Development of a Teaching-Learning-Prediction-Collaboration Model for Human-Robot Collaborative Tasks

Omar Obidat, Jesse Parron, Rui Li, Julia Rodano, and Weitian Wang*, *Member, IEEE*

Abstract—**Human-robot collaboration has been one of the main focuses for both research and usage in advanced manufacturing. In human-robot partnerships, instead of static collaboration for repetitive tasks, it is more significant for the robot to dynamically understand its human partner's intentions and collaborate with them to complete the shared tasks. Motivated by these issues, we develop a model for the robot to learn to complete tasks by watching and analyzing human demonstrations. This allows the robot to become more accurate and customizable with each human's personalized working preference. Based on the long short-term memory method, we propose a new approach to have the robot recognize objects, understand ongoing human actions, and predict human intentions. This will allow the robot to automatically adjust its motions and dynamically pick up and deliver the object to its human partner in the collaborative task. Experimental results suggest that the proposed model can enable robots, like humans, to learn and predict humans' intentions dynamically and intelligently to accommodate customized and personalized collaborative tasks. Future work of this study is also discussed.**

Keywords— **Robotics, human-robot collaboration, learning from demonstrations, smart manufacturing.**

I. Introduction

Robots have served as humans' assistants in numerous application fields [1]. In a manufacturing context, for example, robots working alongside humans to complete shared tasks is not an uncommon sight [2-4]. Robots can handle repetitive tasks, freeing up human workers to focus on more complex and creative work. It is also used in healthcare, where robots can be used to assist with surgery or to provide physical therapy to patients [5]. Education is also a field that has been integrating robots to provide tangible and personalized learning experiences to students [6]. In the entertainment industry, robots can be used to create immersive and interactive experiences for audiences [7]. The use of collaborative robots to assist human workers in the assembly and manufacturing of cars is a widely adopted practice in the automotive industry. Assembling tasks have led to the development of highly specialized robotic systems designed to operate safely when near human co-workers. However, ensuring effective human-robot collaboration while guaranteeing the safety of the human co-worker remains a persistent challenge that researchers and engineers are continuously seeking to overcome. The need to develop robust and reliable control algorithms, coupled with the incorporation of advanced sensing and perception capabilities, is imperative in the design and development of collaborative robots that can operate alongside humans while maintaining safety and task efficiency.

Efficient robot programming improves the productivity and reliability of human-robot collaboration. It also reduces the costs in a wide range of industries [8-10]. Robot programming can be a complex and time-consuming process that requires specialized knowledge and expertise. However, advances in machine learning and artificial intelligence are making it possible to teach robots to perform tasks through human demonstration, which can significantly improve the efficiency of the programming process [11]. This approach involves a human demonstrating a task to the robot, which then learns how to perform the task through imitation. In a manufacturing setting, for example, human co-workers can demonstrate tasks to robots, in turn, producing tasks independently. This approach is able to save time and lower costs by reducing the need for extensive programming and improving the deployment of robot systems.

To facilitate effective human-robot collaboration, it is crucial to have robots work seamlessly with humans while minimizing the risks to human safety. One approach to achieve this is using human intention prediction. This involves developing robots to accurately predict human intentions, thereby enabling them to work in harmony with human partners. Predicting human intentions will improve the speed and accuracy of human-robot collaboration, as well as enhance overall task efficiency. There are different ways to predict human intentions, and recent advances in artificial intelligence and machine learning have made it possible to develop systems that can learn and adapt to human behaviors [12]. For instance, a robot can use machine learning algorithms to analyze human movements and gestures, enabling it to anticipate the next move of its human partner [13]. Additionally, robots can use natural language processing to understand human speech, allowing them to respond appropriately to instructions given by humans [14]. Overall, the development of robotic systems that can predict human intentions and adapt to human actions is essential to enable effective and safe human-robot collaboration. With continued research and development in this area, the potential applications for collaborative robots across a wide range of industries will only continue to expand.

