2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Enhanced Human-Robot Collaboration with Intent Prediction using Deep Inverse Reinforcement Learning

Mukund Mitra[1], Gyanig Kumar[1], Partha Pratim Chakrabarti [2], Pradipta Biswas[1]

Abstract— **In shared autonomy, human-robot handover for object delivery is crucial. Accurate robot predictions of human hand motion and intentions enhance collaboration efficiency. However, low prediction accuracy increases mental and physical demands on the user. In this work, we propose a system for predicting hand motion and intended target during human-robot handover using Inverse Reinforcement Learning (IRL). A set of feature functions were designed to explicitly capture users' preferences during the task. The proposed approach was experimentally validated through user studies. Results indicate that the proposed method outperformed other state-of-the-art methods (PI-IRL, BP-HMT, RNNIK-MKF and $\mathrm{CM_{k=5}}$) with users feeling comfortable reaching upto 60% of the total distance to the target for handover with 90% target prediction accuracy. The target prediction accuracy reaches 99.9% when less than 20% of the task remains.**

I. INTRODUCTION

A collaborative robot or *cobot* operates in direct cooperation with humans within a collaborative workspace [1], [2]. Cobots find intensive applications in domestic settings [3], clinical environments [4] and product assembly in manufacturing [5]. In such environments, it is required for humans to reach various targets for transferring products. However, reaching distant targets can lead to ergonomic fatigue for the user. Consequently, handing over products to robot for correct completion of subsequent actions can improve efficiency of the overall process. Human-robot handovers refers to actions initiated either by the human or their robot counterpart, with the objective of delivering products to each other [6]. These handovers require the robot to anticipate human hand movements and intentions to effectively plan its path, take control of the object from human, and deliver it to the intended destination. Trajectory prediction algorithms are frequently employed to forecast hand movements and intended targets [7]–[9] during human-robot interaction (HRI). Many existing prediction algorithms are constrained by a limited ergonomic collaborative workspace (ECW) [10] with high ergonomic stress on the user. This limitation arises from suboptimal intent or target prediction, leading to increased effort, temporal demands, duration, and intensity of exertion. Instead of improving human-robot interaction, wrong prediction of operators' intention can place higher mental and physical workload on the user [11]. To enhance both the comfort and productivity of human workers during handover, it is important to expand the ECW. However, limited research is reported on ergonomic aspect of human-robot collaborative (HRC) systems [10], [12].

In this work, an Inverse Reinforcement Learning (IRL) based system is proposed for intent prediction during handovers. Results demonstrate that this approach enhances prediction accuracy resulting in improved collaboration and reduced ergonomic stress. In summary, the key contributions of the work are:

1) An IRL-based system with re-tracking is proposed to predict hand motion during human-robot handovers.
2) A comprehensive set of features was introduced to explicitly capture hand motion during the task. A target prediction algorithm was proposed for accurate anticipation of the intended target based on partial hand motion.
3) User studies were performed to compare different prediction algorithms in terms of target and trajectory prediction accuracy, physical workload, and mental workload on the user.

II. LITERATURE SURVEY

In the context of human to robot handover, predicting human intentions is vital. Numerous studies have explored the prediction of human intent based on trajectory data. These algorithms for human trajectory prediction can be categorized into four groups: (1) Physics-based, (2) Probabilistic graphical models, (3) Recurrent neural network (RNN) models, and (4) Inverse reinforcement learning (IRL) or Inverse optimal control (IOC). **Physics-based models** are based on physical modelling of human motion and parameters can be estimated using system constraints. Examples include dynamic prediction using a constant acceleration model with Kalman filter [13], minimum jerk model [14], [15]. While these models can be generalized, they lack confidence information. **Probabilistic models** like Gaussian mixture models (GMMs) and Hidden Markov Models (HMMs) are commonly used for intent prediction. GMM with Gaussian mixture regression [7] was used to estimate occupancy areas using hand motion prediction. To predict hand motion, an online unsupervised learning algorithm with expectation maximization (EM) [16] was proposed to learn the parameters of GMM. A variant of left-to-right HMM [17] was used to model human transition dynamics for trajectory prediction. *Huang and Mun* [18] modeled intention as a Markov process and extended intention tracking to a multi-layer intention hierarchy. Other methods include Conditional Random Fields (CRFs) [19], [20], Support Vector Machines (SVMs) [21], Partially Observable Markov Decision Process (POMDPs) [22]. These models encode task-motion temporal

