Deep Reinforcement Learning to Optimize Task Performance in Human-Robot Co-Manipulation

by

Berk Güler

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Mechanical Engineering

April 17, 2023

Deep Reinforcement Learning to Optimize Task Performance in Human-Robot Co-Manipulation

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Berk Güler

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Çağatay Başdoğan (Advisor)

_____

Assist. Prof. Özgür S. Oğuz

_____

Assist. Prof. Yusuf Aydın

Date: _____

To all my beloved ones

ABSTRACT

Deep Reinforcement Learning to Optimize Task Performance in Human-Robot Co-Manipulation
Berk Güler
Master of Science in Mechanical Engineering
April 17, 2023

We propose a two-layer machine learning (ML) approach, which utilizes an artificial neural network (ANN) model as a precursor for a deep reinforcement learning (DRL) model, to optimize task performance during human-robot co-manipulation of heavy objects. In the first layer, the ANN model estimates the human intention to accelerate or decelerate the object (which precedes the actual acceleration or deceleration of the object due to its large inertia). This probabilistic estimation is then used to calculate the gain of an adaptive admittance controller, which alters the robot's contribution to the task. In the second layer, the DRL model fine-tunes this gain and optimizes the task performance by minimizing the jerk in movement and the physical effort made by human. Since online training of a DRL model for a physical human-robot interaction (pHRI) task is highly time-consuming and can potentially be dangerous for the human due to abrupt changes in controller gain, a data-driven human force model was developed by a conditional variational auto-encoder (C-VAE) for offline training of the DRL model via simulations. For this purpose, experimental data was collected from six subjects under 3 different fixed gains of the admittance controller (minimum, nominal, and maximum) to train and validate the DRL model. The adaptive admittance gain profiles generated by the ANN model alone and the proposed two-layer approach (ANN + DRL) were compared through co-manipulation simulations. The results show that the gain profile obtained by the two-layer approach leads to a decrease in human effort and jerk compared to the initial profile provided by the ANN model.

ÖZETÇE

İnsan-Robot Birlikte Manipülasyonunda, Görev Performansını Optimize Etmek için Derin Pekiştirmeli Öğrenme

Berk Güler

Makine Mühendisliği, Yüksek Lisans

17 Nisan 2023

Ağır nesnelerin insan-robot birlikte manipülasyonu sırasında görev performansını optimize etmek için derin pekiştirmeli öğrenme (DPL) modelinin öncüsü olarak bir yapay sinir ağı (YSA) modeli kullanan iki aşamalı bir makine öğrenimi yaklaşımı öneriyoruz. İlk aşamada, YSA modeli, insanın nesneyi hızlandırma veya yavaşlatma niyetini tahmin eder (büyük eylemsizliği nedeniyle nesnenin gerçek hızlanmasından veya yavaşlamasından önce gelir). Bu olasılıksal tahmin daha sonra robotun göreve katkısını değiştiren uyarlanabilir bir giriş kontrolörünün kazancını hesaplamak için kullanılır. İkinci aşamada, DPL modeli bu kazanca ince ayar yapar ve hareketteki sarsıntıyı ve insan tarafından harcanan fiziksel çabayı en aza indirerek görev performansını optimize eder. Fiziksel insan-robot etkileşimi görevi için bir DPL modelinin çevrimiçi eğitimi oldukça zaman alıcı olduğundan ve kontrolör kazancındaki ani değişiklikler nedeniyle insan için potansiyel olarak tehlikeli olabileceğinden, DPL modelinin simülasyonlar yoluyla çevrimdışı eğitimi için koşullu varyasyonel otomatik kodlayıcı ile veriye dayalı bir insan kuvveti modeli geliştirilmiştir. Bu amaçla, DPL modelini eğitmek ve doğrulamak için admitans kontrolörünün 3 farklı sabit kazancı (minimum, nominal ve maksimum) altında altı denekten deneysel veriler toplanmıştır. Tek başına YSA modeli ve önerilen iki aşamalı yaklaşım (YSA + DPL) tarafından üretilen uyarlanabilir admitans kazanç profilleri, birlikte manipülasyon simülasyonları aracılığıyla karşılaştırılmıştır. Sonuçlar, iki aşamalı yaklaşımla elde edilen kazanç profilinin, YSA modeli tarafından sağlanan başlangıç profiline kıyasla insan çabasında ve sarsıntıda bir azalmaya yol açtığını göstermektedir.

ACKNOWLEDGMENTS

I would like to extend my sincerest appreciation to my wife for being a constant source of strength and encouragement throughout my struggles. Her unwavering support has enabled me to overcome many challenges and obstacles. I am also immensely thankful to my family for their unwavering support and guidance. In particular, my sister has been a pillar of strength, providing me with invaluable motivation and free counseling.

Furthermore, I cannot express enough gratitude to all my friends who have helped me along the way. Their support, encouragement, and assistance were instrumental in my ability to graduate. I am forever grateful to each and every one of you for your kindness and generosity. Without your help and support, it would have been much more difficult to achieve my goals. Thank you from the bottom of my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ANN	Artificial Neural Network
C-VAE	Conditional Variational Autoencoder
DL	Deep Learning
DRL	Deep Reinforcement Learning
LSTM	Long Short-term Memory
ML	Machine Learning
pHRI	physical Human-Robot Interaction
RL	Reinforcement Learning

Chapter 1

INTRODUCTION

1.1 Motivation

Although there are already several studies in the literature on human-robot interaction by focusing on co-existence (human and robot do not share a workspace) and co-operation (human and robot share a workspace but are not in physical contact), the number of studies on collaboration (human and robot share a workspace and are also in physical contact) are still much less. However, many collaborative tasks require continuous physical interaction between human and robot. The most common form of physical interaction between human and robot today is the "hand guidance" by human in which the robot is a passive partner. The operator simply pushes the robot by hand and brings its end-effector to a set of desired locations, which are recorded to construct a continuous trajectory for it to follow later autonomously. This eliminates the need for a programmer to define the trajectory explicitly in a software code as in the case of industrial robots.

One of the most popular applications of hand-guidance is collaborative object manipulation. For example, in manufacturing, the operator guides the robot by hand to pick the object from a rack with the help of an end-effector such as a vacuum gripper and then place it on the assembly line. Once the trajectory for the manipulation task is recorded by hand-guidance, it is performed by the robot autonomously without any human intervention or with minimal intervention. During the execution phase of this so-called "learned" task, a motion controller (such as PID) can be utilized to make the robot follow the desired force/displacement trajectory defined by the hand guidance. During the hand guidance phase, the dynamics of interactions between human and robot is regulated by either an impedance or

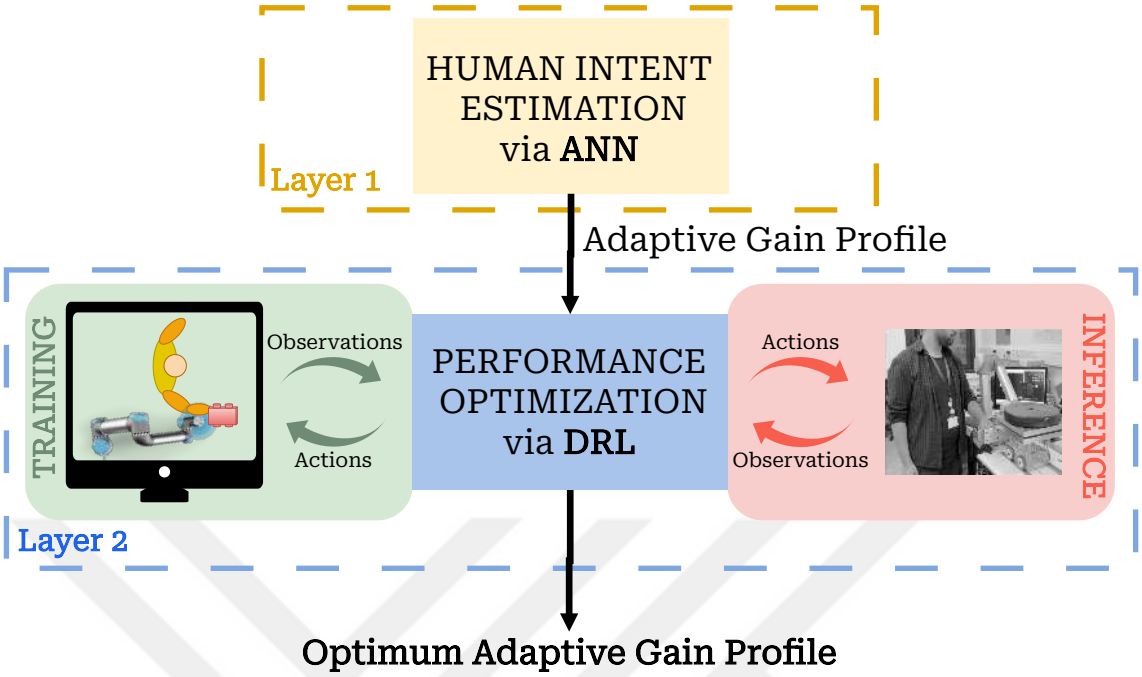

Figure 1.1: The proposed two-layer ML approach for optimizing task performance during human-robot co-manipulation of heavy objects. In the first layer, human intention to accelerate/decelerate the object is detected by an ANN model to adjust the gain of an admittance controller that regulates the physical interactions between human and robot. In the second layer, the admittance gain is further adjusted to optimize the task performance.

admittance controller. In impedance (admittance) control, the displacement (force) is applied by the human to the robot and force (displacement) is returned. Since force/torque sensing (either via an external sensor or by estimating it from the motor currents) is readily available in collaborative robots, admittance control is a more viable option for physical interaction with robots that are motion controlled.

The hand guidance approach discussed above is more appropriate for applications with clear manipulation plans executed in a structured environment exhibiting minimal uncertainties. However, it is important to emphasize that the robot in this approach is passive and simply follows the recorded trajectory. Hence, it does not demonstrate any adaptive behavior during the task. If, for example, the target location changes, the new trajectory must be taught by the human operator again. Moreover, it does not adapt its contribution to the task based on the intentions of human operator.

In some other object manipulation scenarios taking place in unstructured environments involving more uncertainties, the robot must be equipped with a higher level of adaptation to actively contribute to the task. For example, in collaborative manipulation of light but bulky objects or heavy objects (e.g. collaborative transportation of a table at home, patient bed in hospital, car windshield in factory), the robot needs to predict human intention to comply with her/his behavior. Though looking trivial from the outside, understanding human intention is a challenging endeavor and, in our opinion, requires a multi-layered AI architecture to implement properly. For example, at a lower layer, this could be simply the instantaneous detection of human movement intention to translate (such as pull/push) and/or rotate the manipulated object while it could be the detection of interaction patterns of dyad (human and robot) at a higher layer such as whether they work together in harmony or conflict with each other [Al-Saadi et al., 2023b].

Moreover, since both the human and robot must contribute to the task actively for its successful execution, a shared control architecture for role exchange and load allocation is also important, i.e., adjustable autonomy. This architecture determines what roles the partners (human and robot) should take and how much they should contribute to the task during its execution (see a review in [Losey et al., 2018]).

Role exchange can be implemented by assigning leader and follower roles to each partner and then updating those roles according to the forces exchanged between them [Oguz et al., 2010, Kucukyilmaz et al., 2012, Mörtl et al., 2012]. Alternatively, Oguz et al. [Oguz et al., 2012] utilized a game theoretic approach to assign three different roles for the robot, namely competitive, concessive, and retaliative. They, then investigated the utility of these roles and their effect on the quality and end-result of interaction between human and robot using a shared control architecture.

In order to adjust the contribution of human and robot to the task during pHRI, a load-sharing mechanism and a variable (adaptive) interaction controller are necessary since the different phases of the task require different amounts of contribution from the robot for the efficient execution of the task. A standard impedance/admittance controller with fixed parameters cannot accommodate the changes in human intent for different phases of a pHRI task. The need for de-

tecting the human intent during pHRI using the sensed data in order to adapt the parameters of an interaction controller has been already recognized early on. For example, Duchaine and Gosselin [Duchaine and Gosselin, 2007] utilized the derivative of the force applied by human and the velocity of the manipulated object together to predict whether human intends to accelerate or decelerate the object. The admittance damping was decreased (increased) when the human intention was estimated as acceleration (deceleration). Hamad et al. [Hamad et al., 2021] implemented a fuzzy-based approach with end-effector velocity of robot, force applied by human, and its derivative as the inputs to alter the admittance gain in a smooth manner in order to increase/decrease the contribution of robot to the task.

1.1.1 Related Work

As obvious from the discussion above, a pHRI task that requires a proactive robot is significantly more challenging to implement since the robot must make its own decisions based on human intent in real-time using the kinetic and kinematic data acquired by sensors. The rule-based approaches mentioned above may not work well since such variables are highly task-specific, and their values depend on the characteristics of the human operator and the environment even for the same task. Moreover, developing a set of rules to detect human intent becomes highly challenging as the collaborative task becomes more complex.