Human-robot collaboration has been expanding in recent years, but it is still a relatively new field that has many limitations with the current implementation. One limitation is the complexity of the programming process for collaborative robots. While advances in the latest technologies have made it possible to teach robots through human demonstration, the process is still time-consuming and requires specialized expertise. Additionally, the programming of collaborative robots needs to account for the dynamic nature of human behaviors, which can be difficult to predict. Moreover, there is a need to investigate more flexible and adaptable robot systems that can work effectively across a range of different applications and environments [15]. The current research progress has primarily focused on specific use cases, such as manufacturing or healthcare. However, there are still some gaps in developing more generalized systems and approaches to deploy robots across multiple industries and scenarios.

To this end, we propose a teaching-learning-prediction-collaboration (TLPC) model for the robot to learn from human demonstrations of how to cooperatively complete a

The authors are with the Department of Computer Science, Montclair State University, Montclair, NJ 07043 USA (e-mail: wangw@montclair.edu).

task and become more accurate and customizable with each human's personalized working preference. In our approach, the robot can learn different strategies to complete a task using a finite-state machine model. Based on the long short-term memory algorithm, we develop a new approach to have the robot recognize objects, understand ongoing human actions, and predict human intentions. This will allow the robot to automatically adjust its motions and dynamically pick up and deliver the object to its human partner in the collaborative task.

II. Development of the TLPC Model

A. Approach Overview

The overarching vision of this work is to enable robots, like humans, to learn and predict humans' intentions dynamically and intelligently to accommodate customized and personalized collaborative tasks. As shown in Fig. 1, the TLPC model includes human teaching, robot learning, human intention prediction, and human-robot collaboration.

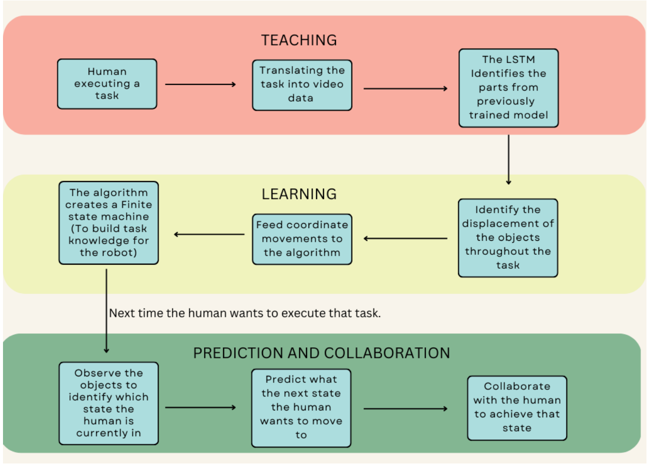

TEACHING
Human executing a task
Translating the task into video data
The LSTM Identifies the parts from previously trained model
LEARNING
The algorithm creates a Finite state machine (To build task knowledge for the robot)
Feed coordinate movements to the algorithm
Identify the displacement of the objects throughout the task
Next time the human wants to execute that task.
PREDICTION AND COLLABORATION
Observe the objects to identify which state the human is currently in
Predict what the next state the human wants to move to
Collaborate with the human to achieve that state

Fig. 1 The teaching-learning-prediction-collaboration (TLPC) model.

In the teaching process, the human demonstrates a task (e.g., assembling a product) and the robot captures the demonstration into video data through a vision system. This data is then processed by a Long Short-Term Memory (LSTM) neural network to identify the locations of the parts required for executing the task. In the learning section, the coordinates of each part are tracked throughout the human demonstration and their movements are fed to an algorithm that creates a finite-state machine to build task knowledge for the robot. A finite-state machine can be used to represent a system or process as a set of states and transitions between those states. In the context of this work, the states represent different configurations of the task (e.g., different positions of the parts) and the transitions represent the movements of the parts between those configurations. In the prediction and collaboration processes, the robot predicts the next state of the task that the human partner wants to operate based on the finite-state machine. By observing which part of the task the human starts with, the robot will collaborate with the human partner by handing the next most suitable part required for completing the task.

