2025 IEEE International Conference on Robotics and Automation (ICRA)
May 19-23, 2025. Atlanta, USA

ForceMimic: Force-Centric Imitation Learning with Force-Motion Capture System for Contact-Rich Manipulation

Wenhai Liu*, Junbo Wang*, Yiming Wang*, Weiming Wang and Cewu Lu†
Shanghai Jiao Tong University
{sjtu-wenhai, sjtuwjb3589635689, sommerfeld, wangweiming, lucewu}@sjtu.edu.cn
(* Equal contribution, † Corresponding author)

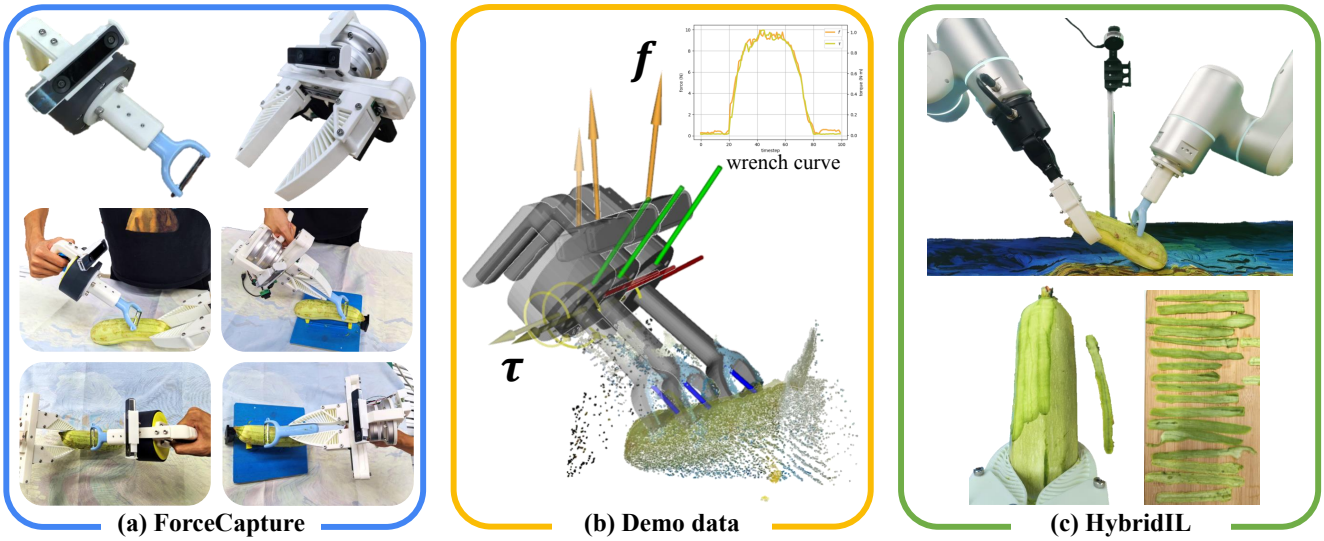


Fig. 1: Overview of **ForceMimic**. **ForceMimic** facilitates (a) **ForceCapture**, a handheld robot-free on-site data collection system, to record (b) high-quality natural force-centric manipulation demonstrations. Leveraging this data, (c) **HybridIL** trains a hybrid force-position control policy to perform contact-rich tasks, such as vegetable peeling.

***Abstract*— In most contact-rich manipulation tasks, humans apply time-varying forces to the target object, compensating for inaccuracies in the vision-guided hand trajectory. However, current robot learning algorithms primarily focus on trajectory-based policy, with limited attention given to learning force-related skills. To address this limitation, we introduce ForceMimic, a force-centric robot learning system, providing a natural, force-aware and robot-free robotic demonstration collection system, along with a hybrid force-motion imitation learning algorithm for robust contact-rich manipulation. Using the proposed ForceCapture system, an operator can peel a zucchini in 5 minutes, while force-feedback teleoperation takes over 13 minutes and struggles with task completion. With the collected data, we propose HybridIL to train a force-centric imitation learning model, equipped with hybrid force-position control primitive to fit the predicted wrench-position parameters during robot execution. Experiments demonstrate that our approach enables the model to learn a more robust policy under the contact-rich task of vegetable peeling, increasing the success rates by 54.5% relatively compared to state-of-the-art pure-vision-based imitation learning. Hardware, code, data and more results can be found on the project website at `https://forcemimic.github.io`.**

