articulated by one participant: *"Well, overall, I just wanted to get the corners and then a few different points along the diagonals in between. I just felt like it kind of covers the majority of the grid"*, which is reflected in Fig. 6. In contrast, two other participants opted for a completely random selection of demonstrations, favoring a higher count of demonstrations (18 and 23) to ensure comprehensive coverage of the starting area. Another participant thought that providing a dense set of points would allow the robot to learn the distance between the points in the starting zone and generalize well to other areas. Another participant provided three demonstrations that formed a triangle. They thought that the two points forming the base of the triangle would specify the width for the learning algorithm to interpolate in, while the third point would specify the height for the algorithm.

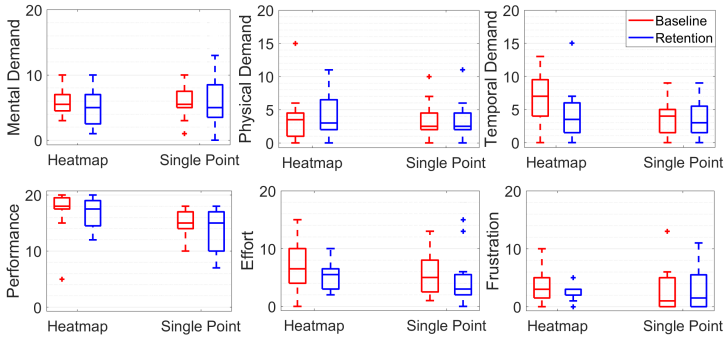


Fig. 8. Participants response to NASA-TLX [19] questions after baseline and retention test. There is no significant difference before and after the training.

After training, the participants approached the task differently and efficiently thought about the number and distribution of the provided demonstrations. Almost all participants provided an average number of demonstrations in retention equal to the average number of demonstrations they provided in the training, regardless of the training condition. As one participant said, *"After the second training trial, I realized that I don't need a really large number of points as long as I pick the four corners of the rectangular area"*. In addition, there was an agreement from several participants in the heatmap group that visualizing the generalization area of each point in the training was very helpful in deciding the location of the next demonstration in the retention test. Some participants counted the number of generalized points around one demonstration in the training and tried to distance the demonstrations in the retention by that number of points.

6 DISCUSSION

This study demonstrated that a few trials of interactive training and guidance for lay users significantly improved their teaching skills, which in turn enhanced robot learning and generalization efficiency. Notably, this learning and generalization occurred online, using demonstrations from teachers with no prior knowledge of robotics or machine learning algorithms. By using the proposed training framework, users develop an understanding of what demonstrations the robot requires to learn efficiently, without necessarily understanding how the learning process occurs. Thus, the overarching goal of this work is realized by decreasing the teaching cost through guiding users to provide a small number of demonstrations that achieve wider generalization.

From the preliminary experiment results and statistical analysis, there is support for **H1(a)** and **H1(b)**. Robot learning efficiency using the information entropy rule was significantly higher than using the heuristic rule with both expert demonstrations and motion planner demonstrations. Furthermore, using the same task as in [32] but with a larger scale, we found a failure case for the heuristic rule: it happens in 61% of the total experiment trials. This failure case happens whenever all the surrounding points are demonstrated and the learning algorithm performs poorly in learning and generalization. This failure case is expected to be even worse with novices who tend to provide low-quality demonstrations [1] that are negatively affecting robot learning and generalization [30].

From the user study results and statistical analysis, there is support for **H2** but no support for **H3** or **H4**. Both training groups showed a significant improvement in learning efficiency from the baseline to the retention and transfer tests.

This underscores the importance of training novice users before teaching the robot, consistent with prior work [8, 13]. Additionally, this study shows the benefit of the proposed entropy guidance system regardless of the visualization used in training. Notably, both training groups achieved similar performance in retention and transfer tests. This is because most participants had a preconception of creating a grid of points to cover the starting zone to help the robot learn better, as shown in Fig. 6. Subsequently, post-training, participants recognized the feasibility of achieving high efficiency with a reduced number of demonstrations, contrary to their initial assumptions. As shown in Fig. 5, the efficacy before and after training is similar. This finding contrasts with previous work [32], which found that most participants tend to underestimate the number of demonstrations required to teach the robot effectively.