This design allows the TLPC model to not require a high level of knowledge from the human to teach the robot to assist the human in collaborative tasks, which makes it a user-friendly and adaptable solution. Additionally, not using any wearable sensors and only using a vision system to identify the tasks and human intentions enhances the naturalness and ease of communication in human-robot partnerships. Overall, the proposed approach has the potential to improve the quality and speed of manufacturing processes, leading to greater productivity and profitability. The TLPC model also has the ability to identify different human participants' working preferences and customize the robot's responses based on how each individual is most likely to complete the task, further increasing efficiency and effectiveness.

B. Long Short-term Memory

The Long Short-Term Memory (LSTM) neural network is utilized in the TLPC model to process the video data obtained from human demonstrations and identify the locations of the parts required for executing the task. The LSTM is a type of recurrent neural network that is particularly useful for modeling sequential data. It can retain information over long periods, making it well-suited for tasks where past inputs are important in making predictions. Fig. 2 shows the basic structure of the LSTM neural network, where X_t is the current input, C_t is the new updated memory, h_t is the current output, C_{t-1} is the memory from the last LSTM unit, h_{t-1} is the output of the last LSTM unit. In the LSTM, a forget gate layer decides what information to keep. Then, an input gate layer and tanh layer determine new information to add. The old state is then updated by forgetting selected information and adding new information. Finally, a sigmoid layer filters the output based on the cell state [16].

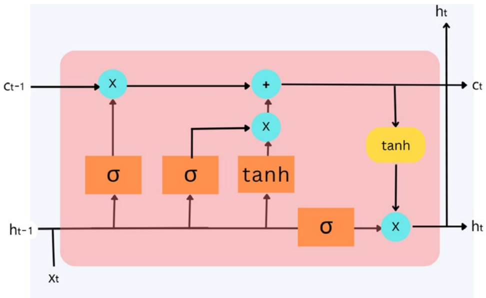


Fig. 2. Structure of the LSTM neural network.

C. Human Teaching and Robot Learning

A web camera is used in the robot vision system to process the human worker's operation movements and the sequence of events required to complete a task in the teaching process. The minimal amount of equipment needed for this method allows for cost reduction and facilitates pairing with other algorithms and technologies to enhance long-term human-robot collaboration. In the robot learning process, we aim to minimize the requirements and dependencies for implementing human-robot collaboration while maintaining stability and consistency. This ensures optimal results in terms of productivity and natural communication in human-robot partnerships. Based on the LSTM, the mathematical representation of human teaching and robot learning is described as follows:

$$rl\_k(t) = LSTM(hd\_v(t), d\_i(t), l(t)) \tag{1}$$

where $rl\_k(t)$ is the robot's learned knowledge at time t, $hd\_v(t)$ means the LSTM processes the video data of the human demonstrations at time t to identify the location of each part in a 3x3 grid and highlights the corresponding blocks, $d\_i(t)$ denotes the LSTM tracks the movements of the parts and detects it when a part is operated, $d\_i(t) = 1$ if part i is missing at time t and 0 otherwise, and $l(t) = [d\_1(t), d\_2(t),$