[1]I3D Lab, Robert Bosch Center for Cyber-Physical System, Indian Institute of Science, Bangalore, India
[2]Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, India

relationships but lack constraints governing the HRC task [23]. With the advancement of neural networks, **RNN models** are increasingly being used to predict human motion. An encoder-recurrent-decoder structure using LSTM as recurrent layer [24] was used to forecast human pose. Other works include [25]–[27]. While effective in their specific prediction contexts, these models often demand substantial data and exhibit limited generalization. **Inverse reinforcement learning (IRL) or Inverse optimal control (IOC)** models humans as agents optimizing a reward function during task demonstrations. The learned reward is used to predict future motions. *Manyar et al.* [28] employed Maximum Margin IRL to model human subtask sequences and transfer them to a robot for desired outcomes. *Ziebert* [29] proposed Maximum Entropy IRL (MEIRL) which maximizes the entropy of demonstrations with feature expectation matching constraint. *Berret et al.* [30] introduced an IOC method that analyzes multiple costs concurrently to predict human hand motion. To predict reaching motion during collaborative assembly, [23] used learned reward for iterative replanning of trajectory. While these methods demand fewer demonstrations and accommodate noisy data, estimating the state-partition function remains challenging [31]. To address this, sampling-based IRL algorithms were introduced which effectively generate sample trajectories to estimate the state partition function. Examples include template-based RRT [32], sampling-based MEIRL (SMEIRL) [33] with discrete elastic band [34] and importance sampling [31]. However, generating random samples to estimate the partition function can be inefficient. Path Integral-IRL (PI-IRL) [35] outperforms importance sampling by sampling in the vicinity of demonstrations. However, such formulations assume linear reward-feature dependency, which may not capture complex non-linear rewards during human-robot collaboration (HRC). Maximum Entropy Deep Inverse Reinforcement Learning (MEDIRL) [36], [37] leverages the representational capacity of neural networks to capture complex reward distribution. **Re-tracking:** Accurate prediction becomes challenging with tracking errors. *Rui* introduced a *re-tracking* method [38] that iteratively smooths tracked motion with prediction to reduce tracking errors. Related work in *tracking-by-prediction* framework includes [39].

However, the discussed prediction models primarily address robot path planning, control, and safety, with limited emphasis on user ergonomic comfort. **Ergonomics in HRI:** *Figueredo et al.* proposed a comfortability index that integrates ergonomic and muscular factors with joint constraints [40]. While it offers real-time filtering for specific task conditions, it emphasizes only physical parameters for comfortability. *Lagomarsino et al.* [11] introduced an online vision-based cognitive load assessment framework, evaluating mental workload during HRI through attention distribution and upper body kinematics monitoring. Recently, *Liu et al.* [41] proposed an optimized assignment technique using tolerance index (TI) derived from physical workload (PWL) and mental workload (MWL) for diverse HRC tasks in O^*Net database [42]. While PWL and MWL were originally computed from Strain Index (SI) [43] and NASA-TLX [44] respectively, the O^*Net database lacked SI and TLX parameters. Instead, the authors employed similar metrics, leading to mixed results for TI values.

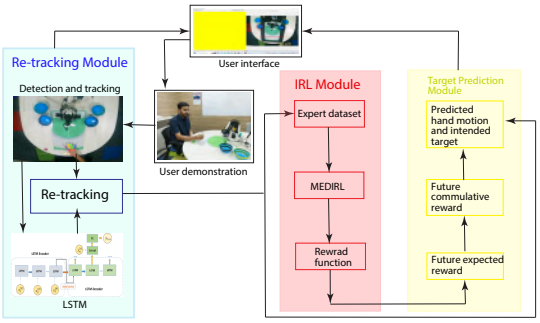


Fig. 1: Framework of the proposed approach using Re-tracking, MEDIRL, and target prediction module along with the user interface

In this work considering the capacity of MEDIRL to model complex reward functions during HRC and re-tracking by prediction paradigm to minimize tracking error, we present an approach to predict human hand motion and intended target during handovers. We designed a set of feature functions to capture hand motion during the task and analyzed a target prediction metric to accurately predict the intended target from partial hand motion. A hand motion dataset was created for reaching task and user studies were conducted to validate the performance of the proposed system. Results show that the proposed approach reports higher prediction accuracy with reduced ergonomic stress.

III. Methodology

The proposed approach comprises three modules: (1) Re-tracking, (2) IRL, and (3) Target prediction, illustrated in Fig. 1. Tracked hand motion was processed by the re-tracking module, generating re-tracked trajectories that served as an expert dataset for learning the reward distribution via MEDIRL. The acquired reward distribution was utilized by the target prediction module to forecast hand motion and the intended target. Users receive real-time visual feedback on hand tracking, robot position, and predicted targets through the user interface.

A. Maximum Entropy Deep Inverse Reinforcement Learning

Human hand motion was modeled as an agent following a Markov Decision Process (MDP). An MDP is defined as $\{S, A, T, \gamma, r\}$ consisting of of states $s \epsilon S$, actions $a \epsilon A$, probability of transition T, discount factor γ, and reward function $r : S \rightarrow \mathbb{R}$. Let $\mathfrak{D} = \{\tau_i\}_{i=1}^{M}$ be the expert dataset consisting of M hand motion trajectories τ given by $\tau = [s_1, a_1, s_2, a_2, ..., a_{N-1} s_N]$. IRL aims to learn a reward function r, under which the expert demonstration is optimal. MEDIRL [36] approximates the reward function using a deep neural network parameterized by ω, i.e., $r_\omega(s) = g(f(s), \omega)$ where $f(s)$ are the features of state s and ω are