With the introduction of machine learning (ML) techniques into robotics, the quality of pHRI can be improved significantly and brought to the level of human-human interaction in the future. In particular, the use of ML techniques in pHRI will enable us to estimate not only the human behavioral intent but also the dyadic interaction patterns of the human and robot as a team. Data-driven techniques such as neural networks, Bayesian networks, and Gaussian Mixture Models can capture the uncertainties in human behavior better than the model-driven approaches proposed in earlier studies.

Compared to the studies on human intent/behavior prediction, the number of studies on detecting the dyadic interaction patterns of human-robot team is

highly limited. Due to the coupled dynamics of pHRI, predicting the human intent/behavior alone is not sufficient for designing effective interaction controllers. One sensory modality that emerges in this regard is haptics. Earlier studies have already demonstrated that haptics play a critical role in detecting interaction patterns in collaborative manipulation tasks involving human dyads using ML techniques [Madan et al., 2015, Townsend et al., 2017, Al-Saadi et al., 2021, Al-Saadi et al., 2023a]. For example, Al-Saadi et al. [Al-Saadi et al., 2021] conducted collaborative object manipulation experiments with human dyads and showed that a set of features derived from force/torque data alone is sufficient for the successful classification of dyadic interaction behaviors using a Random Forest Classifier.

Based on the existing studies in the literature, the applications of ML techniques (supervised, unsupervised, semi-supervised, and reinforcement learning) to pHRI can be grouped into two: a) learning human intent/behavior, and b) performance optimization. The studies in the first group utilize the sensed kinetic and kinematic data to learn human intent/behavior in real-time during the execution of a collaborative task so that the parameters of an interaction controller can be adjusted to comply with human intent/behavior. For example, Li and Ge [Li and Ge, 2014] trained an ANN model with interaction force, position and velocity of the robot arm as the inputs to estimate the position trajectory of the human arm and then adjust the parameters of an impedance controller to follow this trajectory. As a result, the robot actively moves toward human partner's intended position rather than passively comply to the interaction force, which improves the collaboration efficiency. Sharkawy et al. [Sharkawy et al., 2020] trained an ANN model online for a collaborative manipulation task and adjusted the admittance mass of the interaction controller to obtain a manipulation trajectory that is close to the minimum-jerk trajectory. Guler et al. [Guler et al., 2022] trained an ANN model to detect subtasks during collaborative drilling performed with a robot to adjust the admittance damping. They defined 3 subtasks as idle, free motion, and contact and used low damping during the free motion phase to increase transparency (which resulted in 20% lower human effort) and high damping during the contact phase to improve the stability (which resulted in 25% lower oscillations of the drill bit). This

trade-off between transparency (minimal resistance to human) and stability during contact has been investigated extensively in our earlier pHRI studies [Aydin et al., 2018, Aydin et al., 2020].

The second group of studies on pHRI aims to optimize the task performance based on some cost function. One of the featured ML approaches in that regard is Reinforcement Learning (RL). RL models a decision-making process based on a set of states, actions, and rewards. Actions are selected based on a policy, which aims to maximize the cumulative sum of rewards (i.e. minimize cost function). The application of RL to pHRI is challenging for the following reasons; first, there are uncertainties in human behavior and the resulting dynamics of the coupled (human-robot) system. In other words, since the human and robot are in continuous contact in our co-manipulation scenario, the physical behavior of human affects that of the robot which in turn alters the human behavior again and the loop goes on. Hence, selecting the best actions to minimize the cost function (maximize the reward function) for such a coupled system whose dynamics is ever-changing is not trivial. For this reason, the number of applications of RL to such non-stationary problems that involve epistemic uncertainty has been limited so far. Second, if such a minimum cost exists, RL requires a large number of episodes (i.e. number of optimization iterations) to reach that minimum (also known as the curse of dimensionality problem). This is a serious problem in pHRI utilizing an adaptive (variable) interaction controller since the abrupt changes in the parameters of the controller to seek an optimum solution may render the robot unstable and hence jeopardize human safety. Therefore, it is not practical to run many training episodes in pHRI applications involving continuous interaction between human and robot. Third, defining an effective cost function to minimize task performance in pHRI is not trivial. The cost function should provide a good balance between two or more competing metrics of performance. For example, in collaborative object manipulation, these metrics could be the effort made by human and the smoothness of the manipulation trajectory. The former requires a more compliant robot for minimal resistance to the motion while the latter requires the opposite for a less jerky motion trajectory.

The RL approaches used for pHRI can be grouped into two: a) data-based (i.e.

model-free), b) model-based. Ghadirzadeh et al. [Ghadirzadeh et al., 2016] developed a Q-learning approach so that the human operator collaborating with a robot can keep a ball rolling on a beam at a desired target location with minimum effort. The uncertainties in the pHRI task due to the human behavior are modeled by Gausian process and handled by Bayesian optimization. They selected a policy in adjusting the impedance of the controller such that the interaction force is minimized during the task. Dimeas and Aspragathos [Dimeas and Aspragathos, 2015] proposed a variable admittance controller for human-robot co-manipulation based on RL. In order to cope with the curse of the dimensionality problem, they utilized a fuzzy Q-learning (FQL) algorithm for the training of the RL agent. They defined a cost function to adjust the admittance damping of the interaction controller during the manipulation such that the jerk in the movement was minimized. Lu et al. [Lu et al., 2020] also used an FQL algorithm in conjunction with a long short-term memory (LSTM) model to predict human motion intention based on a simplified dynamical model of human arm and then adjust the damping of a variable admittance controller accordingly while the dyad follows a minimum jerk trajectory for point-to-point collaborative movement. Wu et al. [Wu et al., 2019] suggested shared impedance control for collaborative transportation of a rigid object. RL was used to estimate the impedance parameters for the dyad that optimizes a cost function defined by the motion error and the control inputs. Li et al. [Li et al., 2017] proposed an adaptive impedance controller for optimized human-robot cooperation. The optimization problem was formulated as a linear quadratic regulator (LQR), and RL is employed to find the optimized parameters of the impedance controller. Roveda et al. [Roveda et al., 2020] used a model-based RL approach and adjusted the parameters of a variable impedance controller during collaborative object lifting in order to minimize human effort. Since the task was relatively simple, 30 trials were sufficient to train the robot to learn how much to compensate for gravity during the lifting. Modares [Modares et al., 2016] proposed a control architecture with two control loops for pHRI. The inner loop was designed to make the robot dynamics behave like an impedance model while the outer loop was used to calculate the optimal parameters of this impedance model using a RL approach. Although computer

simulations were performed with a two-link arm to show the benefits of the proposed control architecture, only a simple physical demonstration of point-to-point collaborative movements performed with an actual robot was reported. Whitsell and Artemiadis [Whitsell and Artemiadis, 2017] developed an admittance controller that allows human and robot to cooperate in 6-dof Cartesian space. A RL algorithm, taking force and torque as inputs, was utilized to decrease the estimated mechanical power applied by the human to exchange leader/follower roles with the robot.

1.2 Contributions and Novelties

In this paper, we propose a two-layer approach utilizing ML for human-robot co-manipulation of heavy objects (Fig. 1.1). We first estimate the human intention to accelerate or decelerate the object by an ANN using velocity and human force as the input features. Using the probabilistic value returned by the ANN model for this estimation, the gain of an adaptive admittance controller is adjusted in real-time to alter the contribution of the robot to the task. Starting from a nominal value, the gain is increased smoothly during the acceleration phase to reduce the human effort and decreased smoothly during the deceleration phase to help with the parking accuracy by reducing the jerk. Then, a deep reinforcement learning (DRL) approach is utilized to fine-tune this gain in order to optimize the task performance by minimizing the jerk in the movement and the effort made by the human. The contributions of the proposed approach to the current literature are listed below.

- We show that an ANN algorithm can successfully estimate the human intent to accelerate/decelerate a manipulated heavy object in real-time during a co-manipulation task before the object actually accelerates/decelerates.

- A data-driven model for human force response was developed using a Conditional Variational Encoder (C-VAE) model for the simulation of co-manipulation task. The training of our DRL model was performed in a simulated environment using the developed model, and the resulting agent was deployed on the real robot. This approach enables more efficient and safe RL training, while

maintaining a high degree of realism and accuracy in the learned behavior.

- We show that the proposed two-layer approach (ANN+DRL) can return an optimal profile for the admittance gain once an initial gain profile is provided. Compared to this initial profile provided by ANN in our case, which can be used to adjust the admittance gain as a standalone approach without any optimization, the gain profile optimized by the proposed two-layer approach led to significant reductions in human effort and jerk.

1.3 Outline of the Thesis

Chapter 2, Approach, presents our methods which include the control architecture, ANN model, DRL model, and modeling of human force response. In Chapter **??**, Experiments, we detail our experimental design and procedures. For this, we invited six participants to define the goal, derive the cost function, and train both the ANN model and human force response model. Chapter 4 explains the results of our experiments and simulations. We analyze performance metrics and define cost parameters for our proposed controllers. In Chapter 5, we discuss our proposed approach for the co-manipulation task. Finally, Chapter 6 concludes this thesis with a brief summary of our findings and outlines future research directions.

Chapter 2

APPROACH

2.1 Control Architecture for pHRI

We utilize an adaptive admittance controller to adjust the contribution of the robot to the pHRI task based on the acceleration/deceleration intent of human operator. Fig. 2.1 shows our control and learning architecture for the admittance adaptation.

Based on this diagram, human exerts a force to move a heavy object with a reference velocity υ_{ref}. The force applied by human F_h is measured by a force sensor, amplified by a gain (scaling factor) G_H and then sent to the admittance controller $Y(s)$ of robot to adjust its contribution to the task. The admittance controller outputs a reference velocity υ_{ref} for the internal motion controller of robot. The motion controller of robot transmits sufficient torques to its joints to move the manipulated object with a velocity of υ. Here, $G(s) = \upsilon(s)\ /\upsilon_{ref}(s)$ is the transfer function for the robot, which is available in Aydin et al. [Aydin et al., 2018].

The admittance controller can be written as follows:

$$Y(s) = \frac{\upsilon_{\text{ref}}(s)}{F_h(s)} = \frac{1}{ms + b} \tag{2.1}$$

where m and b represent the admittance mass and damping, respectively. Eq. 2.1 can be rearranged as,

$$Y(s) = \frac{K}{\tau s + 1} \tag{2.2}$$

where, $K = 1/b$ is the admittance gain and $\tau = m/b$ is the admittance time constant. In this study, we alter the admittance gain based on human intent to adjust the contribution of the robot to the task while keeping the time constant of the controller unchanged. Note that increasing (decreasing) the admittance gain decreases (increases) the contribution of the robot to the task.

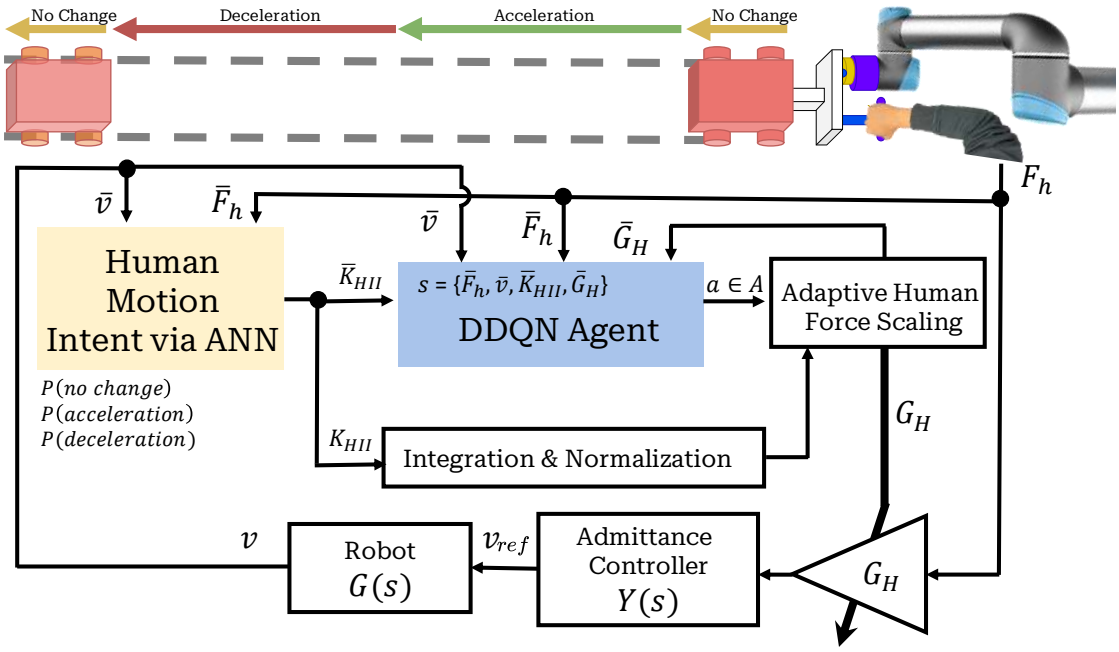

Figure 2.1: The proposed control architecture for human-robot co-manipulation of heavy objects. Human intention to accelerate/decelerate the manipulated object is estimated by an ANN model to adjust the gain of an adaptive admittance controller of the robot, which is later fine-tuned by a DRL model (i.e. DDQN Agent) for performance optimization. The admittance controller takes human force as the input and outputs a desired velocity for the manipulated object to be followed by the internal motion controller of the robot. Note that the input variables to the ANN and DRL models are time series data and represented with an overbar to distinguish them from the same variables that represent a single time step.