I. Introduction

Humans can take use of force sensing, fine muscle force control to achieve better manipulations, from grasping [1], lifting [2] to peeling [3]. The exploitation of force can detect and correct the error brought by vision-based motion planning. Inspired by these neuroscience results, we want to explore the utility of force in robot learning. However, the force-centric manipulation demonstration data is hard to collect. Substantial human videos exist on the Internet, but no interaction force data is recorded. Teleoperation [4] is a popular data collection approach, enabling operators remotely to control robot finishing manipulation tasks. Particularly, force-feedback teleoperation shows a potential path to the force-centric data collection. But it cannot give the operator a natural manipulation experience, harmful to smooth action execution and precise force control. Recently, portable handheld devices [5, 6] make in-the-wild learning possible. They make use of SLAM tracking camera to record human hand or handheld gripper pose trajectory. In addition to the removal of real robot, it provides an additional advantage as almost direct interaction between human and objects, which is critical in contact-rich force-centric manipulation.

On the other hand, robotic imitation learning with force involved is under-explored. Imitation policy learning mimics the function of human cerebellum, and it has been found that the central nervous system can predict the force load and even fuse this dynamics information into the inner model of human motor [1]. So we wonder whether the introduction of

2025 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3315-4139-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICRA55743.2025.11128061

force can help the model learn better and guide the low-level robotic control.

To handle the above challenges, we propose **ForceMimic**, a force-centric robot learning system, providing natural, force-aware and robot-free robotic demonstration collection experience and force-centric imitation learning algorithm equipped with hybrid force-position control to achieve robust contact-rich manipulation (see Fig. 1). We first develop **ForceCapture**, a handheld robot-free data collection system to record high-quality pure interaction wrench, i.e. combination of force and torque, by ratchet locking and gravity compensation. After that, **HybridIL** leverages the data to train a force-aware policy that outputs wrench-position parameters. And we apply hybrid force-position control to fit not only predicted pose trajectory but also the predicted force parameters, achieving robust execution against error-prone visual guide. Using ForceCapture, operators can peel a zucchini in just 5 minutes, while force-feedback teleoperation takes over 13 minutes and struggles with smooth peeling. Robot experiments show that HybridIL achieves a 54.5% higher success rate during contact-rich peeling compared to state-of-the-art vision-based imitation learning.

Overall, our contributions can be summarized as follows:

- We develop ForceCapture, a handheld robot-free data collection system, providing natural, force-aware and on-site force realism collecting experience.
- We provide the HybridIL algorithm, a force-centric imitation learning model that outputs wrench-position parameters and utilizes orthogonal hybrid force-position control primitives to fit the model's predictions.
- We conduct experiments on robot zucchini peeling and achieve more robust performance using our data and model than state-of-the-art pure-vision-based imitation learning algorithms.

II. Related Work

a) Robotic Data Collection System: A direct approach to collect robot manipulation demonstrations is teleoperation [4], where a human operator remotely controls the robot to execute the manipulation task, by various user interfaces, including haptic devices [7], exoskeleton [8–10], virtual reality [11–14] and leader-follower paradigm [15–19]. Teleoperation can gather real robot data, with no domain gap between training and rollout data, but it poses the unintuitive controlling nature between human operators and robots, even when added force feedback. Recently, hand-held grippers [5, 6, 20–22] make in-the-wild learning possible. However, although the hand-held gripper offers a more natural experience during data collection, it does not make the policy model aware of this interaction, with no interaction force recorded.

b) Robot Imitation Learning: Imitation learning (IL) from human expert collected demonstrations has been widely applied in robot learning tasks. Behavior cloning (BC) [23], as one of the simplest methods in IL, directly learns the policy mapping from observations to corresponding robot actions in a supervised manner. Despite its simplicity, BC has shown many exciting results in various robot manipulations. Most methods parameterize the policy using neural networks [17, 24, 25], mapping 2D raw image pixels to the action space, while some non-parametric approaches [26] leverages the nearest neighbor to retrieve actions from the demonstration dataset. Recently, Diffusion Policy [27] conditions on the vision representations and uses diffusion model to denoise the action trajectory. Built upon it, several approaches [28, 29] have been adapted to 3D point clouds as observation. However, current imitation learning approaches focus predominantly on trajectory-based skills, lacking exploration of action spaces such as interaction forces.