TABLE I: Velocity Curve Fit

	Average R^2	Average error (cm/s)
Linear	0.80	10.33
Quadratic	**0.97**	**7.21**
Cubic	0.79	11.45

the weights of the network or reward parameters. Function $g : \mathbb{R}^N \rightarrow \mathbb{R}$ maps the feature space of dimension N to a real-valued reward r. The IRL problem was framed by maximizing the joint posterior distribution of observing the expert dataset and reward parameters under a given reward structure.

$$L(\omega) = logP(\mathfrak{D}, \omega | r) = logP(\mathfrak{D}|r) + logP(\omega) \quad (1)$$

The optimal reward parameters can be obtained by backpropagating the gradient with a regularization technique [36] as given below:

$$\frac{\partial L}{\partial \omega} = (\mu_{\mathfrak{D}} - \mathbb{E}[\mu]) \frac{\partial g(f, \omega)}{\partial \omega} \quad (2)$$

where $\mu_{\mathfrak{D}}$ is the State Visitation Frequency (SVF) from the expert data and $\mathbb{E}[\mu]$ is the expected SVF [29] given the learned reward at each iteration. Approximate Value Iteration and Policy Propagation algorithm [45] were used to estimate SVF. A feedforward neural network with 4 hidden layers with 128 neurons each and a ReLU activation function was employed for feature parameterized reward network. The state space was defined by discretizing the collaborative workspace of $60 \times 60cm$ into a 30×30 uniform grid. Variation of z-coordinate of the hand was in the range $[0, 1.8]cm$ and assumed constant to ease computation. The action space consists of five actions *(up, left, right, up-left, up-right)* considering only forward motion of hand.

B. Features

From equation (2), it may be noted that reward distribution is a function of feature space. The following features were used to model human preferences during the task:

- Distance feature (f_d): To minimize effort, humans prefer the shortest path. The distance feature captures deviation from this path:

$$f_d(s) = e^{-\left\{d(s) - d_{shortest}\left(s'\right)\right\}} \quad (3)$$

  where $d_{shortest}$ is the straight line path from start to end of the trajectory with states s'.
- Velocity feature (f_v): Given the close proximity of targets, relying solely on the distance feature is inadequate for distinguishing between targets. To address this, the velocity feature was introduced, which captures deviations from the desired hand velocity v_{des}. Desired velocity was determined based on previous work [46], [47], which reported velocity profiles concerning the distance to the end-point. Linear, quadratic, and cubic velocity profiles were fit with respect to the target's distance. Table I, indicates that quadratic velocity profile gives the lowest error and highest R^2. Therefore, the desired velocity at each state is given by:

$$v_{des}(s) = aX(s)^2 + bX(s) + c \quad (4)$$

  where $X(s)$ represents the distance of state s from the end state, and a,b,c are constants derived from quadratic polynomial fitting. Deviations from this desired velocity is the velocity feature:

$$f_v(s) = -\left(v_{des}(s) - v(s)\right)^2 \quad (5)$$

Considering hand dynamics and user comfort, acceleration $f_a(s)$ and jerk $f_j(s)$ features were introduced. These features were computed as the sum of squared acceleration and jerk at each state, respectively. Feature of a trajectory was expressed as the sum of features of each individual state of the trajectory. All the features were normalized to (0,1) to have equal contribution to the learned reward function.

C. Re-tracking Module

The re-tracking module iteratively smoothens tracked hand trajectory by integrating data from the camera and predictions as the task progresses. This module comprises three components:

- Detection and tracking: Google Mediapipe was employed to detect and track hand movements [48].
- Prediction: An LSTM recurrent neural network with linear encoder and decoders, with 3 LSTM layers of 500 units each was used for prediction. Previous work [24] used similar model for human motion prediction.
- Re-tracking: Re-tracking method proposed in [38] was adopted. The method is adaptable to various tracking and prediction pipelines. This re-tracking model also accounts for suboptimal hand motion at the task's beginning.

Hand trajectory information from Google Mediapipe was sent to the LSTM network. The trajectory predicted by the LSTM and that tracked by the camera were combined for re-tracked trajectory.

D. Target Prediction Module

MEDIRL learns the underlying reward distribution r that maximizes expert demonstrations. The reward function was made goal explicit (r_G) by assigning higher reward to goal states G. The reward distribution was obtained such that the expectation of features of the expert with respect to the probability distribution π_G over the actions (i.e., hand movements) at a given state equals that of the learner. Using the maximum entropy formulation of IRL [29], [46], π_G is given by:

$$\pi_G(a_i|s_i) \propto e^{Q_G(s_i, a_i)} \quad (6)$$

$$V_G(s_i) = \underset{a_i}{\text{softmax}} \{r_G(s_i) + Q_G(s_i, a_i)\} \quad (7)$$

$$Q_G(s_i, a_i) = \mathbb{E}_{T(s_{i+1}|s_i, a_i)}[V_G(s_{i+1})|s_i, a_i] \quad (8)$$

where V_G and Q_G are the future expected and cumulative rewards respectively. They are recursively calculated using equations (7), (8) and the MDP given in Section III-A. The probability distribution over a partial trajectory $\psi = \{s_1, ..., s_m\}$, $m < N$ is given by:

$$\begin{aligned} p(\psi|G) &= \prod_{i=1}^{m} \pi_G(a_i|s_i) \\ &= e^{\left\{\sum_{i=2}^{m} r_G(s_i)\right\} + V_G(s_m) - V_G(s_1)} \end{aligned} \quad (9)$$