Hence, Eq. 2.2 can be rewritten as follows:

$$G_H \cdot Y(s) = \frac{G_E}{\tau s + 1} \tag{2.3}$$

where $G_E = G_H K$ is the effective admittance gain of the controller. Hence, the contribution of the robot to the task can be adjusted by changing the scaling factor G_H without affecting the first-order dynamics of the admittance controller. The maximum value of G_H is determined based on the stability analysis in Hamad et. al. [Hamad et al., 2021].

2.2 Human Intent Estimation via ANN

Collaborative transportation of a heavy object with a robot involves three major phases: a) starting, b) driving, and c) parking. The human operator starts the motion by pushing the object with the intent to accelerate it toward the target location. As the object gets closer to the target, the human intends to decelerate it to park at the target location. In order to adjust the contribution of the robot to the task for more efficient collaboration, the human intention to accelerate/decelerate the object should be detected before the actual acceleration/deceleration of the object takes place since it occurs with a delay due to the large inertia of the manipulated object. During the driving phase, the robot should increase its contribution to the task to reduce the human effort. On the other hand, during the parking phase, maximizing the precision is more critical than reducing human effort and hence, the robot should display more resistance to the velocity of movement, enabling the human to decelerate the object in a less jerky manner. Based on the discussion above, one can carefully design a profile for the scaling factor G_H, which is multiplied with the admittance gain K to obtain the effective admittance gain G_E and hence adjust the contribution of the robot to the co-manipulation task based on its different phases (starting, driving, parking). Such a profile (Fig. 2.2) has been already suggested by Hamad et al. [Hamad et al., 2021] in our earlier study. The scaling factor starts from a nominal value G_H^{nom} and increases slowly reaching its highest possible value at G_H^{max}. In the driving phase, the dissipative (i.e., resistive) behavior of robot should be avoided as human desires to accelerate the object easily. At the end of this phase, human intends to decelerate the object (note that since the object has a large inertia, it may continue to accelerate for a while depending on the force applied by the human) to park it accurately while avoiding jerky oscillations around the target. From this point on (G_H^{max}), the dissipation capacity of the controller should be increased by reducing the scaling factor to a value lower than the nominal one (G_H^{min}).

We utilize an Artificial Neural Network (ANN) to estimate human movement intention. In this context, detecting human intent is considered a time-series clas-

sification problem. As typical in time-series problems, our inputs are sequences of data collected from previous time steps (velocity and human force) using a sliding time window of 40-time steps, and our output is the predicted intent (*accelerate*, *decelerate*, *no change*) and the associated probability values at the current time step. As shown in Eq. 2.4, the probability value for deceleration is subtracted from that of the acceleration to generate "Human Intention Index" (K_{HII}), which takes a value between -1 and 1.

$$K_{HII} = \mathbb{P}(acceleration) - \mathbb{P}(deceleration) \tag{2.4}$$

In order to obtain the desired profile shown in Fig. 2.2, the instantaneous values of K_{HII} are first integrated over time, then bounded by a sigmoid function, and finally normalized to obtain the scaling factor G_H as a function of time. The related mathematical relations can be found in [Hamad et al., 2021].

2.3 Performance Optimization via DRL

In our study, DRL is utilized to further adjust the effective gain of the admittance controller for optimizing the task performance by minimizing the jerk in the movement and the human effort (defined as the human force multiplied by the velocity of the manipulated object).

In RL, the agent interacts with an environment. At time t, the agent observes a state s_t and takes a discrete action a_t from a finite set of legal actions A under a policy π. In our case, the states are the velocity of the object (υ), human force (F_h), scaling factor for admittance gain (G_H), and the human intention index (K_{HII}). All of these variables are represented using a sliding time window of 10-time steps. The discrete actions are to increase or decrease human intention index K_{HII} by multiplying it with the following factors $A = \{5, 2, 1, 0.5, 0.2\}$. This action execution leads to a state transition from s_t to s_{t+1} and a scalar reward $r(s_t, a_t) \in R$. The ultimate goal of the agent is to learn the policy π that maximizes the expected total return.

$$R(s_t, a_t) = [\sum_{t'=t}^{T} \gamma^{(t'-t)} r(s_{t'}, a_{t'})] \tag{2.5}$$

where T and $\gamma \in (0, 1]$ correspond to the terminal time and discount factor, respectively. In our case, the reward function to be maximized is defined as

$$r(s_t, a_t) = -[w\mathrm{P}_t + (1 - w)\mathrm{J}_t] \tag{2.6}$$

where, P_t is the normalized effort made by the human, J_t is the normalized jerk in the movement of the cart, and w is the weighting factor. The parameters of the reward function, J_t and P_t, are selected based on the analysis performed on the data collected under the controllers utilizing three fixed scaling factors (G_H^{min}, G_H^{nom}, G_H^{max}), see Section 4.2.2. To train a policy that balances the trade-off between minimum jerk and human effort, we define a weighted sum of these two objectives as our reward function. We calculate the cumulative sum of human effort and jerk after every 40-time steps of a trial (episode) to provide intermediate rewards for the DRL algorithm in order to improve the overall task performance. The choice of 40-time steps for providing intermediate rewards is motivated by the following factors; First, giving rewards accelerates the learning process by providing more frequent feedback to the agent, but this should not be done at the frequency of the control loop, as this would be computationally expensive. Second, since the RL actions are applied to the human intention index K_{HII}, which is integrated over time to obtain the scaling factor G_H, a longer time interval is needed to observe the effect of those actions on the scaling factor. Therefore, the selected interval (40-time steps) provides a good balance between accelerating learning and managing computational costs while allowing sufficient time for the RL actions to affect the scaling factor.

To ensure that our reward function is properly balanced, we normalize the parameters J_t and P_t after every 40 time steps of a trial. Normalizing the parameters at the end of a trial could result in an imbalance between the two objectives. Specifically, a large value of J at the beginning of the co-manipulation task could dominate the final reward calculation, while a small value of P at the end of the task would have little effect on the reward. We ensure that each objective is equally weighted throughout the task by normalizing J and P after every 40 time steps. This normalization process also provides a well-defined range for the weighting factor (w) in

Eq. (2.6).

In this study, we selected Double Deep Q-Learning Network (DDQN) as our DRL model due to its potential to manage non-stationary nature of our application and improve the stability of the learning process [Hasselt et al., 2016]. In particular, our co-manipulation task involves coupled and non-linear dynamics and requires the robot to adapt to different force and velocity profiles that vary over time. Compared to a conventional Deep Q-Learning Network (DQN), DDQN utilizes a separate target network to estimate the optimal action values, which helps to reduce the overestimation bias and stabilize the learning process.

2.4 Modeling Human Force Response

Online training of a DRL model for our pHRI task is highly time-consuming and can be potentially dangerous for human operator due to the abrupt changes in admittance gain during the exploration of action space by the DRL agent.

In order to perform offline co-manipulation simulations utilizing the proposed DRL agent, we developed a data-driven model to generate the force applied by human to the manipulated object. Once the estimated human force is available, the control architecture shown in Fig. 2.1 can be used to simulate the co-manipulation task.

Most of the earlier studies on human behavior modeling in pHRI have focused on the estimation of motion trajectories rather than force response. The earlier studies on human force modeling in HRI followed a) model-driven and b) data-driven approaches [Hiatt et al., 2017]. One popular model-driven approach is to minimize jerk of movement to obtain a point-to-point motion trajectory (i.e. minimum jerk model). Then, a simple mass-spring-damper model could be utilized for the arm dynamics to estimate the model parameters for the given (measured) end-point forces [Rahman et al., 2002]. There are also more sophisticated dynamical models of human arm in the biomechanics literature considering multiple degrees of freedom and EMG measurements for integrating muscle activities [Holzbaur et al., 2005], but they are computationally expensive. Moreover, estimating human force response is

still highly challenging even for a relatively simple pHRI task such as ours due to the coupled and unknown dynamics of human-robot interaction, and the uncertainties in the mass and inertial properties of the manipulated object and its interaction with the environment.

A data-driven model could be an alternative solution since this approach aims to find a nonlinear mapping (functional relation) between pre-defined input and output variables based on the measured data only. Early data-driven approaches for human motion modeling have utilized several techniques. Gaussian Process Regression (GPR) models are often used for continuous motion estimation and prediction [Li et al., 2020], while Neural network-based models are typically used for more complex and high-dimensional motion data captured by vision systems [Formica et al., 2021, Landi et al., 2019, Ravichandar and Dani, 2017]. Principal Component Analysis (PCA) and Gaussian Mixture Model(GMM)-based models are used for dimensionality reduction and clustering of the motion data over a short time horizon [Callens et al., 2020], [Mainprice and Berenson, 2013]. Hidden Markov Model (HMM) and Probabilistic Movement Primitives (ProMP) are commonly used for probabilistic human motion modeling which in turn facilitates segmentation and prediction of movements [Ding et al., 2011], [Gómez-González et al., 2018]. Inverse optimal control is utilized to learn the underlying reward function that explains the observed human motion data [Mainprice et al., 2015]. However, none of those prior studies have focused specifically on modeling human force response in the context of pHRI.

In this study, we present a data-driven approach to model the complex nature of human force response during human-robot interaction where a coupled dynamical behavior is observed. In particular, the model should generate stochastically diverse force profiles for the co-manipulation simulations to train the DRL model.

We utilized a Long-Short Term Memory (LSTM) network for the decoding phase of a C-VAE model to capture the stochastic nature of human force response [Sohn et al., 2015]. While LSTM networks are frequently used to model time-series signals as standalone models, they fail to generate stochastically diverse data if the generated signals do not share the same statistical distribution as the input [Crnkovic-Friis

and Crnkovic-Friis, 2016]. The use of a C-VAE model in tandem with an LSTM network solves this problem and generates diverse force profiles under the same simulated conditions, which is critical for training of a DRL model.

For our C-VAE model, we define the loss function as the summation of reconstruction loss and Kullback–Leibler divergence (D_{KL}). Reconstruction loss measures the difference between the original input and the reconstructed output, while D_{KL} ensures that the learned latent space follows a prior distribution. Hence, our loss function is given as:

$$\mathcal{L} = \mathcal{L}_{\text{Reconstruction-Loss}} + \mathcal{L}_{D_{KL}} \tag{2.7}$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(|F_{hi} - F^*_{hi}|^2 + D_{KL}(Q(\mathbf{z}_i|\mathbf{x}_i)||\mathbb{P}(\mathbf{z})) \right) \tag{2.8}$$

where F_{hi} is the human force for the i-th sample, F^*_{hi} is its reconstruction, N is the total number of samples, $Q(\mathbf{z}_i|\mathbf{x}_i)$ is the variational posterior distribution, $\mathbb{P}(\mathbf{z})$ is the prior distribution. Note that $\mathbf{x}_i$ represents the inputs to the C-VAE for encoding and $\mathbf{z}_i$ is the latent code for the i-th sample of decoding.

During the training phase, we feed our model with a sliding window of prior data corresponding to 40-time steps of human force $\overline{F}_h$, object velocity $\overline{v}$, force scaling gain $\overline{G}_H$, and distance to the target ($\overline{d2t}$) for encoding. The model outputs (estimates) the human force for the next time step (Fig. 2.3). We assumed that the human operator utilizes $d2t$ to regulate her/his force and hence was selected as the conditional feature to learn the hidden representation of input space for decoding. To implement the C-VAE model, we designed an encoder network consisting of 1D convolution layers to map input space to a latent space as shown in Fig. 2.3. The purpose of using these layers is to reduce the dimensionality of the input data, making the model computationally more efficient and less prone to overfitting. Additionally, convolutional layers are useful for recognizing patterns in the input data. For the decoder network, the C-VAE model integrates an LSTM network, which captures long-term dependencies in the latent space and the conditional feature (i.e. $d2t$). LSTM networks are also known for their robustness to noise, which was helpful when we perturb the latent space with Gaussian noise having the $d2t$ as the

main parameter, $\xi(d2t)$. In our approach, the noise is applied to the latent variables from the very beginning of the co-manipulation task. However, as the object approaches to the target position, it decreases exponentially with distance to prevent any parking problems. This approach enables us to generate stochastically different force profiles for the training of our DRL model via end-to-end simulation of the co-manipulation task.