Force perception and control plays a crucial role in manipulation tasks, providing valuable and complementary information with visual guidance [30]. Several works have explored the force in contact-rich robotic manipulation, ranging from opening bottle caps [31], assembling [32] to playing Jenga [33]. Recently, MOMA-Force [34] utilizes the visual representation similarity to retrieve target action and wrench from the expert database and uses a PID-based controller [35, 36] to control the robot. ForceSight [37] presents a transformer-based robotic planner that generates force-based objectives given a text input and an RGBD image. In this paper, We propose a new paradigm of using orthogonal hybrid force-position control primitives to fit the model's predicted continue wrench-position parameters.

c) Robot Peeling: While peeling is an important instrumental activity of daily living (IADL), it is relatively under-explored in current robot research field. Dong et al. [38] attempts peeling five types of food by calculating the cutting plane and controlling the constant contact force along the planning trajectory, but this method depends heavily on preset assumption. MORPHeus [39] introduces neural networks to release the hand-crafted perception assumption, but it separates the peeling procedure into several individual modules and preset skills, focusing on high-level skill arrangement. There also exist other works dealing with the peeling problem but using knife [40] or dexterous hand [41], instead of peelers in our setup. In contrast to the aforementioned methods, we approach the peeling task as a force-related skill for end-to-end learning.

III. Method

ForceMimic first employs ForceCapture, a handheld robot-free data collection system (detailed in Sec. III-A), to naturally gather force-centric human demonstration data. Then, we transfer the robot-free data to (pseudo-)robot data (detailed in Sec. III-B), bridging the domain gap. Leveraging this data, HybridIL learns to predict wrench-pose trajectory, and applies hybrid force-position control to fit the predicted force-position parameters (detailed in Sec. III-C), enabling robust performance in contact-rich manipulation tasks. The overall pipeline is illustrated in Fig. 2.

A. Hardware Design: ForceCapture

Accurately, naturally, and cost-effectively capturing force data during contact-rich manipulation remains a significant

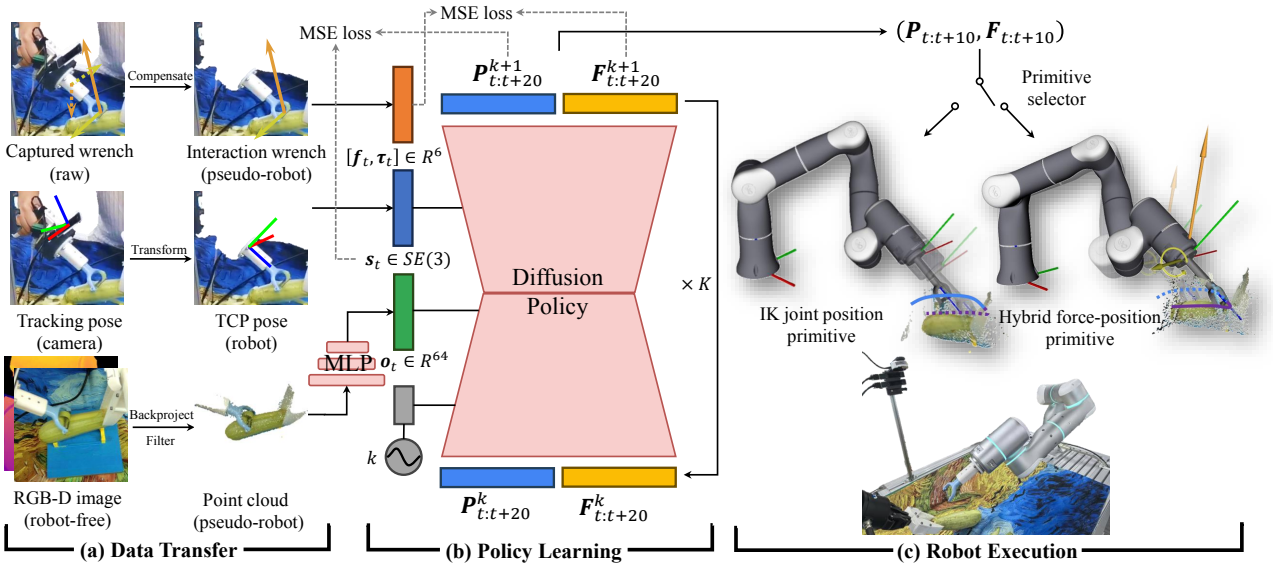


Fig. 2: Overview of the pipeline. (a) We first transfer the collected robot-free data to (pseudo-)robot data, bridging the domain gap. The captured wrench is compensated to account for self-gravity effects. The pose recorded by SLAM camera is transformed as the robot TCP pose. And RGB-D observation images are backprojected into point cloud and filtered out unrelated points. (b) Leveraging this data, a diffusion-based policy is learned, with both pose and wrench predicted, conditioned on the encoded point cloud features, history pose and diffusion timestep embeddings. (c) According to the predicted force value, either IK joint position primitive or hybrid force-position primitive is selected, and fits the output force-position parameters to conduct execution actions.