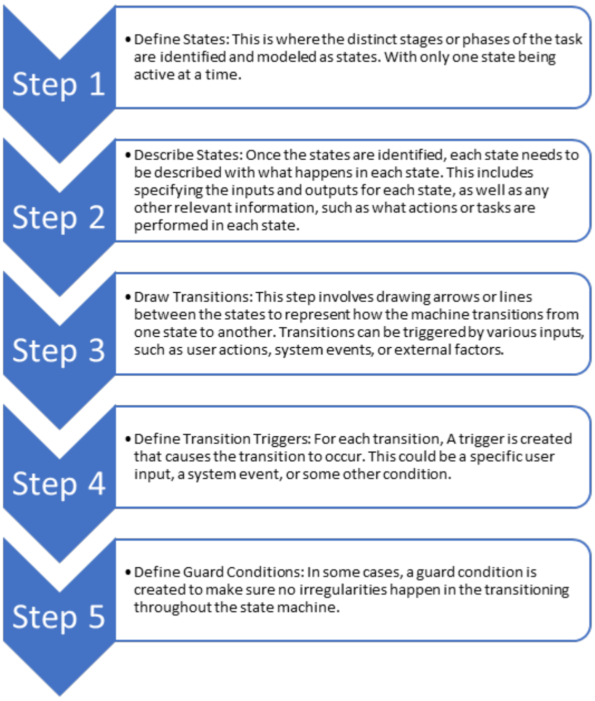
d_3(t), d_4(t)] represents the LSTM registers each missing part in a list at time *t*. Based on its learned knowledge and the missing parts, the robot then determines the appropriate actions and sequences to complete the task with its human partner.

D. Human Intention Prediction

In the human intention prediction process, a finite-state machine (FSM) is created for the robot to build task strategies and predict human intentions based on the working preference of a specific human worker employed in the teaching process. To create a finite-state machine for a task, as shown in Fig. 3, we follow 5 main steps. The created finite-state machine is expressed as:

$$M = (S, I, O, f, g, s_0) \tag{2}$$

which includes a finite set S of states, a finite input I, a finite output O, a transition function f $(f : S \times I \rightarrow S)$, an output function g $(g : S \times I \rightarrow O)$, and an initial state s_0 [17].



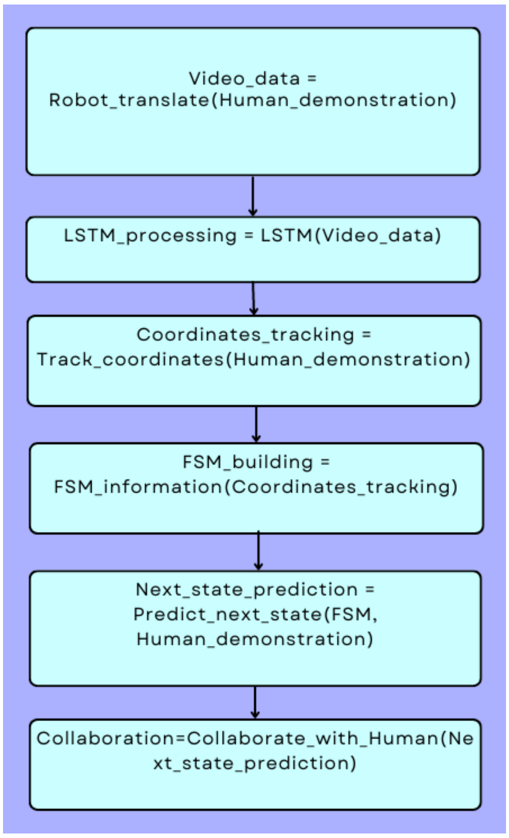
- Define States: This is where the distinct stages or phases of the task are identified and modeled as states. With only one state being active at a time.
- Describe States: Once the states are identified, each state needs to be described with what happens in each state. This includes specifying the inputs and outputs for each state, as well as any other relevant information, such as what actions or tasks are performed in each state.
- Draw Transitions: This step involves drawing arrows or lines between the states to represent how the machine transitions from one state to another. Transitions can be triggered by various inputs, such as user actions, system events, or external factors.
- Define Transition Triggers: For each transition, A trigger is created that causes the transition to occur. This could be a specific user input, a system event, or some other condition.
- Define Guard Conditions: In some cases, a guard condition is created to make sure no irregularities happen in the transitioning throughout the state machine.

Fig. 3. Steps to create a finite-state machine for human intention prediction.

E. Human-robot Collaboration

Based on the robot's prediction of human intentions in shared tasks, we design a human-robot collaboration model for the robot to accommodate its human partners' working preferences and assist humans. This model is not only flexible and adaptable but also easily customizable to suit the specific needs of other human-robot collaborative applications. The collaboration model building process based on the TLPC framework is presented in Fig. 4.