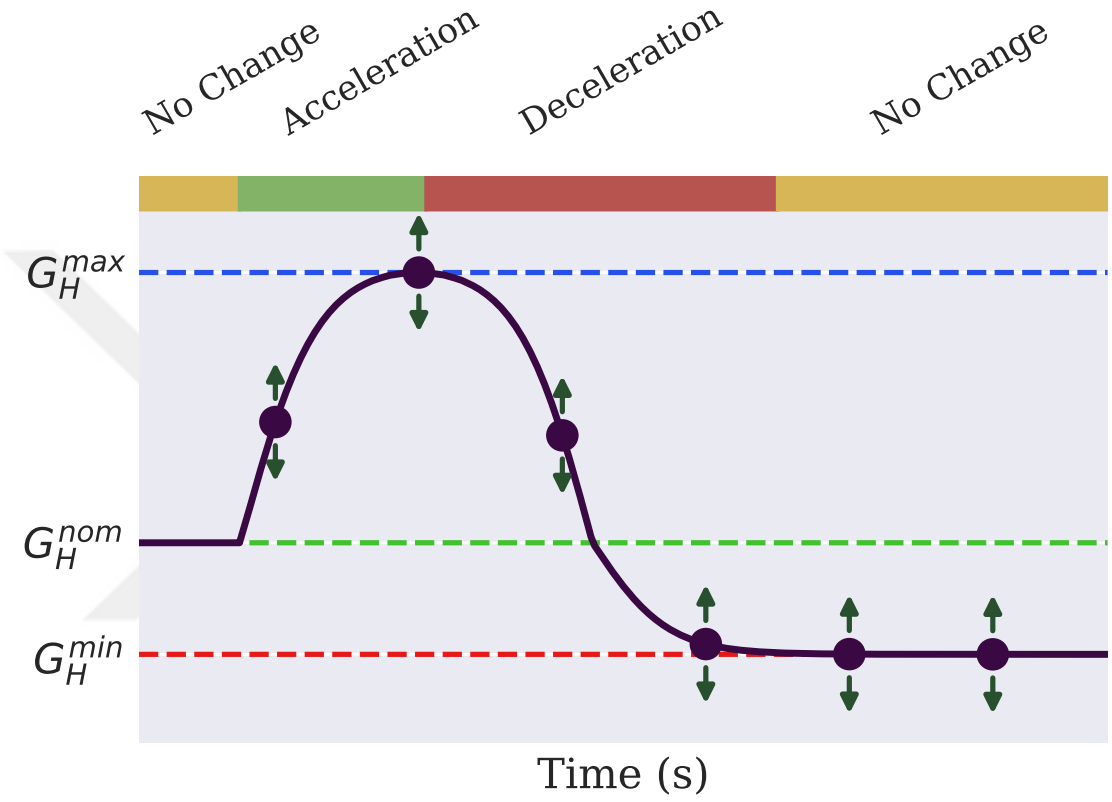

Figure 2.2: The gain profile targeted by the ANN model: the scaling factor G_H is used to adjust the gain of the admittance controller and hence the contribution of the robot to the co-manipulation task. The scaling factor starts from a nominal value (G_H^{nom}) and increases until it reaches a maximum value (G_H^{max}) during the driving phase of the task to accelerate the object. Afterward, it is reduced to a minimum value (G_H^{min}) to assist with the deceleration phase and parking the object. Knot points on the gain profile represent the time stamps when the DRL model is called to take a new action to further adjust the scaling factor (and hence the admittance gain) in order to optimize the task performance. This optimization aims to minimize the jerk in movement and human physical effort simultaneously.

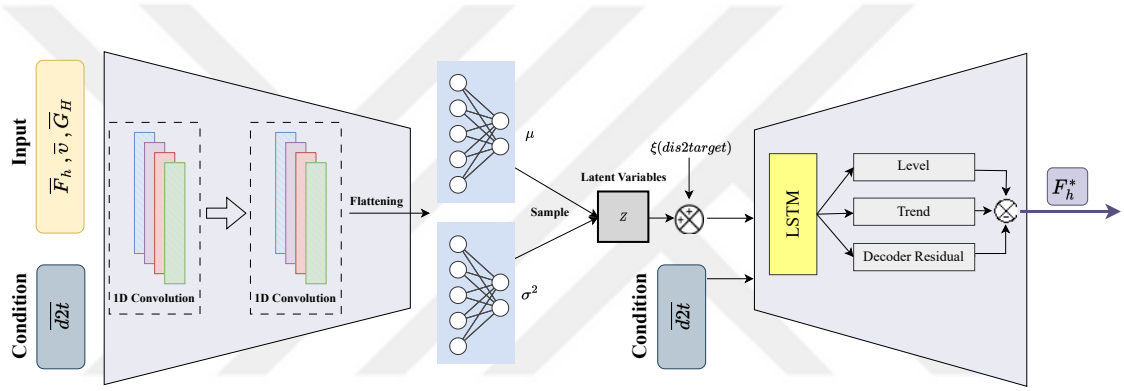

Figure 2.3: Data-Driven Human Force Modeling: In order to train the DRL agent via offline simulations, a data-driven model was developed by using a Conditional Variational Autoencoder (C-VAE) to generate the force applied by human to the manipulated object (F_h^*). Once the estimated human force is available, the control architecture shown in Fig. 1.1 can be used to simulate the co-manipulation task in order to train the DRL agent.

Chapter 3

EXPERIMENTS

All experiments were performed using the pHRI set-up developed by Hamad et al. [Hamad et al., 2021]. The main components of this setup are a collaborative robot (UR5, Universal Robot Inc.), a mobile cart that is rigidly connected to the robot and moves on a rail while carrying a mass of 26 kg, and two force sensors (Mini40, ATI Inc.), each measuring the force applied to the cart separately by subject and robot (Fig. 3.1). The experimental data was collected at 125 Hz, which is also the update rate of the closed-loop control system shown in Fig. 2.1, using a DAQ card (USB-6343, National Instruments Inc.). During the experiments, the instantaneous force applied by human on the cart is scaled by $G_H(t)$ and fed into the interaction controller, which outputs the cart's desired velocity that the internal trajectory controller of the robot to follow. The robot aims to push the cart with this desired velocity until a new force input from the human arrives (see the control architecture in Fig. 2.1).

The experiment was designed to mimic Fitts's reaching movement task [Hamad et al., 2021]. Fitts [Fitts, 1992] proposed a linear relationship between movement time MT and the index of task difficulty ID, $MT = a + bID$, where a and b are the constant coefficients. The index of difficulty (in bits) can be defined as a function of reaching distance D and parking width W as $ID = log2(D/W + 1)$ via the Shannon formula [Soukoreff and MacKenzie, 2004].

In our experiments, subjects were instructed to hold the handle depicted in Fig. 3.1 and move the cart as quickly and accurately as possible from a starting point to a parking zone. Subjects were required to keep the cart stable in the parking zone for 2 seconds. During the experiment, subjects were provided with visual feedback in the form of a moving box on a computer screen, which simply simulated the behavior of cart.

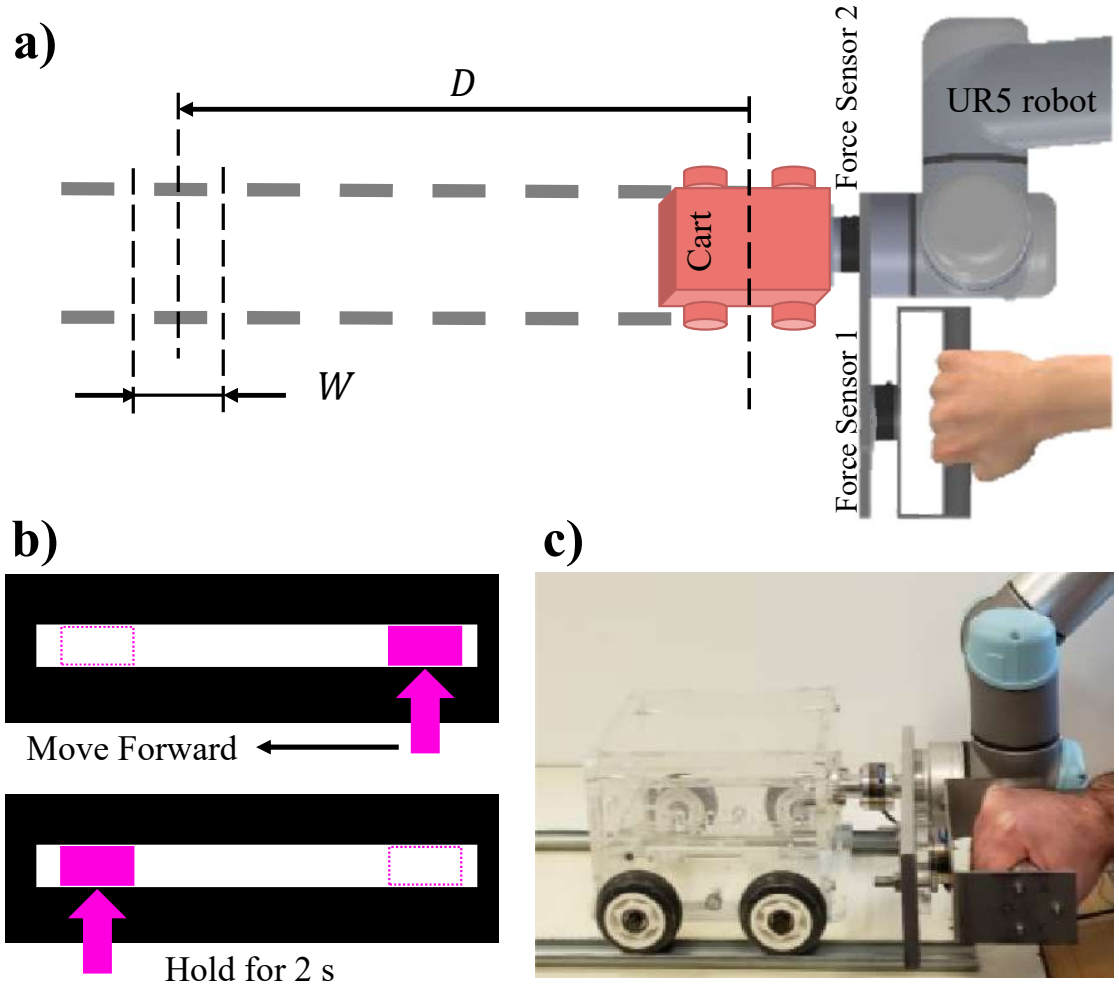

Figure 3.1: he experimental setup [(a), (c)] and visual feedback provided to subjects during the experiment (b). The subjects were asked to push a heavy mass placed in a cart (not shown) forward for a distance of "D" and then park in a region with a width of "W".

3.1 Data Collection for Training

We selected a reaching distance of D = 40 cm and two indices of difficulty (ID = 4, and 6 bits) for training our ANN and DLR models. This selection resulted in 2 different parking widths (W= 2.667 and 0.635 cm). The experimental data (velocity, human force, and robot force) was collected from 6 subjects. The nominal values of the time constant and admittance gain of the admittance controller were set to τ= 0.4 and K= 0.01. The experiments were performed under 3 different scaling factors:

G_H^{min}=0.33, G_H^{nom}= 1.0, G_H^{max}= 2.68. These scaling factors were multiplied by the admittance gain K before transmitted to the robot (see Eq. 2.3). There were a total of 720 trials (6 subjects x 3 scaling factors for admittance gain x 2 parking widths x 20 repetitions) in the experiment. Each subject conducted the experiment in 10 sessions and there were 12 trials in each session (3 effective admittance gains x 2 parking widths x 2 repetitions). The trials of each session were randomized while the same order was displayed for each subject.

3.2 Human Intent Estimation via ANN

In order to estimate human intent as accelerate, decelerate, or no change using an ANN model successfully, an offline training and validation phases were both required.

1. Offline training and validation: In order to develop an ANN model, the collected data was labeled using the following procedure: the *acceleration* phase starts when the velocity is greater than the threshold (0.015 m/s) which is close to zero. The *acceleration* phase ends and the *deceleration* phase starts when the human force reaches to the maximum value. The *deceleration* phase ends when the velocity is less than the threshold value close to zero. After labeling, 75% of the data was used as for training the ANN model, and the remaining 25% was used for validation. The Keras deep learning library was used in Python for the development of ANN model. We chose Softmax to be the activation function of our output layer, as it is typical in classification problems, and LeakyReLU (Leaky Rectified Linear Unit) as the activation function of the hidden layers, as it is one of the most popular activation functions for neural networks. We selected ADAM as the optimization algorithm, validation accuracy as the main metric, and cross-entropy as the cost function to be minimized. For preventing over-fitting, L2 regularization was utilized. Weights are randomly initialized from a normal distribution. Different ANN configurations were trained, and the configuration shown on Table 3.1 was chosen eventually based on classification accuracy of the model.