The above equation assigns higher probabilities to trajectories which maximizes future expected rewards. Using Equation (9), Bayes' theorem was employed to calculate the probability of a goal given partial hand trajectory:

$$p\left(G|\psi\right) = \frac{p\left(\psi|G\right)p\left(G\right)}{\sum_{G\epsilon\mathbb{G}} p\left(\psi|G\right)p\left(G\right)} \tag{10}$$

where $\mathbb{G}$ is the set of 4 targets. Equation (10) assigns higher probabilities to targets towards which the partial trajectory approaches. The goal corresponding to maximum probability was predicted. A summary of target prediction is given in Algorithm 1. Future hand trajectory to the predicted target G_{pred} was obtained from $\pi_{G_{pred}}$.

Algorithm 1 Target Prediction

Input: Partial hand trajectory $\psi = \{s_1, ..., s_m\}$, Goal states $G\epsilon\mathbb{G}$, MDP = $\{S, A, T, \gamma, r\}$
Output: Predicted goal G_{pred}

1: **for** G in $\mathbb{G}$ **do**
2: Initialize $V_G = 0$
3: Update $V_G \leftarrow (r, S, A, T, \gamma)$, [Equation (7), (6)]
4: $p\left(\psi|G\right) = e^{\left\{\sum_{i=2}^{m} r(s_i)\right\} + V_G(s_m) - V_G(s_1)}$
5: $p\left(G|\psi\right) = \frac{p(\psi|G)p(G)}{\sum_{G\epsilon\mathbb{G}} p(\psi|G)p(G)}$
6: **end for**
7: $G_{pred} \leftarrow \arg\max_G p\left(G|\psi\right)$

IV. User Study

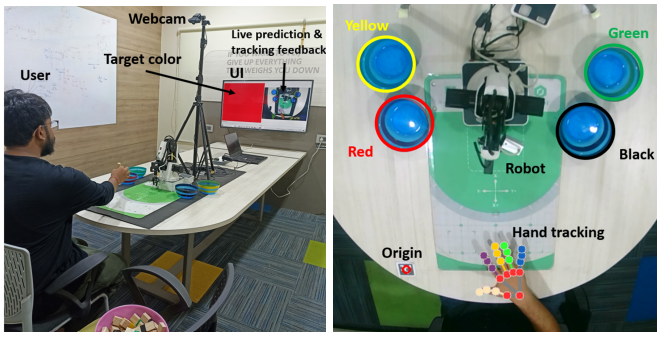
Two user studies were conducted: (1) Reaching task without robot and (2) Human-robot handover.

A. Reaching task without robot

This study investigates the physical and mental workload on users when reaching distant targets. Data collected through this experiment was used to train the MEDIRL model.

1) Participants: 10 participants (8 males and 2 females) were recruited from our university, averaging 27.9 years of age (std: 4.2). Their average arm length was 57.83 cm (std: 2.5), with 7 being right-handed and 3 left-handed individuals. None of the participants had color blindness. All participants provided necessary permissions and consent for the trials.

2) Setup: The experimental setup is depicted in Fig. 2a. Four target bowls of diameter $14cm$ were colored yellow, black, red and green. $\{yellow\ , green\}$ were placed at a distance of $115cm$ and $\{black\ , red\}$ at a distance of $95cm$ from the user, symmetrically at close proximity to each other. The user interface displays a random color from the target set, continuous hand tracking, target positions, and the robot's configurations. The target bowl size increases upon prediction, providing visual feedback. Additionally, it shows the number of iterations and the predicted target. The interface ran on a TV positioned at $3.2m$ away from the participant. A fixed-base robotic manipulator, Dobot Magician, was situated at a distance of $100cm$ from the user. The robot and targets were on a $70cm$ high table. The user sat on a chair with a height of $45.7cm$, positioned $40cm$ from the nearest edge of the robot's workspace. The camera



(a) (b)

Fig. 2: (a) Experimental setup (b) Hand tracking using Google Mediapipe

[Fig. 2b] placed on a tripod $33cm$ from the robot, tracked hand coordinates. These coordinates were transformed into Cartesian coordinates relative to the April tag located at the bottom-left corner of the workspace. Colored blocks matching the targets were kept in a bowl $50cm$ away from the user's dominant hand.

3) Design: The UI randomly presents a color when the *Next target* button was clicked. Users then collect a block of the same color from the bowl and place it in the corresponding target, completing one iteration. Each user was required to complete 20 iterations, with 5 for each target randomly assigned. Hand motion tracking was exclusively focused on forward-reaching motion.

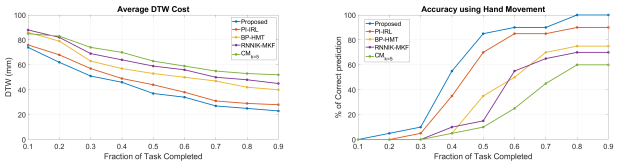
4) Procedure: Participants were briefed about the task and then conducted trials. Following each trial, the robot was set to its home position. After each participant completed their trial, subjective feedback was collected using Strain Index (SI) for estimating physical load and NASA-TLX for assessing cognitive load.