As shown in Fig. 4, *Human_demonstration* represents the input data provided by the human participant to demonstrate a task to the robot. *Robot_translate( )* is a function that translates the human demonstration into video data, which is assigned to the variable *Video_data*. *LSTM( )* is a function that processes the *Video_data* using an LSTM neural network to identify the locations of the parts required for executing the task. The output of this function is assigned to the variable *LSTM_processing*. *Track_coordinates( )* is a function that tracks the coordinates of each part throughout the human demonstration, and assigns the tracked coordinates to the variable *Coordinates_tracking*. *FSM_information( )* is a function that takes in the *Coordinates_tracking* data and builds a finite-state machine to create task knowledge for the robot. The output of this function is assigned to the variable *FSM_building*. *Predict_next_state( )* is a function that uses the *FSM_building* and *Human_demonstration* data to predict the next state that the human participant wants to move to based on the finite-state machine. The output of this function is assigned to the variable *Next_state_prediction* to predict the human intention in the task. *Collaborate_with_Human( )* is a function for the robot to collaborate with its human partner by handing the next most suitable part required for completing the task.



```
Video_data =
Robot_translate(Human_demonstration)

LSTM_processing = LSTM(Video_data)

Coordinates_tracking =
Track_coordinates(Human_demonstration)

FSM_building =
FSM_information(Coordinates_tracking)

Next_state_prediction =
Predict_next_state(FSM,
Human_demonstration)

Collaboration=Collaborate_with_Human(Ne
xt_state_prediction)
```

Fig. 4. The collaboration model building process based on the TLPC framework.

III. Experimental Results and Analysis

A. Experimental Platform

We develop a high-fidelity advanced manufacturing context to testify the proposed model. As shown in Fig. 5, the experimental platform includes a collaborative robot, a vehicle model to be assembled, a web camera, and a shared workspace. The human participant will assemble the parts of the vehicle to teach the robot how to perform the task alongside the human participant the next time the human participant is performing the task. The robot used is a Franka Emika Panda robot, which is a 7-DoF collaborative robot [18]. The web camera used is a generic webcam. A ThinkStation P520 workstation is employed to process human demonstration data, run the TLPC model, and send robot control commands in the human-robot collaborative experiments. The Robot Operating System (ROS) is utilized in managing our robot system [19]. ROS is an open-source framework for inter-platform maneuvering and communication on a large scale. In addition, this work utilizes MoveIt! and runs it with the ROS operating system [20]. To plan the robot's movements in human-robot collaborative tasks, the control commands are sent to the libfranka interface, which is a ROS package that allows the collaborative robot to communicate with the FCI controller. The FCI will provide the current robot states and enable the robot to be directly controlled by the commands derived from the TLPC model.

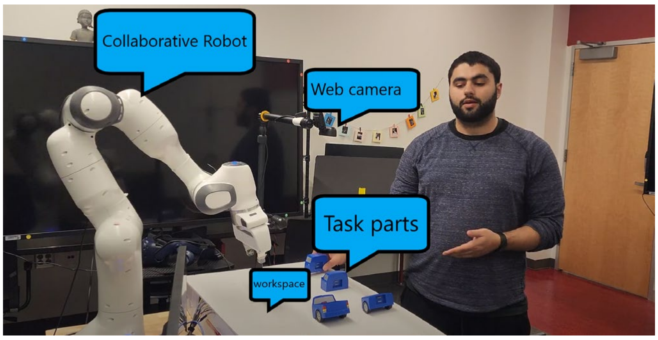


Fig. 5. Experimental platform.