Table 3.1: Hyperparameters chosen for the ANN model

Hyperparameter	Value
Number of hidden layers	5
Number of units on every hidden layers	50
Sequence length (number of time steps in input)	40
Leaky ReLU coefficient	0.01
L_2 regularization parameter	0.001
Batch Size	1000
Epochs	1000

3.3 Performance Optimization via DRL

1. Offline training and validation:

   As proposed in the model, offline training for DRL was performed by simulations. In these simulations, the admittance gain estimated by the ANN model was utilized as a precursor for the DRL model. Furthermore, human force response was generated by the C-VAE model. The simulations were performed for the index of difficulties of ID = 4 and ID = 6 and the movement distance of D = 40 cm with the configuration shown on Table 3.2. The total number of simulation trials (episodes) was 21600 (2 parking widths x 10800 repetitions).

Table 3.2: Hyperparameters chosen for the DDQN Agent

Hyperparameter	Value
Number of episodes	21000
Batch size	512
Memory size	50000
Learning rate	0.01
Discount	0.9
Target network synchronization	every 30 episodes
Training policy	ϵ-*greedy* ($\epsilon_{min} = 0.1$)
Epsilon (ϵ) Decay	0.92
Type of neural network	Multi Layer Perceptron
Number of hidden layers	4
Number of units on hidden layers	[200, 500, 200, 100]

Chapter 4

RESULTS

We compare the performance of the adaptive admittance controllers utilizing the ANN and DRL models with the admittance controller utilizing a fixed gain of nominal value (G_H^{nom}). Hence, there are 3 controller cases, which are compared for the index of difficulties of ID = 4 and ID = 6 and the movement distance of 40 cm. We present the results of our simulations and the online testing experiments. For each, we report a) normalized admittance gain profiles b) parameters making up the cost function (jerk in movement, human effort) and c) the metrics measuring task performance under the 5 different force scalings. For the performance metrics, we used the average force applied by human subject ($F_h^{ave} = 1/(t_f - t_i) \int_{t_i}^{t_f} |Fh(t)|dt$), average velocity of the cart attached to the robot ($\upsilon^{ave} = 1/(t_f - t_i) \int_{t_i}^{t_f} |\upsilon(t)|dt$, effort made (power consumed) by human subject ($\mathrm{P} = \int_0^{t_e} |F(t) \cdot \upsilon(t)|dt$), jerk of the motion ($\mathrm{J} = \int_0^{t_e} |\ddot{\upsilon}^2(t)|dt$), movement time, where t_i and t_f were the time when the velocity of the cart just exceeded 2% of the maximum velocity of the trial and the time at the end of the driving of trial (until the parking zone), t_e was the time at the end of the trial. In addition to the aforementioned methods, we assessed the parking performance using the average parking overshoot metric. This metric is calculated by determining the number of times the cart reenters the parking zone and dividing it by the total number of trials. This evaluation provides a quantitative measure of the parking accuracy.

4.1 Initial Analysis of Collected Data

The change in human force and object velocity as functions of normalized time are given in Fig. 4.1 under three different fixed scaling factors for the admittance gain (G_H^{min},G_H^{nom},G_H^{max}). The first row in the figure shows the data for ID = 4, while

the second one is for ID = 6. This graph clearly shows that the force and velocity profiles are different for different admittance gains.

In general, the subjects exert a larger force at the beginning of the co-manipulation task and gradually decrease it as they approach the parking buffer. Moreover, in some trials, the subjects even apply force in the direction opposite to the movement in order to overcome the inertia of the object. As the scaling factor increases, the magnitude of the force applied by the subjects decreases and the robot contributes to the task more. Moreover, subjects apply force for a longer duration when the scaling factor is reduced, which results in a delay in reducing the object's velocity. In addition, when the parking buffer is narrow (ID = 6), the subjects reduce their force earlier to lower the object's velocity more quickly, compared to the case of large buffer (ID = 4).

Fig. 4.2 depicts the analysis of the collected data in terms of the performance metrics defined in Section 3.1. Specifically, we report the distribution of average force applied by the subjects, F_h^{ave} (a,e), average velocity of the object, v^{ave} (b,f), parking overshoot (c,g), and the movement time (d,h) in the form of box plots for ID = 4 (first row) and ID = 6 (second row) under the controllers utilizing fixed admittance gains (G_H^{min},G_H^{nom},G_H^{max}). The average force by the subjects (object's velocity) increases (decreases) as the admittance gain increases. Both (force and velocity) are slightly higher for ID = 4 than those for ID = 6. Also, when the admittance gain increases, the task completion time (i.e. movement time) decreases. However, the trials for the narrow buffer (ID = 6) takes longer to complete than those for the large buffer (ID = 4). Also, parking overshoots occur more frequently when the parking buffer is narrow (ID = 6).

4.2 Offline Training and Validation

4.2.1 Human Intent Estimation using ANN

The ANN model successfully classified the subjects' intention as to accelerate, to decelerate, or no change with an accuracy of 95.4% and F1 Score of 0.95. Fig. 4.3 displays the confusion matrix for the trained ANN model. The values in parenthesis

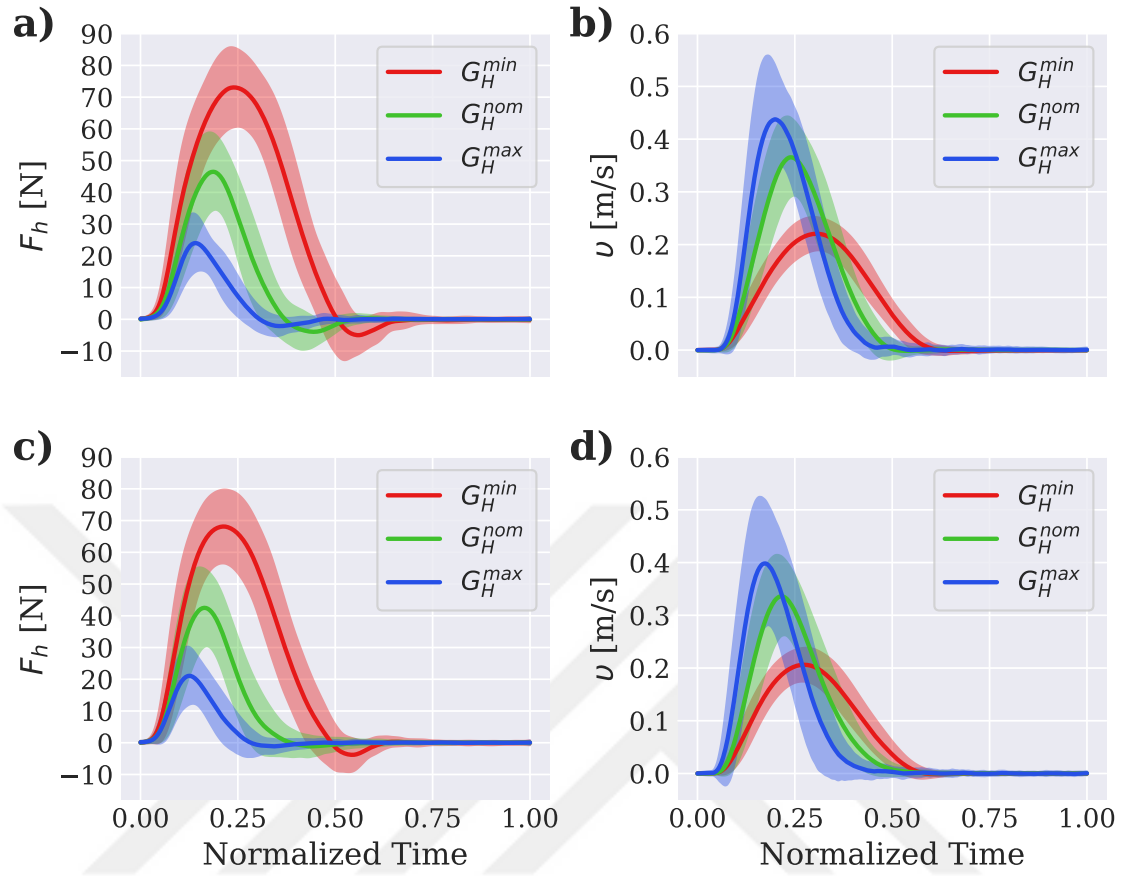

Figure 4.1: Change in human force and object velocity as a function of normalized time for ID = 4 (first row) and ID = 6 (second row) under the controllers utilizing fixed admittance gains. The solid lines in the plots are the mean values of all subjects while the shaded regions around them represent the standard deviations.

on the y-axis reports the percentage of the data in that particular class (i.e. intent). The largest confusion is between the deceleration intent and the no change during the parking phase since the force and velocity are very low and slightly fluctuates at this stage.

Fig. 4.4a depicts the force and velocity profiles of a representative subject for one trial under the controller utilizing nominal admittance gain. Fig. 4.4b shows the corresponding probability values estimated by the ANN model for each class. It can be observed from this figure that the ANN model successfully detects no change in intention initially, but as the force applied by the subject increases, the probability of the acceleration intent also increases until the force profile reaches its peak value.

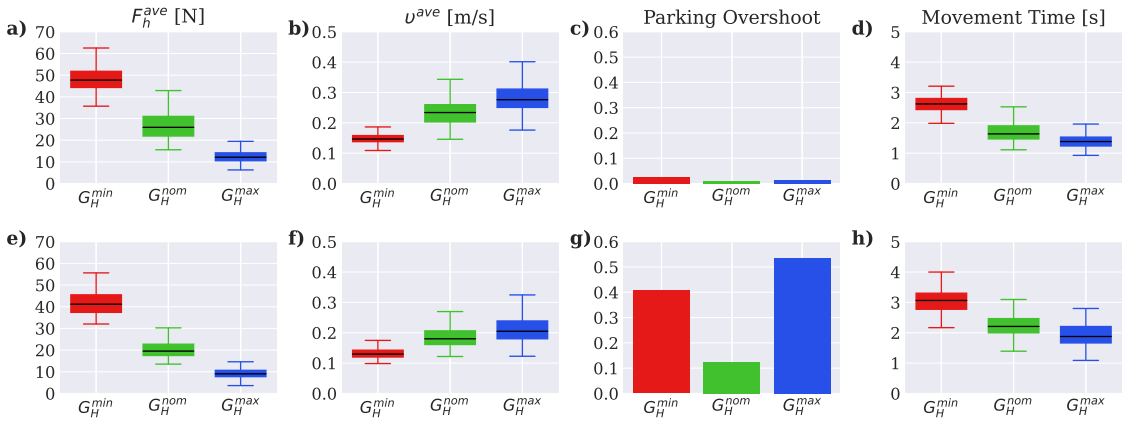

Figure 4.2: Average human force, average object velocity, parking overshoot, and movement time for ID = 4 (first row) and ID = 6 (second row) under the controllers utilizing fixed scaling factors for admittance gain.

Subsequently, the ANN model detects deceleration intent until the velocity of object is close zero and the parking phase starts. The gradual changes in the probability values support our claim that an adaptive admittance gain profile can be derived based on the classification of human motion intent.

4.2.2 Performance Optimization using DRL

Fig. 4.5 shows the bar plots for the effort made by subjects and the jerk in their movement under 3 different (fixed) scaling factors for the admittance gain. These plots demonstrate a clear trade-off between minimizing jerk and reducing human effort in co-manipulation tasks, as suggested in Section 2.3. As the gain increases, the human effort decreases (since the contribution of robot to the task increases) but the jerk in movement increases. Specifically, we found that when the force applied by the robot is amplified (i.e. maximum admittance gain), the operator spends less effort to perform the co-manipulation task, but the movement becomes jerkier. Conversely, when the force is attenuated (i.e. minimum admittance gain), the movement becomes smoother but the operator must put more effort to push the object.