B. Human-robot handover

The experimental setup, procedure, and participants were the same as those in the *reaching task without robot*, with slight modifications in the experiment design.
The user's task involves reaching and placing a color block at the corresponding target. Prediction was made at each time instant. The robot advances along the predicted trajectory to a fixed point for takeover. However, hand tracking was limited to a set threshold distance indicated by the red horizontal line in the UI. Based on this partial motion, final intended target was predicted and the robot places the block at the predicted target. This completes one iteration of the task. Each user must perform 20 such iterations, with 5 assigned randomly for each target.

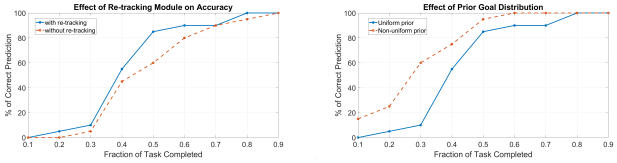
V. Results

The expert dataset consists of 200 hand motion trajectories with velocity, acceleration and jerk data at each time instant. Each trajectory in the dataset was annotated with the corresponding target location. Data was recorded at 30Hz. A total of 180 trajectories were used for training and remaining



(a) (b)

Fig. 3: (a) Average DTW costs of predicted trajectories with respect to the ground truth at various fractions of task completed (b) Average target prediction accuracy as the task progresses



(a) (b)

Fig. 4: Average target prediction accuracy using the proposed approach (a) With and without re-tracking (b) Uniform and non-uniform prior

20 for testing. 10-fold cross-validation was used to evaluate the results. Average values were reported. The proposed approach was validated with state-of-the-art methods under each category of prediction methods as described in Section II - (1) Physics-based model: Classical kinematics of motion ($CM_{k=5}$) [49] (2) Probabilistic graphical model: Bayesian predictor for human motion trajectory (BP-HMT) [50] (3) RNN model: Recurrent neural network-Inverse kinematics-Modified Kalman filtering (RNNIK-MKF) [27], and (4) IRL/IOC model: Path Integral-Inverse Reinforcement Learning (PI-IRL) [51].

To evaluate trajectory prediction, dynamic time warping (DTW) [52] was used to compare two time-series data that might vary in length. It calculates an optimal match between two time-series based on distance metrics. In this work, Euclidean distance was used to estimate DTW cost with the fraction of task completed as shown in Fig 3a. Fraction of task completed is ratio of the distance traversed by the hand to the total distance between the start and the target. To evaluate target prediction, accuracy metric [53] was used, given by:

$$accuracy = \frac{Number\ of\ correct\ target\ predictions}{Total\ number\ of\ predictions} \quad (11)$$

Accurate prediction results in a colored block matching its corresponding colored target. Target prediction accuracy as the task progresses is shown in Fig. 3b. The proposed approach was evaluated with and without re-tracking as shown in Fig. 4a. Note that in all cases, the reward learned and accuracy achieved were based on re-tracked hand motion. For *without re-tracking* analysis, only the partial trajectory was obtained without further re-tracking. Equation (10) demonstrates that the choice of an appropriate prior goal distribution influences target prediction. The proximity of the black and red targets to the user results in faster reach, leading to challenges in measuring the corresponding feature distribution due to high variance in user demonstrations. Consequently, this affects reward distribution and reduces prediction accuracy. To address this, a higher non-uniform prior probability of 0.3 was assigned to both black and red targets and 0.2 to each of the yellow and green targets. This non-uniform prior distribution of goal was analyzed with a uniform prior distribution of 0.25 each as shown in Fig. 4b. Note that target and trajectory prediction accuracy was evaluated using uniform prior. To estimate the ergonomic stress on the user throughout the task, PWL and MWL were used to compute the tolerance index (TI) [41] given by:

$$TI_i = \alpha PWL_i + MWL_i \quad (12)$$

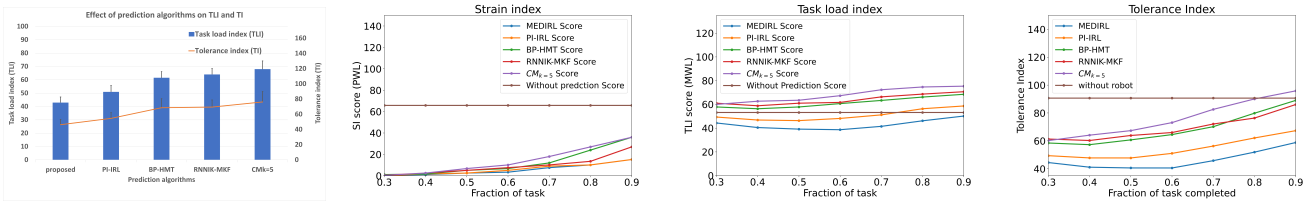
where PWL represents strain index (SI) score, MWL corresponds to the task load index (TLI) score and i denotes the fraction of task completion. Effect of various prediction algorithms on TLI and TI is shown in Fig. 5a. To account for the differing ranges of SI and TLI, a normalization weight of $\alpha = 0.569$ was used which is the ratio of maximum range of TLI (100) to maximum possible range of SI (175.5) for the task. Variation of SI, TLI and TI values throughout the task is shown in Fig. 5b, 5c, 5d.