challenge. Inspired by existing handheld motion data collection devices [5, 6], we developed a low-cost, versatile, and robot-free force-position capture device, ForceCapture. To design ForceCapture, we consistently adhered to the following objectives:

(1) **Scalability**. Key factors for scalability include low cost, compatibility with different force sensors, ease of fabrication and maintenance.

(2) **On-site force realism**. Unlike teleoperation systems that create a sense of presence through force feedback, our goal is to directly capture real-time force data from human operations without requiring users to learn how to interact with artificial environments created by the device.

(3) **Ergonomic comfort**. The device must adhere to ergonomic principles, including an appropriate center of gravity and the convenience of operation, to ensure it does not interfere with the user's natural operating habits. Since accurate interaction force data needs to be recorded, poor ergonomics could alter muscle exertion patterns or cause discomfort, leading to non-natural force data during operation.

The overall design is shown in Fig. 3, which illustrates two versions, one with a fixed tool and the other with an adaptive gripper. At its core, both designs share the feature of a six-axis force sensor placed between the end-effector and the user's gripping handle, which can be used to capture the effector-environment interaction wrench. Additionally, a SLAM camera positioned near the center of the force sensor records the motion data during interaction. The user grips the handle to directly operate the tool or control the fingers for grasping and manipulation tasks. The rack-and-pinion mechanism of the gripper version at the base of

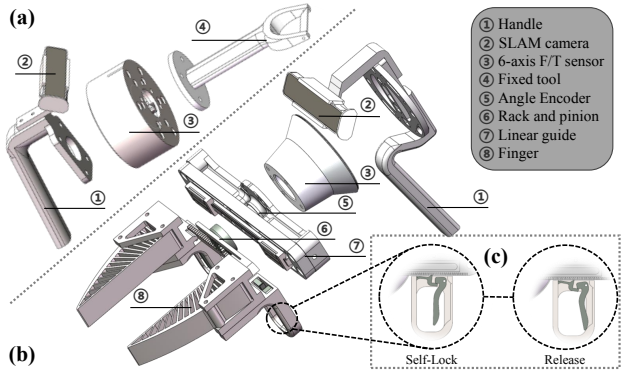


Fig. 3: Structure of ForceCapture. It consists of (a) a fixed-tool end-effector version, and (b) a movable gripper version, which provides (c) a unique self-lock function.

both fingers ensures synchronized movement of the grippers. The pinion is connected to an encoder, which records the opening distance of the grippers. The continuous width value is determined based on the calibrated relationship between the encoder angle and the gripper's width.

It is important to note that during the manual control of the gripper's opening and closing, the forces exerted by the hand on the grippers are also applied to the force sensor. To address this, we designed a unidirectional locking mechanism, as shown in Fig. 3 (c). Once the fingers are closed, they cannot be opened from the fingertip. Instead, they can only be released using a lever mechanism to unlock the gripper. This design aligns with the natural logic of opening and closing the gripper and adheres to ergonomic principles. Additionally, the overall design of ForceCapture,

with its center of mass positioned above the handle, conforms to the natural force application habits of the human hand.

ForceCapture is quite straightforward to manufacture, with the main body fully produced using 3D printing. The total cost of the printed parts and encoder is approximately $50, aligning with the design goal of cost-effectiveness. The weight of the device equipped with the gripper is only 0.8kg, of which the force sensor weighs 0.5kg, and our accessories weigh only 0.3kg, which is even lighter than a can of cola. For more details about ForceCapture, including CAD models, installation instructions and 3D printing materials, please refer to the project website.

B. Data Collection and Transfer

The data collection system includes a six-axis F/T sensor, a RealSense T265 SLAM camera, and an external RealSense L515 RGB-D camera. For the gripper version, encoder angle data is also collected. Their respective sampling frequencies are 1000 Hz, 200 Hz, 30 Hz, and 30 Hz. Each sensor collects data at its own frequency, and during data processing, all frequencies are aligned to match the frequency of L515 observation. At the initial stage, T265 is placed on the L515 mount, and the relative position between the T265 and L515 is determined by their mounting positions. Once data collection begins, the T265 is detached from the mount and placed on ForceCapture. This process is similar to DexCap [6], where the initial position of the T265 relative to the L515 is used to track the position of ForceCapture.