In order to calculate the weighting factor (w) that will simultaneously minimize the jerk and the human effort (i.e. maximize the reward function given in Eq. 2.6),

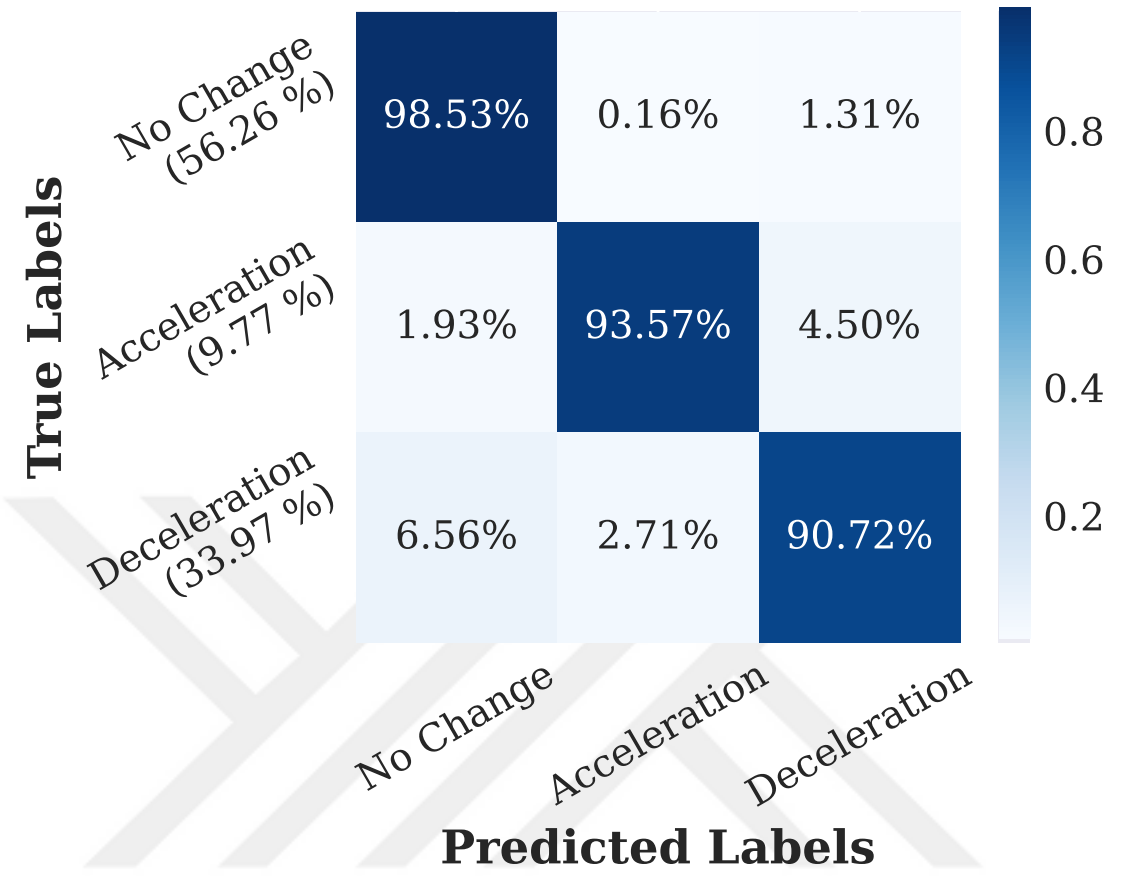

Figure 4.3: Confusion Matrix for the ANN model

we performed a Pareto analysis as implemented in [Aydin et al., 2020] for the design of an interaction controller to be utilized for pHRI tasks. Fig. 4.6 presents the Pareto front curve obtained from the scaled experimental data of jerk and human effort under 3 different (fixed) scaling factors for the admittance gain. Jerk exhibits high volatility and a wide range of values, which necessitates a nonlinear scaling approach using a quantile transform. On the other hand, a simple min-max scaling was applied to the data of human effort.

The Pareto front is the set of solutions where no objective can be improved without worsening at least one of the other objectives. In other words, the solutions on the Pareto front represent the best possible trade-offs between the conflicting objectives. By analyzing the Pareto front, we identify a set of non-dominated solutions that represent optimal trade-offs between our objectives of minimizing jerk and

human effort. This resulted in a total of 11 solutions. We employ multi-objective optimization to determine the optimal weighting factor (w). For this purpose, we perform an iterative searching process that seeks to minimize the total cost (i.e. maximize the reward function) rather than focusing solely on single objective.

We have chosen the weighting factor as w = 0.65, which results in the largest number of non-dominated solutions (i.e. 8 black dots lying on the Pareto front in Fig. 4.6).

In order to train the DRL model using the reward function given in Eq. 2.6, we developed a data-driven model for human force response as detailed in Section 2.4. We used a C-VAE model, trained by the experimental data collected for 3 fixed scaling gains (G_H^{min}, G_H^{nom},G_H^{max}) to generate synthetic data for time series of human force. Fig. 4.7 shows the force profiles as a function normalized time (a, c) generated by our model and their distribution in the form of box plot (b, d) for ID = 4 (first row) and ID = 6 (second row) under the controllers utilizing 3 different fixed scaling factors. These simulated results show similar patterns to those reported earlier in Fig. 4.1 for the actual experimental data, except the simulations performed for the maximum admittance gain (G_H^{max}) when ID = 6. The force response appears to oscillate more in the simulations during the parking phase of the task.

Once the human force was generated satisfactorily by the proposed C-VAE model, data-driven co-manipulation simulations were performed using the control architecture given in Fig. 2.1 to investigate the overall task performance under the controller utilizing the proposed two-layer ML approach (ANN+DRL) in comparison to those of the standalone ANN model and the fixed admittance gains. Fig. 4.8 reports the simulated performance of all 5 controllers for the jerk in movement, human effort, and the total cost. These simulations show that the controllers utilizing the ANN+DRL and ANN alone result in low human effort (a, d) while the jerk is low under the controllers utilizing the nominal (G_H^{nom}) and minimum (G_H^{min}) gains. In terms of the total cost, ANN+DRL outperforms all the others.

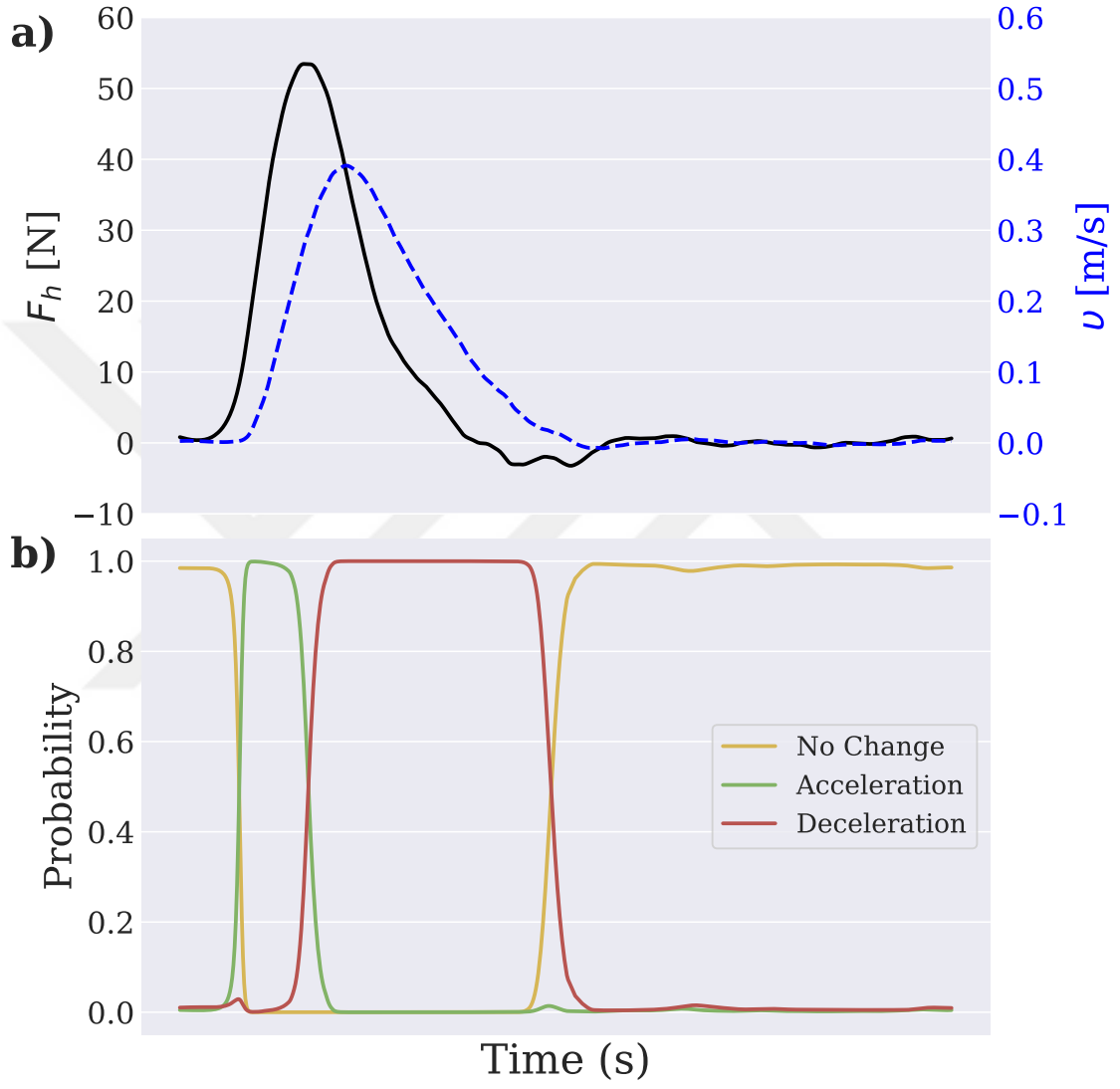

Figure 4.4: The force (solid black) and velocity (dashed blue) profiles of a subject for one trial under the controller utilizing nominal admittance gain (a) and the corresponding probability values of human motion intention classified by the ANN model (b). The ANN model classifies the human intention in real-time as 1) to accelerate the object (green), 2) to decelerate the object (red), or 3) no change (yellow).

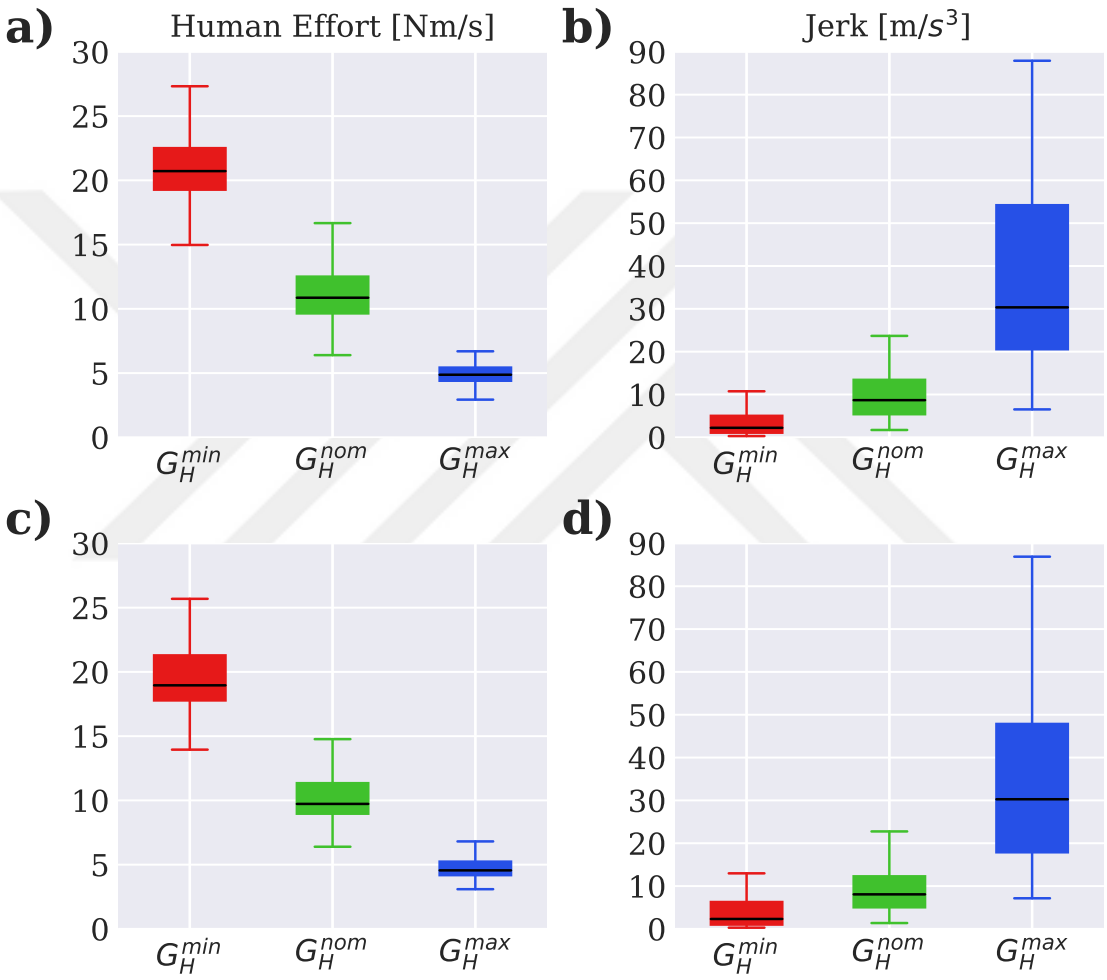

Figure 4.5: Average human effort and jerk in movement for ID = 4 (first row) and ID = 6 (second row) under the controllers utilizing fixed admittance gains.