VI. Discussion

As the task advances, prediction accuracy improves as evident from the decreasing DTW cost and increasing target prediction accuracy. From Fig. 3a, it is noted that MEDIRL and PI-IRL exhibit lower DTW, ranging from $[74, 76]mm$ at the beginning to $[23, 28]mm$ at the end, in contrast to other methods, which show DTW values ranging from $[85, 88]mm$ at the start to $[45, 52]mm$ at the end. The proposed method accurately predicted trajectories with DTW ranging from $[74, 23]mm$ from start to end, outperforming PI-IRL with values of $[76, 28]mm$ throughout the task. This improvement is attributed to the neural network's ability in MEDIRL to capture complex reward distributions, which also contributes to higher target prediction accuracy as seen in Fig. 3b. At 50% completion of the task, MEDIRL achieves the highest average target prediction accuracy of 85%, surpassing PI-IRL, BP-HMT, RNNIK-MKF and $CM_{k=5}$ which achieve accuracies of 75%, 35%, 15% and 10% respectively. At 60% of the task, accuracy reaches 90% using the proposed method. Accuracy approaches 99.9% when more than 80% of the task is completed.

Cases of human error, where the user selects a block of a different color than displayed, were not considered. However, even in such cases, the target was predicted based on the demonstrated partial motion (Experimental video: Link). This means that the robot makes predictions solely based on the demonstrated motion, without knowledge of the color. In some instances, the robot may initially move to an incorrect takeover position due to incorrect trajectory prediction from the initial tracking. However, it soon corrects itself as more tracking data becomes available.

Fig. 4a illustrates that without re-tracking, target prediction



(a) (b) (c) (d)

Fig. 5: (a) Effect of various prediction algorithms on TLI and TI. Variation of (b) Strain index (c) Task load index (d) Tolerance index, as the task progresses

accuracy drops by 25% compared to re-tracking at 50% task completion. This decrease occurs because of noisy and spurious trajectories tracked due to fast and quick hand motion. Re-tracking improves accuracy by smoothing and aligning trajectories with the ground truth through prediction. From Fig. 4b, employing a non-uniform prior by assigning higher probabilities to proximal targets increases prediction probability by 10% at 50% task completion. This indicates that re-tracking has a more significant impact on target prediction accuracy compared to the choice of prior goal distribution.

One-way ANOVA was performed to assess the impact of prediction algorithms on SI, TLI, and TI. Results indicate that while prediction algorithms had no significant impact on strain index (SI), they did influence TLI ($F(4,30) = 30.7788, p < 0.05, \eta^2 = 0.8047$) and TI values ($F(4,30) = 10.0868, p < 0.05, \eta^2 = 0.5737$). In Fig. 5a, the proposed method exhibits the lowest average TLI and TI compared to other methods. Paired t-test confirms greater effect of the proposed method ($p < 0.05$) on TI and TLI compared to other prediction algorithms. Fig. 5b indicates that the physical workload on the user during *human-robot handover* increases as the task progresses, with SI scores ranging from $[0.125, 36]$. However, this remains lower than *reaching task without robot*, which has an SI score of 65.813. This difference was due to the increased user effort required to reach distant targets, leading to physical stress. In Fig. 5c, except for $CM_{k=5}$, MWL initially decreases until reaching 55% of the task, after which it starts increasing. For $CM_{k=5}$, TLI values increase from the beginning of the task. However, the MWL in *reaching task without a robot*, with a TLI of 53.167, was lower than that of $CM_{k=5}$, RNNIK-MKF, and BP-HMT. This difference can be attributed to users feeling uncomfortable with incorrect target predictions and actively working to improve their performance with each iteration, leading to increased mental workload. Tolerance index from Fig. 5d indicates that until 80% of the task, the TI value (90.66) for *reaching task without robot* was the highest compared to prediction-based algorithms. This highlights the potential of prediction-based handovers for reaching tasks. Notably, TI consistently increases for $CM_{k=5}$ from the start of the task, while it decreases until 40% for RNNIK-MKF and BP-HMT, and 45% for PI-IRL. With the proposed method, TI decreases until 60% of the task and increases thereafter. This TI decrease suggests an improved ergonomic collaborative workspace (ECW), where users can comfortably work with reduced ergonomic stress. This is attributed to the system's improved prediction performance compensating for increased physical workload as users reach greater distances, thereby reducing user mental workload and enhancing the collaborative workspace's ergonomics.