ForceCapture is designed to record only interaction forces between end-effector and external environment. However, the force sensor measures the combined forces, including the tool's gravitational and inertial forces. Therefore, it is necessary to subtract the external forces generated by the tool or gripper from the force sensor data. We assume that the data collection process with ForceCapture is quasi-static, meaning that at each position, the forces are in static equilibrium, and we only need to compensate for the tool's gravity. To perform the gravity compensation, we first move ForceCapture in a quasi-static manner for a certain period while recording the pose and wrench data. Using the static equilibrium forces at each position, we construct an overdetermined system of equations to estimate the tool's center of mass and weight using least-squares solution.

Additionally, RGB-D images recorded by L515 camera are backprojected into point clouds. To reduce discrepancies between the point clouds during data collection and those used in robot deployment, we uniformly exclude point clouds above the operational background and end-effector coordinate systems, retaining only the consistent end-effector and object point clouds. And the point clouds are voxelized to a size of 10,000. The example data transfer process is shown in Fig. 2 (a).

C. Learning Algorithm: HybridIL

This section introduces HybridIL, an end-to-end imitation learning method centered on force, which maps from perception to a force-position hybrid control strategy, as shown in Fig. 2 (b). HybridIL takes point clouds as visual input, which are represented as one-dimensional visual features via an MLP encoder. These features are then cascaded with the robot's TCP pose to form a joint representation of multiple modalities. The strategy generation utilizes modified diffusion policy [27] to predict both position and wrench parameters over the next 20 time steps.

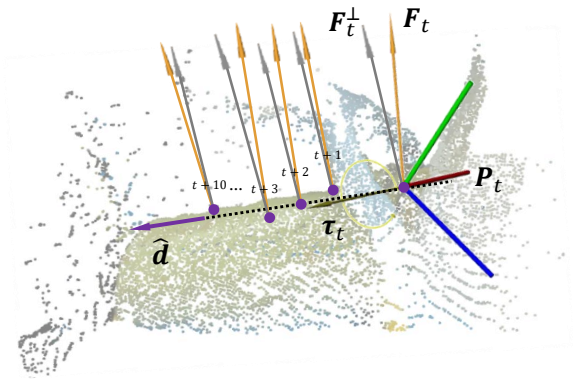


Fig. 4: Illustration of the interface between policy and control primitive. When the hybrid force-position control primitive is active, the motion direction $\hat{\mathbf{d}}$ is calculated based on the pose trajectory $\mathbf{P}_{t:t+10}$ from policy, and the predicted forces $\mathbf{F}_{t:t+10}$ are orthogonalized to $\mathbf{F}^{\perp}_{t:t+10}$. Hybrid force-position control primitive then takes $\hat{\mathbf{d}}$ and $\mathbf{F}_t^{\perp}$ as parameters and controls the robot to track both pose and force.

It is important to note that wrench and position control must be orthogonal. While our model does not explicitly model the orthogonality of wrench and position, we achieve this through an orthogonal force-position hybrid controller that aligns with the model's predicted force-position parameters. This approach differs from conventional imitation learning methods, which typically use a fixed lower-level position controller to track the position commands prediction by the model. HybridIL employs two distinct control primitives to fit the model's predicted force-position parameters, demonstrated in Fig. 2 (c). When the predicted force is below a threshold of 6N, an IK-based [42] joint position controller is used. If the predicted force exceeds 6N over consecutive steps, a hybrid force-position controller is employed to execute the model's predicted parameters. The force threshold of 6N was empirically determined. The orthogonal force-position matching approach is illustrated in Fig. 4. For force-position actions where the force exceeds 6N continuously, the motion direction is determined based on the positional information before and after. The corresponding predicted force information is projected onto the orthogonal plane of the motion direction, which defines the force control parameters during execution. For the initial step of hybrid force-position control, if the end-effector has not yet made contact with the object, a pressing control in the opposite direction of force control is applied to achieve stable contact. These functionalities are realized using Flexiv RDK[1] of joint

[1]https://github.com/flexivrobotics/flexiv_rdk

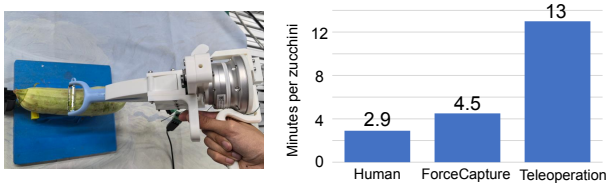
position control and hybrid force-position control primitives to execute the force-position actions of HybridIL.