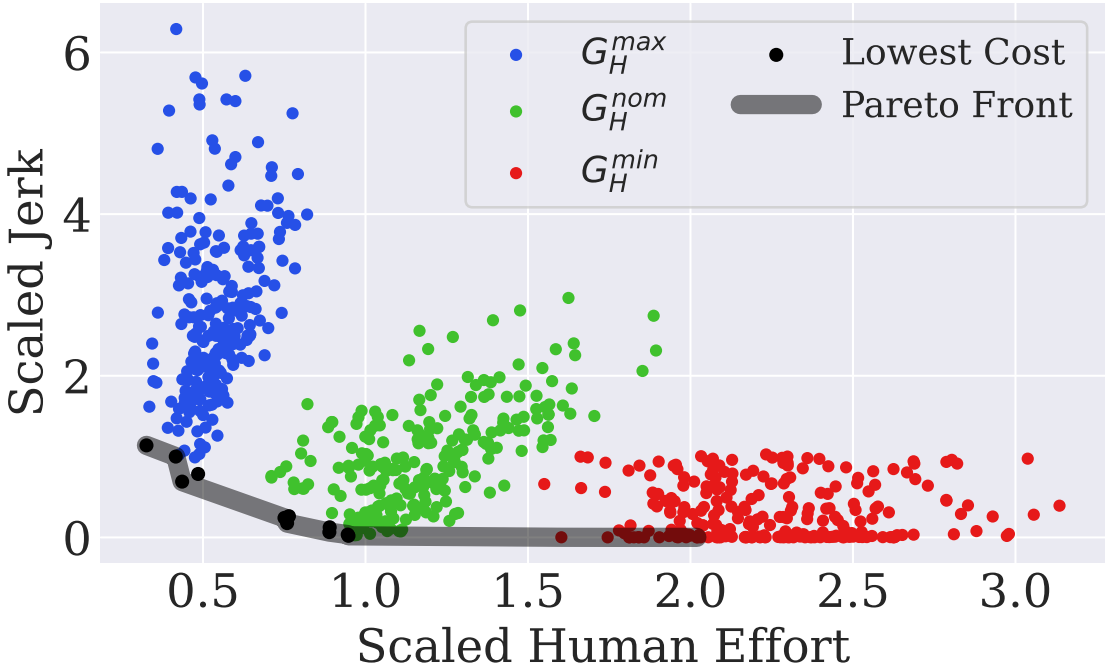

Figure 4.6: Pareto front curve (thick gray-colored) was constructed by using the experimental data of scaled human effort and scaled jerk. A linear min-max scaling was applied to the human effort while a nonlinear scaling was utilized for the jerk due to the large variations in the data. The black colored dots correspond to the experimental trials which has the lowest total cost when the weighting factor is taken as w = 0.65.

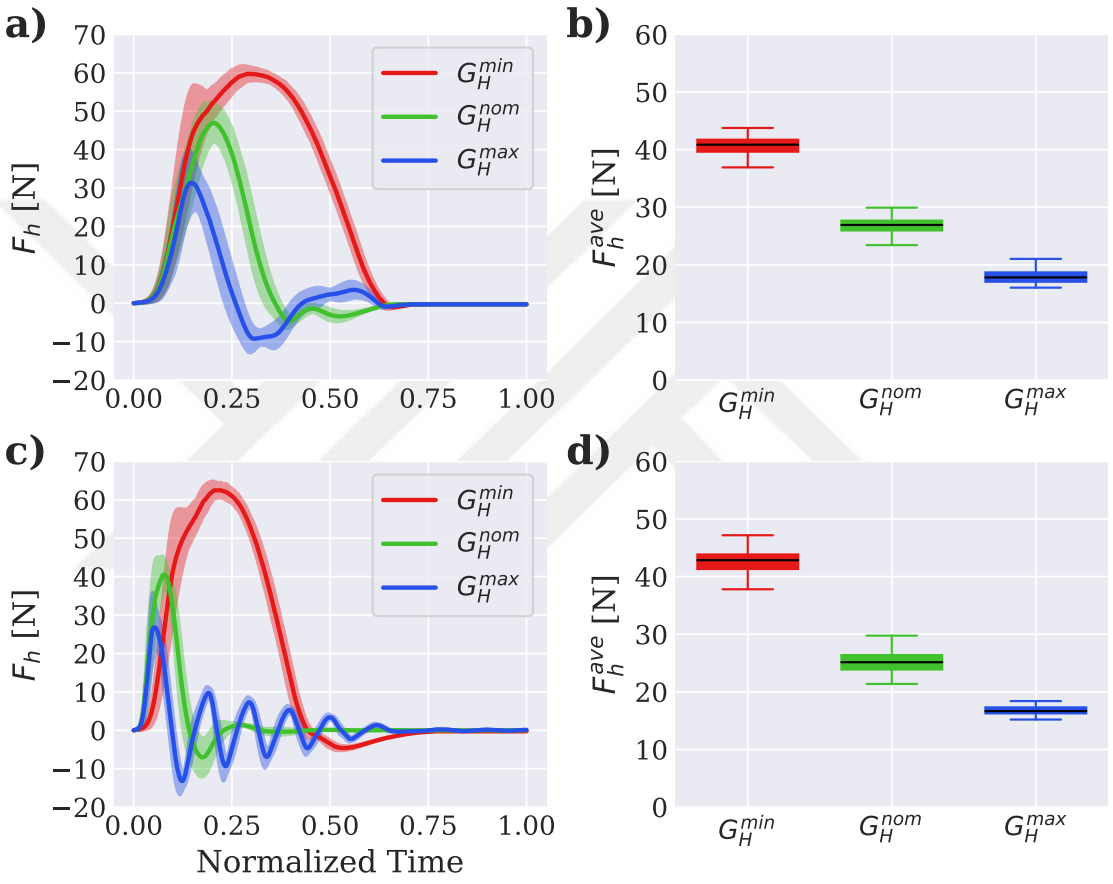

Figure 4.7: Results of the simulations: change in human force and object velocity as a function of normalized time (a). Average human force and object velocity (b) and the admittance gain profiles (c) under the controller utilizing adaptive gain profiles adjusted by the proposed two-layer ML approach (ANN+DRL) and the ANN model alone and the fixed admittance gains for ID = 4 (first row) and ID = 6 (second row).

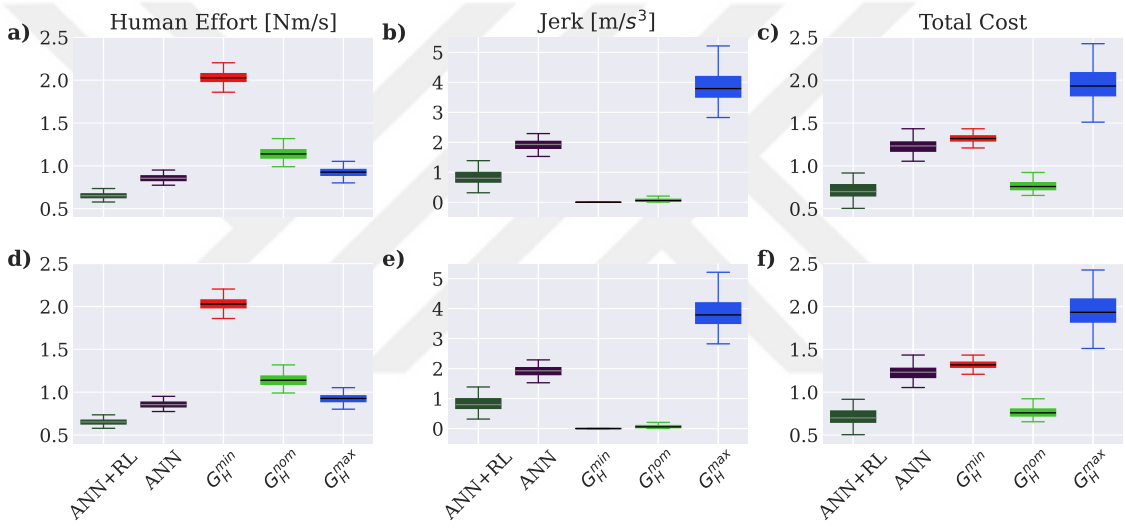

Figure 4.8: Results of the simulations: average scaled human effort, scaled jerk, and the total cost under the controllers utilizing the adaptive gain profiles adjusted by the proposed two-layer ML approach (ANN+DRL) and the ANN model alone, and the fixed admittance gains for ID = 4 (first row) and ID = 6 (second row).

Chapter 5

DISCUSSION

We present a two-stage ML approach for co-manipulation of heavy objects performed by a human and a robot. In our approach, the effective admittance gain of the interaction controller is initially adjusted by an ANN algorithm adaptively based on human intention and then fine-tuned by a DRL algorithm to optimize task performance. In fact, the results showed that the proposed ANN model could be utilized as a standalone approach to change the admittance gain adaptively based on the estimation of human intent to accelerate or decelerate the manipulated heavy object. The main benefit of DRL in our study was to optimize the task performance by maximizing a reward function. Since an adaptive admittance gain is already estimated by the ANN model for each control loop at 125 Hz, the DRL model is called less frequently. This two-layer approach is not only computationally more efficient than a standalone DRL approach but also enables us to utilize DRL algorithms that work with discrete action space, such as DDQN in our implementation.

In our study, the objective functions for optimization were to minimize the jerk in the movement and the human effort. We showed that there is a clear trade-off between them as shown in Fig. 4.5. Some of the earlier studies have focused on a single objective only such as the minimization of jerk in the movement [Dimeas and Aspragathos, 2015]. However, this approach does not take into account the other possible objectives that might be critical in pHRI tasks such as the minimization of human effort. We argue that there should be multiple objectives with trade-offs between them to optimize task performance in pHRI as proposed in [Aydin et al., 2020]. However, since the magnitude and range of the variables used in the cost function (or the reward function) could be different from each other (e.g. jerk and human effort as in our case), the weights applied to them must be determined by multi-objective optimization techniques. In our study, we opted to utilize Pareto

analysis for this purpose, which resulted in a weighting factor of w = 0.65 (Fig. 4.6).

Once the reward function was defined, we performed co-manipulation simulations to train and validate our DRL model. While RL has shown great promise in many domains, it can be challenging to apply to problems that involve pHRI. In these scenarios, the safety of the human operator is of paramount importance, and the consequences of abrupt changes in control parameters can jeopardize the safety of the human operator. Researchers have typically relied on simulations to train their RL models before deploying them to real-life scenarios, which is also the approach followed in our study. In order to perform pHRI simulations based on the control architecture shown in Fig. 2.1, a model for the force response of human was developed. Developing a dynamical model of human arm to estimate the force response for during a co-manipulation task is highly challenging due to the nonlinear and stochastic nature of human behavior. Hence, we opted to develop a data-driven model using a C-VAE model. To assess the performance of the human force model, force profiles obtained experimentally (Fig. 4.1 a,c) were compared to those generated by the simulations (Fig. 4.7 a, c) under the controllers utilizing fixed admittance gains. The results show that the data-driven human force model successfully simulated the human force response up to a certain degree. Although the mean force magnitudes for real and simulated data are very similar, the profiles are slightly different. For example, we observed that the simulated force response appears to oscillate more during the parking phase of the task when the parking width is narrow (Fig. 4.7c), when it is compared to that of the real experimental data (Fig. 4.1c). Despite these shortcomings, our human force model performed well in the co-manipulation simulations performed with the proposed adaptive controllers, though it was trained with the controllers utilizing fixed admittance gains and with a few hundred trials only.

The results of our co-manipulation simulations showed that the task performance obtained by the admittance controller utilizing the proposed two-layer approach (ANN+DRL) is better than those of the others (see the total cost in Fig. 4.8 c, f). In particular, the differences from the other controllers are more prominent as the task gets more difficult (ID = 6).