VII. Conclusion

This work presents an IRL-based method with re-tracking for predicting hand movement and intended target for handovers. Maximum-Entropy Deep IRL was leveraged to learn the reward distribution along with a set of feature functions specific to the handover task. A target prediction metric was analyzed with uniform and non-uniform prior goal distributions. User studies were conducted for reaching task without-robot and with robot-assisted handovers with prediction. Quantitative and qualitative analysis were performed to calculate prediction accuracy and estimate physical and mental workload throughout the task. While the proposed approach achieves a higher average target prediction accuracy, individual accuracy for nearby targets was lower than for distant ones. This was due to a high variance in reward function due to suboptimal demonstrations of different users. Future work will explore the use of various modalities to enhance handover predictions and address higher-dimensional workspace challenges in industrial environments.

References

[1] J. E. Colgate and M. A. Peshkin, "Cobots," Patent US5 952 796A, 1999.

[2] "ISO 8373:1994 - manipulating industrial robots - vocabulary," International Organization for Standardization, 1994, standard.

[3] R. Parween, M. V. Heredia, M. M. Rayguru, R. E. Abdulkader, and M. R. Elara, "Autonomous self-reconfigurable floor cleaning robot," *IEEE Access*, vol. 8, pp. 114 433–114 442, 2020.

[4] Y.-E. Lee, H. M. Husin, M.-P. Forte, S.-W. Lee, and K. J. Kuchenbecker, "Learning to estimate palpation forces in robotic surgery from visual-inertial data," *IEEE Transactions on Medical Robotics and Bionics*, 2023.

[5] S. Li, P. Zheng, J. Fan, and L. Wang, "Toward proactive human–robot collaborative assembly: A multimodal transfer-learning-enabled action prediction approach," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 8, pp. 8579–8588, 2021.

[6] W. Wang, R. Li, Y. Chen, Y. Sun, and Y. Jia, "Predicting human intentions in human–robot hand-over tasks through multimodal learning," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2339–2353, 2021.

[7] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 299–306.

[8] M. Ragaglia, A. M. Zanchettin, and P. Rocco, "Safety-aware trajectory scaling for human-robot collaboration with prediction of human occupancy," in *2015 international conference on advanced robotics (ICAR)*. IEEE, 2015, pp. 85–90.

[9] J. Mainprice, R. Hayne, and D. Berenson, "Predicting human reaching motion in collaborative tasks using inverse optimal control and iterative re-planning," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 885–892.

[10] A. Ajoudani, P. Albrecht, M. Bianchi, A. Cherubini, S. Del Ferraro, P. Fraisse, L. Fritzsche, M. Garabini, A. Ranavolo, P. H. Rosen *et al.*, "Smart collaborative systems for enabling flexible and ergonomic work practices [industry activities]," *IEEE Robotics & Automation Magazine*, vol. 27, no. 2, pp. 169–176, 2020.

[11] M. Lagomarsino, M. Lorenzini, P. Balatti, E. De Momi, and A. Ajoudani, "Pick the right co-worker: Online assessment of cognitive ergonomics in human-robot collaborative assembly," *IEEE Transactions on Cognitive and Developmental Systems*, 2022.

[12] W. Kim, M. Lorenzini, P. Balatti, P. D. Nguyen, U. Pattacini, V. Tikhanoff, L. Peternel, C. Fantacci, L. Natale, G. Metta *et al.*, "Adaptable workstations for human-robot collaboration: A reconfigurable framework for improving worker ergonomics and productivity," *IEEE Robotics & Automation Magazine*, vol. 26, no. 3, pp. 14–26, 2019.

[13] A. Elnagar, "Prediction of moving objects in dynamic environments using kalman filters," in *Proceedings 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation (Cat. No. 01EX515)*. IEEE, 2001, pp. 414–419.

[14] Y. Maeda, T. Hara, and T. Arai, "Human-robot cooperative manipulation with motion estimation," in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, vol. 4. Ieee, 2001, pp. 2240–2245.

[15] B. Corteville, E. Aertbeliën, H. Bruyninckx, J. De Schutter, and H. Van Brussel, "Human-inspired robot assistant for fast point-to-point movements," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3639–3644.

[16] R. Luo and D. Berenson, "A framework for unsupervised online human reaching motion recognition and early prediction," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 2426–2433.

[17] Z. Wang, P. Jensfelt, and J. Folkesson, "Modeling spatial-temporal dynamics of human movements for predicting future trajectories," in *Workshop at the Twenty-Ninth AAAI Conference on Artificial Intelligence," Knowledge, Skill, and Behavior Transfer in Autonomous Robots", AAAI Conference on Artificial Intelligence, Austin, USA, January 25, 2015*. Association for the advancement of Artificial Intelligence, 2015.

[18] Z. Huang, Y.-J. Mun, X. Li, Y. Xie, N. Zhong, W. Liang, J. Geng, T. Chen, and K. Driggs-Campbell, "Hierarchical intention tracking for robust human-robot collaboration in industrial assembly tasks," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9821–9828.

[19] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 14–29, 2015.

[20] Y. Jiang and A. Saxena, "Modeling high-dimensional humans for activity anticipation using gaussian process latent crfs." in *Robotics: Science and systems*. Berkeley, CA, 2014, pp. 1–8.

[21] C.-M. Huang, S. Andrist, A. Sauppé, and B. Mutlu, "Using gaze patterns to predict task intent in collaboration," *Frontiers in psychology*, vol. 6, p. 1049, 2015.