For guidance system usability, the heatmap group reported a higher SUS score than the single-point group, while both training groups showed an inconclusive difference with the global mean score. We posit that this is because participants in the heatmap condition had access to a larger amount of information during training than the single-point group. The heatmap visualization provided feedback on robot learning progress and an efficient way for deciding their next demonstration. Only one participant commented that they wished to have a highlight of the highest uncertainty point in the heatmap. This is because they sometimes found the colors in the heatmap were very close to each other, making it hard to define the highest uncertainty point. Participants in the single-point group did not have access to the robot's learning progress and therefore had to depend on the demonstration point suggestions to find a pattern on how to efficiently teach the robot.

A surprising observation emerged concerning users' perceived performance post-training. Despite a significant quantitative improvement in performance, 8 out of the 24 participants reported a decline in perceived performance compared to before training. This discrepancy was investigated by analyzing their interview data related to their approach to determining the number and locations of demonstrations. These participants experienced significant shifts in their strategies after training, which impacted their confidence and perception of performance. For instance, one participant provided 16 demonstrations in the baseline. After training, they provided four demonstrations in the retention test and remarked in the interview, *"I just used four points, although I'm not sure I covered 90%, I don't know."* Others changed their strategies from dense point sets to a more distributed approach, albeit with skepticism. These findings suggest the potential need for adaptive training tailored to individual participants' requirements, especially for those with strong pre-training strategy convictions, aligning with previous research [30].

Another interesting observation is that one participant without prior knowledge of robotics realized the importance of the demonstration's quality in conjunction with the number and location of the provided demonstrations for robot learning and generalization [25]. They tried different starting points in the training and then realized that providing smooth demonstrations without abrupt corners allows the robot to learn and generalize faster with a smaller number of demonstrations, as shown in Fig. 7. However, it is challenging to extrapolate that all participants would naturally arrive at this realization without proper training, as in [29]. This underscores the necessity for an instructional framework to guide participants in offering demonstrations of both high quality and optimal efficiency, echoing the emphasis of previous research [1].

7 CONCLUSION AND FUTURE WORK

Toward our goal of achieving efficient robot learning and generalization, we proposed using information entropy as a criterion for guiding human teachers to provide the next demonstration that achieves the highest efficiency. This proposed approach reduces the teaching dimension by suggesting the minimum number of demonstrations that achieve the highest learning efficiency. The approach was validated in two experiments to explore its contribution to both

efficient robot learning and generalization, as well as improving novices' teaching skills. We found that the proposed approach significantly improves robot learning efficiency compared to the state-of-the-art heuristic rule. Moreover, information entropy was integrated into two training schemes and tested in a user study. The results show a significant improvement in robot learning efficiency after training in both groups.

We believe that the results of this paper open up several directions for future research. The significant improvement in efficiency by merely guiding users to strategically distribute the demonstrations suggests that guiding users to provide high-quality demonstrations, as in [29], along with their good distribution, could further boost learning efficiency. It would also be interesting to test the proposed approach in facilities with real users without controlling the conditions. For instance, users would have the freedom to decide how long they need guidance to ensure they provide the most beneficial demonstrations to the robot.

REFERENCES

[1] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57, 5 (2009), 469–483.

[2] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*. 41–48.

[3] Aude G Billard, Sylvain Calinon, and Rüdiger Dillmann. 2016. Learning from humans. *Springer handbook of robotics* (2016), 1995–2014.

[4] Erik A Billing and Thomas Hellström. 2010. A formalism for learning from demonstration. *Paladyn, Journal of Behavioral Robotics* 1, 1 (2010), 1–13.

[5] Julian Brooke. 2006. SUS - A quick and dirty usability scale.

[6] Maya Cakmak, Crystal Chao, and Andrea L Thomaz. 2010. Designing interactions for robot active learners. *IEEE Transactions on Autonomous Mental Development* 2, 2 (2010), 108–118.

[7] Maya Cakmak and Manuel Lopes. 2012. Algorithmic and human teaching of sequential decision tasks. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

[8] Maya Cakmak and Leila Takayama. 2014. Teaching people how to teach robots: The effect of instructional materials and dialog design. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 431–438.

[9] Maya Cakmak and Andrea L Thomaz. 2012. Designing robot learners that ask good questions. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 17–24.

[10] Maya Cakmak and Andrea L Thomaz. 2014. Eliciting good teaching from humans for machine learners. *Artificial Intelligence* 217 (2014), 198–215.

[11] Sylvain Calinon. 2016. A tutorial on task-parameterized movement learning and retrieval. *Intelligent service robotics* 9, 1 (2016), 1–29.

[12] Sylvain Calinon and Aude Billard. 2007. Incremental learning of gestures by imitation in a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. 255–262.