B. Task Description

In this work, we design a typical co-assembly task in advanced manufacturing contexts, in which the robot learns from its human partner to assemble a vehicle model and assists the human in the task through predicting their intentions based on the learned knowledge. The flow of our experiment is the same as our proposed TLPC model. Starting with human teaching where the human participant would perform the given task which in this case is assembling a vehicle model that has four parts (the proposed model is also applicable for more parts). The program starts by asking the user if it should go into the knowledge-building mode where the robot would observe the human participants' operational actions and learn from them how to perform the task or to predict and collaborate with the human participant using the knowledge it accumulated so far. During our experiment, we ran the program in the knowledge-building mode where the webcam fed the LSTM model with real-time human demonstration video data and the LSTM recognized the locations of the vehicle parts. For each part the human participant assembled, the robot built its knowledge with the sequence of how the task was achieved. Following that, we went into prediction and collaboration mode showcasing how the robot learned from human demonstrations. By having the robot observe the human participants' operation movements, it could help the human participant with the following parts based on the learned task strategies. The robot can predict based on as many demonstrations as previously given. Once the whole task course is learned, the human would start the collaborative task with the robot to assemble the vehicle.

C. Results and Analysis of Learning from Demonstrations

The results of robot learning from human demonstrations are shown in Fig. 6. The system uses a camera to capture video data of a human participant performing a task. The video data is then sent to a Long Short-Term Memory (LSTM) to identify the location of each part in a 3x3 grid and highlight the corresponding blocks (the blue parts in each subfigure). This allows the system to track the movement of each part and identify when a part is operated. Following that the human participant starts with the task in any possible route according to his assembly preference and the LSTM tracks the movements of the parts and detects when they are absent. Each time a part is picked up, the LSTM registers this information in a list. Once all parts are out of the camera range, the LSTM writes the created route into a file. The output file is then used to construct a finite-state machine to build the robot's task knowledge. Fig. 6 indicates that the robot accurately learns the task execution procedure.

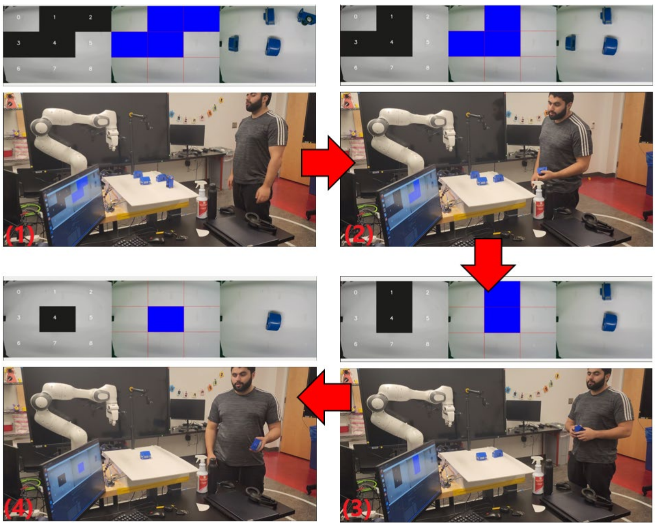


Fig. 6. Learning task from human demonstrations.

D. Results and Analysis of Human Intention Prediction and Human-robot Collaboration

Fig. 7 presents the results of human intention prediction and human-robot collaboration of a co-assembly route for the vehicle model in real-world contexts. In the prediction and collaboration mode of our approach, the robot waits for the human participant to start the task (Fig. 7(1)). Once the first part's movement from its original location is detected (Fig. 7(2)), the robot predicts and picks up what the next part the human participant will require to continue the assembly based on its learned task knowledge (Fig. 7(3)). Then the robot hands it to the human participant to help with conducting the task (Fig. 7(4)). As shown in Fig. 7(5), the human picks up another part to continue the assembly task. After that, the robot picks up the following predicted part (Fig. 7(6)) and delivers it to its human partner to accomplish the task (Fig. 7(7) and Fig. 7(8)). The more demonstrations and assembly routes the robot learns, the more accurate personalized strategies it can employ to assist the human in collaborative tasks. While the human participant is assembling the parts, the robot anticipates and provides the next required part based on the human's assembly preference, thereby facilitating task completion in an optimal time.