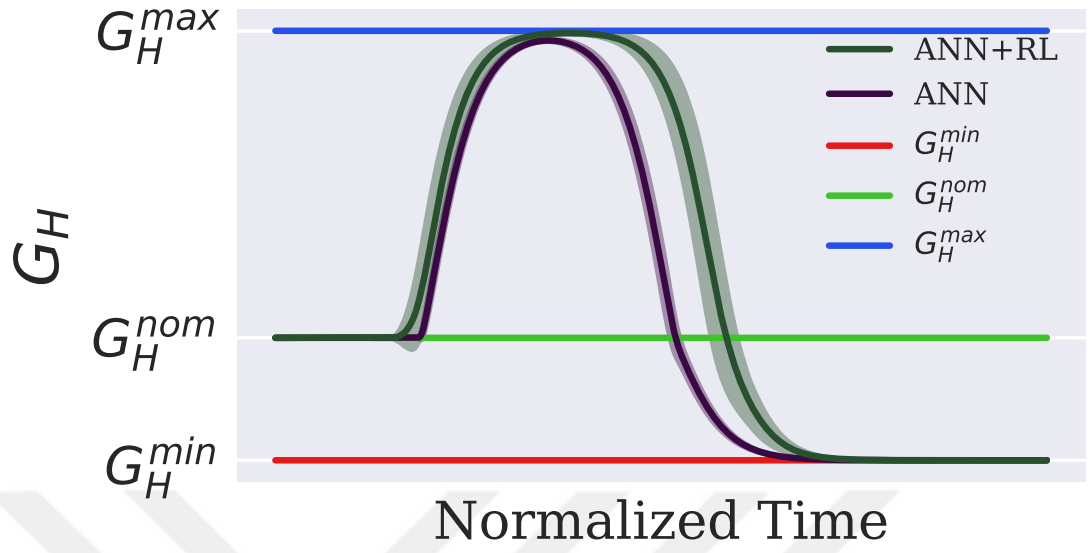

Figure 5.1: Results of the simulations: Gain profiles under 5 different controllers.

The improvement in task performance can be further justified if the simulated gain profiles shown in Fig. 5.1 are inspected. First of all, the adaptive gain profiles (ANN, ANN+DRL) obtained through the simulations match with the one suggested in Fig. 2.2. Moreover, we observe that the controller utilizing the proposed two-layer approach (ANN+DRL) exhibited a more proactive response during the whole task, amplifying the gain earlier during the acceleration phase to reduce human effort while rapidly attenuating it during the deceleration phase, mitigating the jerkiness in the movement and helping with the parking accuracy.

Chapter 6

CONCLUSION

The proposed two-layer ML approach, with the first layer (ANN model) as the precursor for the second layer (DRL model) reduces human effort in human-robot co-manipulation. Moreover, the proposed data-driven model for human force response paves the way for the offline training of RL models via simulations for pHRI tasks, reducing the potential safety risks for the human operator during online training.

In summary, the proposed approach successfully demonstrates that a) the ANN model can successfully estimate the human intent to accelerate (decelerate) a manipulated heavy object before the object actually starts to accelerate (decelerate); b) the human force modeling helps to reduce the number of episodes executed in the real world for DRL training since the training can be done in the simulation domain; c) the DRL model can return an optimal gain profile for the admittance controller once an initial profile, obtained by the ANN model, is supplied to it.

Further research could investigate the feasibility of the proposed approach in more complex manipulation tasks involving 6-dof physical interactions between human and robot. All the ML approaches utilized in this study estimate the model outputs for the current state, but predicting the future outputs could be helpful. For example, predicting the future human force response could help with early adaptation of the admittance gain for further improvement in task performance.

BIBLIOGRAPHY

[Al-Saadi et al., 2023a] Al-Saadi, Z., Hamad, Y., Aydin, Y., Kucukyilmaz, A., and Basdogan, C. (2023a). Resolving conflicts during human-robot co-manipulation.

[Al-Saadi et al., 2023b] Al-Saadi, Z., Hamad, Y. M., Aydin, Y., Kucukyilmaz, A., and Basdogan, C. (2023b). Resolving conflicts during human-robot co-manipulation.

[Al-Saadi et al., 2021] Al-Saadi, Z., Sirintuna, D., Kucukyilmaz, A., and Basdogan, C. (2021). A novel haptic feature set for the classification of interactive motor behaviors in collaborative object transfer. *IEEE Transactions on Haptics*, 14(2):384–395.

[Aydin et al., 2018] Aydin, Y., Tokatli, O., Patoglu, V., and Basdogan, C. (2018). Stable physical human-robot interaction using fractional order admittance control. *IEEE Transactions on Haptics*, 11(3):464–475.

[Aydin et al., 2020] Aydin, Y., Tokatli, O., Patoglu, V., and Basdogan, C. (2020). A computational multicriteria optimization approach to controller design for physical human-robot interaction. *IEEE Transactions on Robotics*, 36(6):1791–1804.

[Callens et al., 2020] Callens, T., van der Have, T., Rossom, S. V., De Schutter, J., and Aertbeliën, E. (2020). A framework for recognition and prediction of human motions in human-robot collaboration using probabilistic motion models. *IEEE Robotics and Automation Letters*, 5(4):5151–5158.

[Crnkovic-Friis and Crnkovic-Friis, 2016] Crnkovic-Friis, L. and Crnkovic-Friis, L. (2016). Generative choreography using deep learning. *CoRR*, abs/1605.06921.

[Dimeas and Aspragathos, 2015] Dimeas, F. and Aspragathos, N. (2015). Reinforcement learning of variable admittance control for human-robot co-manipulation.

[Ding et al., 2011] Ding, H., Reißig, G., Wijaya, K., Bortot, D., Bengler, K., and Stursberg, O. (2011). Human arm motion modeling and long-term prediction for safe and efficient human-robot-interaction. In *2011 IEEE International Conference on Robotics and Automation*, pages 5875–5880.

[Duchaine and Gosselin, 2007] Duchaine, V. and Gosselin, C. M. (2007). General model of human-robot cooperation using a novel velocity based variable impedance control. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pages 446–451.

[Fitts, 1992] Fitts, P. M. (1992). The information capacity of the human motor system in controlling the amplitude of movement. 1954. *Journal of experimental psychology. General*, 121 3:262–9.

[Formica et al., 2021] Formica, F., Vaghi, S., Lucci, N., and Zanchettin, A. M. (2021). Neural networks based human intent prediction for collaborative robotics applications. In *2021 20th International Conference on Advanced Robotics (ICAR)*, pages 1018–1023.

[Ghadirzadeh et al., 2016] Ghadirzadeh, A., Bütepage, J., Maki, A., Kragic, D., and Björkman, M. (2016). A sensorimotor reinforcement learning framework for physical human-robot interaction. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2682–2688.

[Gómez-González et al., 2018] Gómez-González, S., Neumann, G., Schölkopf, B., and Peters, J. (2018). Adaptation and robust learning of probabilistic movement primitives. *CoRR*, abs/1808.10648.

[Guler et al., 2022] Guler, B., Niaz, P. P., Madani, A., Aydin, Y., and Basdogan, C.

(2022). An adaptive admittance controller for collaborative drilling with a robot based on subtask classification via deep learning. *Mechatronics*, 86:102851.

[Hamad et al., 2021] Hamad, Y. M., Aydin, Y., and Basdogan, C. (2021). Adaptive human force scaling via admittance control for physical human-robot interaction. *IEEE Transactions on Haptics*, 14(4):750–761.

[Hasselt et al., 2016] Hasselt, H. V., Guez, A., and Silver, D. (2016). Deep reinforcement learning with double q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).

[Hiatt et al., 2017] Hiatt, L. M., Narber, C., Bekele, E., Khemlani, S. S., and Trafton, J. G. (2017). Human modeling for human–robot collaboration. *The International Journal of Robotics Research*, 36(5-7):580–596.

[Holzbaur et al., 2005] Holzbaur, K. R., Murray, W. M., and Delp, S. L. (2005). A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control. *Annals of biomedical engineering*, 33:829–840.

[Kucukyilmaz et al., 2012] Kucukyilmaz, A., Oguz, S. O., Sezgin, T. M., and Basdogan, C. (2012). *Improving Human-Computer Cooperation Through Haptic Role Exchange and Negotiation*, pages 229–254. Springer London, London.

[Landi et al., 2019] Landi, C. T., Cheng, Y., Ferraguti, F., Bonfè, M., Secchi, C., and Tomizuka, M. (2019). Prediction of human arm target for robot reaching movements. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5950–5957.

[Li et al., 2020] Li, Q., Zhang, Z., You, Y., Mu, Y., and Feng, C. (2020). Data driven models for human motion prediction in human-robot collaboration. *IEEE Access*, PP:1–1.

[Li and Ge, 2014] Li, Y. and Ge, S. S. (2014). Human–robot collaboration based

on motion intention estimation. *IEEE/ASME Transactions on Mechatronics*, 19(3):1007–1014.

[Li et al., 2017] Li, Z., Liu, J., Huang, Z., Peng, Y., Pu, H., and Ding, L. (2017). Adaptive impedance control of human–robot cooperation using reinforcement learning. *IEEE Transactions on Industrial Electronics*, 64(10):8013–8022.

[Losey et al., 2018] Losey, D. P., McDonald, C. G., Battaglia, E., and O'Malley, M. K. (2018). A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. *Applied Mechanics Reviews*, 70(1).

[Lu et al., 2020] Lu, W., Hu, Z., and Pan, J. (2020). Human-robot collaboration using variable admittance control and human intention prediction. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pages 1116–1121.

[Madan et al., 2015] Madan, C. E., Kucukyilmaz, A., Sezgin, T. M., and Basdogan, C. (2015). Recognition of haptic interaction patterns in dyadic joint object manipulation. *IEEE Transactions on Haptics*, 8(1):54–66.

[Mainprice and Berenson, 2013] Mainprice, J. and Berenson, D. (2013). Human-robot collaborative manipulation planning using early prediction of human motion.

[Mainprice et al., 2015] Mainprice, J., Hayne, R., and Berenson, D. (2015). Predicting human reaching motion in collaborative tasks using inverse optimal control and iterative re-planning. volume 2015.

[Modares et al., 2016] Modares, H., Ranatunga, I., Lewis, F. L., and Popa, D. O. (2016). Optimized assistive human–robot interaction using reinforcement learning. *IEEE Transactions on Cybernetics*, 46(3):655–667.

[Mörtl et al., 2012] Mörtl, A., Lawitzky, M., Kucukyilmaz, A., Sezgin, M., Basdogan, C., and Hirche, S. (2012). The role of roles: Physical cooperation between humans and robots. *The International Journal of Robotics Research*, 31(13):1656–1674.

[Oguz et al., 2012] Oguz, S. O., Kucukyilmaz, A., Metin Sezgin, T., and Basdogan, C. (2012). Supporting negotiation behavior with haptics-enabled human-computer interfaces. *IEEE Transactions on Haptics*, 5(3):274–284.

[Oguz et al., 2010] Oguz, S. O., Kucukyilmaz, A., Sezgin, T. M., and Basdogan, C. (2010). Haptic negotiation and role exchange for collaboration in virtual environments. In *2010 IEEE Haptics Symposium*, pages 371–378.

[Rahman et al., 2002] Rahman, M. M., Ikeura, R., and Mizutani, K. (2002). Investigation of the impedance characteristic of human arm for development of robots to cooperate with humans. *JSME International Journal Series C*, 45(2):510–518.

[Ravichandar and Dani, 2017] Ravichandar, H. C. and Dani, A. P. (2017). Human intention inference using expectation-maximization algorithm with online model learning. *IEEE Transactions on Automation Science and Engineering*, 14(2):855–868.

[Roveda et al., 2020] Roveda, L., Maskani, J., Franceschi, P., Abdi, A., Braghin, F., Tosatti, L. M., and Pedrocchi, N. (2020). Model-based reinforcement learning variable impedance control for human-robot collaboration. *Journal of Intelligent & Robotic Systems*, 100(2):417–433.

[Sharkawy et al., 2020] Sharkawy, A.-N., Koustoumpardis, P. N., and Aspragathos, N. (2020). A neural network-based approach for variable admittance control in human–robot cooperation: online adjustment of the virtual inertia. *Intelligent Service Robotics*, 13(4):495–519.

[Sohn et al., 2015] Sohn, K., Lee, H., and Yan, X. (2015). Learning structured output representation using deep conditional generative models. In Cortes, C.,

Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

[Soukoreff and MacKenzie, 2004] Soukoreff, R. W. and MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *Int. J. Hum. Comput. Stud.*, 61:751–789.

[Townsend et al., 2017] Townsend, E. C., Mielke, E. A., Wingate, D., and Killpack, M. D. (2017). Estimating human intent for physical human-robot comanipulation. *CoRR*, abs/1705.10851.

[Whitsell and Artemiadis, 2017] Whitsell, B. and Artemiadis, P. (2017). Physical human–robot interaction (phri) in 6 dof with asymmetric cooperation. *IEEE Access*, 5:10834–10845.

[Wu et al., 2019] Wu, M., He, Y., and Liu, S. (2019). Shared impedance control based on reinforcement learning in a human-robot collaboration task. In *Advances in Service and Industrial Robotics*, pages 95–103. Springer International Publishing.