[22] S. Pellegrinelli, H. Admoni, S. Javdani, and S. Srinivasa, "Human-robot shared workspace collaboration via hindsight optimization," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 831–838.

[23] J. Mainprice, R. Hayne, and D. Berenson, "Goal set inverse optimal control and iterative replanning for predicting human reaching motions in shared workspaces," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 897–908, 2016.

[24] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4346–4354.

[25] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2891–2900.

[26] P. Kratzer, M. Toussaint, and J. Mainprice, "Prediction of human full-body movements with motion optimization and recurrent neural networks," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1792–1798.

[27] R. Liu and C. Liu, "Human motion prediction using adaptable recurrent neural networks and inverse kinematics," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1651–1656, 2020.

[28] O. M. Manyar, Z. McNulty, S. Nikolaidis, and S. K. Gupta, "Inverse reinforcement learning framework for transferring task sequencing policies from humans to robots in manufacturing applications," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 849–856.

[29] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey *et al.*, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[30] B. Berret, E. Chiovetto, F. Nori, and T. Pozzo, "Evidence for composite cost functions in arm movement planning: an inverse optimal control approach," *PLoS computational biology*, vol. 7, no. 10, p. e1002183, 2011.

[31] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization," in *International conference on machine learning*. PMLR, 2016, pp. 49–58.

[32] S. Hosoma, M. Sugasaki, H. Arie, and M. Shimosaka, "Rrt-based maximum entropy inverse reinforcement learning for robust and efficient driving behavior prediction," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 1353–1359.

[33] Z. Wu, L. Sun, W. Zhan, C. Yang, and M. Tomizuka, "Efficient sampling-based maximum entropy inverse reinforcement learning with application to autonomous driving," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5355–5362, 2020.

[34] T. Gu, J. Atwood, C. Dong, J. M. Dolan, and J.-W. Lee, "Tunable and stable real-time trajectory planning for urban autonomous driving," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2015, pp. 250–256.

[35] M. Kalakrishnan, P. Pastor, L. Righetti, and S. Schaal, "Learning objective functions for manipulation," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1331–1336.

[36] M. Wulfmeier, P. Ondruska, and I. Posner, "Maximum entropy deep inverse reinforcement learning," *arXiv preprint arXiv:1507.04888*, 2015.

[37] L. Gan, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "Energy-based legged robots terrain traversability modeling via deep inverse reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8807–8814, 2022.

[38] R. Yu and Z. Zhou, "Towards robust human trajectory prediction in raw videos," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 8059–8066.

[39] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Tracking by prediction: A deep generative model for mutli-person localisation and tracking," in *2018 IEEE Winter conference on applications of computer vision (WACV)*. IEEE, 2018, pp. 1122–1132.

[40] L. F. Figueredo, R. C. Aguiar, L. Chen, S. Chakrabarty, M. R. Dogar, and A. G. Cohn, "Human comfortability: Integrating ergonomics and muscular-informed metrics for manipulability analysis during human-robot collaboration," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 351–358, 2020.

[41] L. Liu, A. J. Schoen, C. Henrichs, J. Li, B. Mutlu, Y. Zhang, and R. G. Radwin, "Human robot collaboration for enhancing work activities," *Human Factors*, p. 00187208221077722, 2022.

[42] O. R. Center. (2020) O*net. (2020). national center for o*net development. the o*net® content model. o*net resource center. [Online]. Available: https://www.onetcenter.org/content.html

[43] J. Steven Moore and A. Garg, "The strain index: a proposed method to analyze jobs for risk of distal upper extremity disorders," *American Industrial Hygiene Association Journal*, vol. 56, no. 5, pp. 443–458, 1995.

[44] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[45] N. Deo and M. M. Trivedi, "Trajectory forecasts in unknown

environments conditioned on grid-based plans," *arXiv preprint arXiv:2001.00735*, 2020.

[46] B. Ziebart, A. Dey, and J. A. Bagnell, "Probabilistic pointing target prediction via inverse optimal control," in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, 2012, pp. 1–10.

[47] E. Lank, Y.-C. N. Cheng, and J. Ruiz, "Endpoint prediction using motion kinematics," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2007, pp. 637–646.

[48] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[49] N. M. Gamage, D. Ishtaweera, M. Weigel, and A. Withana, "So predictable! continuous 3d hand trajectory prediction in virtual reality," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 332–343.

[50] Q. Li, Z. Zhang, Y. You, Y. Mu, and C. Feng, "Data driven models for human motion prediction in human-robot collaboration," *IEEE Access*, vol. 8, pp. 227 690–227 702, 2020.

[51] S. Tian, X. Liang, and M. Zheng, "An optimization-based human behavior modeling and prediction for human-robot collaborative disassembly," in *2023 American Control Conference (ACC)*. IEEE, 2023, pp. 3356–3361.

[52] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the 3rd international conference on knowledge discovery and data mining*, 1994, pp. 359–370.

[53] P. Biswas and P. Langdon, "Multimodal target prediction model," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 2014, pp. 1543–1548.