[13] Sylvain Calinon and Aude G Billard. 2007. What is the teacher's role in robot programming by demonstration?: Toward benchmarks for improved learning. *Interaction Studies* 8, 3 (2007), 441–464.

[14] Sonia Chernova and Andrea L Thomaz. 2014. Robot learning from human teachers. *Synthesis lectures on artificial intelligence and machine learning* 8, 3 (2014), 1–121.

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[16] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[17] Sally A Goldman and Michael J Kearns. 1995. On the complexity of teaching. *J. Comput. System Sci.* 50, 1 (1995), 20–31.

[18] Dylan Hadfield-Menell, Stuart J Russell, Pieter Abbeel, and Anca Dragan. 2016. Cooperative inverse reinforcement learning. *Advances in neural information processing systems* 29 (2016).

[19] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[20] Sandy H Huang, David Held, Pieter Abbeel, and Anca D Dragan. 2019. Enabling robots to communicate their objectives. *Autonomous Robots* 43, 2 (2019), 309–326.

[21] Michael Kelly, Chelsea Sidrane, Katherine Driggs-Campbell, and Mykel J Kochenderfer. 2019. Hg-dagger: Interactive imitation learning with human experts. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 8077–8083.

[22] Faisal Khan, Bilge Mutlu, and Jerry Zhu. 2011. How do humans teach: On curriculum learning and teaching dimension. *Advances in neural information processing systems* 24 (2011).

[23] Bernard C Kress and William J Cummings. 2017. Towards the Ultimate Mixed Reality Experience: HoloLens Display Architecture Choices. In *SID Symposium Digest of Technical Papers*, Vol. 48. 127–131.

[24] James J Kuffner and Steven M LaValle. 2000. RRT-connect: An efficient approach to single-query path planning. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, Vol. 2. IEEE, 995–1001.

[25] Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. 2021. What matters in learning from offline human demonstrations for robot manipulation. *arXiv preprint arXiv:2108.03298* (2021).

[26] Anahita Mohseni-Kabir, Charles Rich, Sonia Chernova, Candace L Sidner, and Daniel Miller. 2015. Interactive hierarchical task learning from a single demonstration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 205–212.

[27] Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, and Jan Peters. 2018. An algorithmic perspective on imitation learning. *arXiv preprint arXiv:1811.06711* (2018).

[28] Affan Pervez and Dongheui Lee. 2018. Learning task-parameterized dynamic movement primitives using mixture of GMMs. *Intelligent Service Robotics* 11, 1 (2018), 61–78.

[29] Maram Sakr, Martin Freeman, H F Machiel Van der Loos, and Elizabeth Croft. 2020. Training Human Teacher to Improve Robot Learning from Demonstration: A Pilot Study on Kinesthetic Teaching. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 800–806.

[30] Maram Sakr, Zexi Jesse Li, H. F. Machiel Van der Loos, Dana Kulix0107;, and Elizabeth A. Croft. 2022. Quantifying Demonstration Quality for Robot Learning and Generalization. *IEEE Robotics and Automation Letters* 7, 4 (2022), 9659–9666. https://doi.org/10.1109/LRA.2022.3191950

[31] Jeff Sauro and James R. Lewis. 2012. *Quantifying the User Experience*. Morgan Kaufmann. https://doi.org/10.1016/C2010-0-65192-3

[32] Aran Sena and Matthew Howard. 2020. Quantifying teaching behavior in robot learning from demonstration. *The International Journal of Robotics Research* 39, 1 (2020), 54–72.

[33] Claude Elwood Shannon. 2001. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review* 5, 1 (2001), 3–55.

[34] Aaron P Shon, Deepak Verma, and Rajesh PN Rao. 2007. Active imitation learning. In *AAAI*. 756–762.

[35] Jonathan Spencer, Sanjiban Choudhury, Matthew Barnes, Matthew Schmittle, Mung Chiang, Peter Ramadge, and Siddhartha Srinivasa. 2020. Learning from interventions: Human-robot interaction as both explicit and implicit feedback. In *16th Robotics: Science and Systems, RSS 2020*. MIT Press Journals.

[36] Astrid Weiss, Judith Igelsbock, Sylvain Calinon, Aude Billard, and Manfred Tscheligi. 2009. Teaching a humanoid: A user study on learning by demonstration with hoap-3. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 147–152.

[37] Xiaojin Zhu. 2015. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.

[38] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. 2008. Maximum entropy inverse reinforcement learning.. In *Aaai*, Vol. 8. Chicago, IL, USA, 1433–1438.