The proposed approach enhances collaboration between humans and robots, enabling them to work together more efficiently and effectively. By predicting and delivering the required parts accurately, the robot saves time and effort that would otherwise be spent searching for and retrieving parts. The results of our study evince the successful implementation of our TLPC model in real-world scenarios. Our findings show that the TLPC model can significantly improve existing solutions, advancing efficient human-robot collaboration across diverse shared tasks. In this experiment, we tested the effectiveness of the proposed model by assembling a four-part vehicle, which may appear to be a straightforward task; however, our results demonstrate that the TLPC model is scalable for more complex tasks.

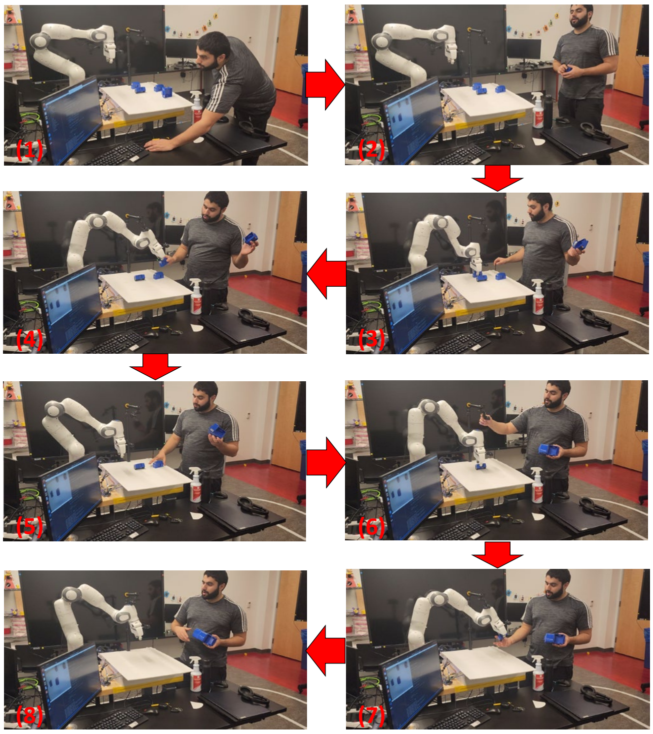


Fig. 7. Human intention prediction and human-robot collaboration for vehicle model assembly.

IV. Conclusions and Future Work

In this study, we have developed a TLPC model, which is an effective solution to program robots to collaborate with humans on tasks that can be completed in different ways. The model utilizes machine learning techniques to identify the locations of parts required for executing a task and employs a finite-state machine-based approach to predict the next state of the shared task in the human-robot collaboration process. This prediction feature allows the robot to collaborate with humans in many ways by handing over the next most suitable part/object required for completing the task, thereby reducing task errors and improving collaboration productivity. One of the significant advantages of the TLPC model is its ability to be customized by human participants according to their working preferences while executing a task. This customization enhances the naturalness and ease of communication between humans and robots to improve the quality and speed of manufacturing processes, leading to greater productivity and profitability. The use of an LSTM neural network to process video data obtained from human demonstrations allows the TLPC model to retain information over long periods, making it well-suited for tasks where past inputs are important in making current predictions. Experimental results suggest that the TLPC model could minimize the requirements for implementing human-robot collaboration while maintaining the stability and consistency of the partnership. This approach eliminates the need for wearable sensors and allows the model to be implemented with minimal equipment, helping to cut costs or pair it with other algorithms and technologies to improve human-robot collaboration in the long term. In our future work, first, we will verify it with more complex tasks in different scenarios. Additionally, we will explore more metrics and conduct subjective evaluation experiments by recruiting participants from various backgrounds and working preferences to assess the developed model, collect their feedback, and iteratively enhance the performance of our approach.

Acknowledgment

This work is supported in part by the National Science Foundation under Grant CMMI-2138351 and in part by the National Science Foundation under Grant CNS-2117308.

References

[1] T. B. Sheridan, "Human–Robot Interaction:Status and Challenges," *Human Factors,* vol. 58, no. 4, pp. 525-532, 2016, doi: 10.1177/0018720816644364.

[2] C. C. Kemp, A. Edsinger, and E. Torres-Jara, "Challenges for robot manipulation in human environments [grand challenges of robotics]," *IEEE Robotics & Automation Magazine,* vol. 14, no. 1, pp. 20-29, 2007.

[3] C. Heyer, "Human-robot interaction and future industrial robotics applications," in *2010 ieee/rsj international conference on intelligent robots and systems*, 2010: IEEE, pp. 4749-4754.

[4] C. Paxton, A. Hundt, F. Jonathan, K. Guerin, and G. D. Hager, "CoSTAR: Instructing collaborative robots with behavior trees and vision," in *2017 IEEE international conference on robotics and automation (ICRA)*, 2017: IEEE, pp. 564-571.

[5] E. Z. Goh and T. Ali, "Robotic surgery: an evolution in practice," vol. 2022, ed: Oxford University Press, 2022, p. snac003.

[6] W. Wang, C. Coutras, and M. Zhu, "Empowering computing students with proficiency in robotics via situated learning," *Smart Learning Environments,* vol. 8, no. 1, pp. 1-18, 2021, doi: 10.1186/s40561-021-00167-6.

[7] R. Bogue, "The role of robots in entertainment," *Industrial Robot: the international journal of robotics research and application,* 2022.

[8] L. Pérez, S. Rodríguez-Jiménez, N. Rodríguez, R. Usamentiaga, D. F. García, and L. Wang, "Symbiotic human–robot collaborative approach for increased productivity and enhanced safety in the aerospace manufacturing industry," *The International Journal of Advanced Manufacturing Technology,* vol. 106, pp. 851-863, 2020.

[9] E. Appleton and D. J. Williams, *Industrial robot applications*. Springer Science & Business Media, 2012.

[10] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," in *Springer handbook of robotics*: Springer, 2008, pp. 1371-1394.

[11] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling Object Hand-Over in Human–Robot Collaboration Via Natural Wearable Sensing," *IEEE Transactions on Human-Machine Systems,* vol. 49, no. 1, pp. 59-71, 2019.

[12] 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, 2022, doi: 10.1109/TASE.2021.3074873.

[13] R. Li, W. Wang, Y. Chen, and Y. Jia, "Natural Language and Gesture Perception Based Robot Companion Teaching for Assisting Human Workers in Assembly Contexts," in *ASME 2019 Dynamic Systems and Control Conference*, 2019, doi: 10.1115/dscc2019-9177.

[14] P. Fung *et al.*, "Towards empathetic human-robot interactions," in *Computational Linguistics and Intelligent Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part II 17*, 2018: Springer, pp. 173-193.

[15] C. Wong, E. Yang, X.-T. Yan, and D. Gu, "An overview of robotics and autonomous systems for harsh environments," in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017: IEEE, pp. 1-6.

[16] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review,* vol. 53, pp. 5929-5955, 2020.

[17] K. H. Rosen, *Discrete mathematics and its applications*, 7th ed. New York: McGraw-Hill (in eng), 2012.

[18] H. Diamantopoulos and W. Wang, "Accommodating and Assisting Human Partners in Human-Robot Collaborative Tasks through Emotion Understanding," in *2021 International Conference on Mechanical and Aerospace Engineering (ICMAE)*, 2021: IEEE, pp. 523-528.

[19] M. Quigley *et al.*, "ROS: an open-source Robot Operating System," in *ICRA workshop on open source software*, 2009, vol. 3, no. 3.2: Kobe, Japan, pp. 1-6.

[20] S. Chitta, I. Sucan, and S. Cousins, "Moveit!," *IEEE Robotics & Automation Magazine,* vol. 19, no. 1, pp. 18-19, 2012.