IV. Experiments

In this section, we perform a zucchini peeling experiment to validate the data collection efficiency of ForceCapture and the effectiveness of HybridIL. All data were collected in an on-site manner, without any involvement of robots in the data acquisition process.

A. Collection Efficiency: ForceCapture vs. Teleoperation

Currently, simultaneous collection of pose trajectory and six-axis wrench data primarily relies on teleoperation. To compare the efficiency of teleoperation with ForceCapture, we conducted a case study of peeling a zucchini using a single-arm. The experimental setup is shown in Fig. 5 (a). The procedure involved picking up the peeler, peeling the zucchini on a stand, placing the peeler down, then grasping the zucchini to adjust its orientation for peeling until the entire vegetable was peeled. Since the task involved force capture and finger movements, we used the gripper version of ForceCapture for data collection. The teleoperation setup follows the configuration described in RH20T [7].



(a) Setup **(b) Efficiency comparison**

Fig. 5: Experimental setup for data collection efficiency comparison and the time required to fully peel a zucchini by different methods.

Fig. 5 (b) shows the time comparison for completing the peeling task. The results indicate that teleoperation took approximately three times longer than ForceCapture, while the time taken by ForceCapture was very close to that of direct human peeling. It is worth noting that teleoperation requires additional training, whereas ForceCapture requires minimal training, with users becoming proficient after just one attempt. Furthermore, during teleoperation, there were three interruptions due to operational errors that caused workspace disruptions, none of which occurred with ForceCapture. ForceCapture demonstrates a more natural and streamlined data collection process, without requiring extensive user training or robotic involvement, contrasting with the more structured teleoperation setup.

B. Manipulation Performance: Zucchini Peeling

a) Setup: To evaluate the effectiveness of ForceMimic, we formulated the peeling action as an end-to-end skill learning task. The data collection scene is exemplified in Fig. 1 (a), utilizing the fixed-tool version of ForceCapture. The user held the zucchini steady with the left gripper and peeled with the right ForceCapture. The robot experiment setup is illustrated in the top of Fig. 1 (c), where the L515 RGB-D camera is mounted externally to the robotic arm. The L515 camera was positioned consistently during both data collection and the robot experiment, though it can be positioned flexibly for portable in-the-wild data collection like DexCap [6]. The left robot, equipped with a gripper, was used for rule-based stabilization of the zucchini, while the right arm's fixed peeler identical to the one used in ForceCapture, performed the peeling skill via HybridIL. The robotic arm used in the experiments is the Flexiv Rizon 4, which features precise force sensing and force control capabilities.

We processed 15 zucchinis, collecting 438 peeling skill segments, resulting in a total of 30,199 action sequences. The actions advanced by 3 time steps relative to the perception data. Both the HybridIL model and the baseline methods were trained for 500 epochs each.

b) Methods: In addition to HybridIL, we compared three other baseline methods. **Raw DP** used raw visual perception and robot pose as inputs, outputting the end-effector pose sequence based on diffusion policy. **Force DP** incorporated visual perception, robot pose, and robot force sensing as inputs, also outputting the end-effector pose sequence. **Force+Hybrid DP** used visual perception, robot pose, and robot force sensing as inputs, but output both pose and wrench sequences. For baselines that output wrench-position parameters, hybrid force-position control primitives were employed to match and switch between control modes. **Raw DP** and **HybridIL** were tested for 20 peeling actions, while other two models were tested for 10 peeling actions for their poor performance. The robot's initial TCP pose is consistent with the dataset, positioned slightly above and behind the zucchini.

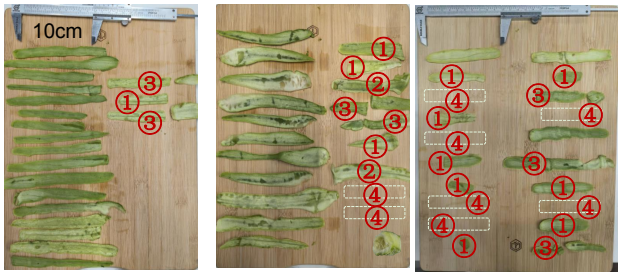
c) Metrics: We defined success using two evaluation criteria. The first criterion is whether the trajectory of the motion is correct, meaning that any length of zucchini peel is successfully removed without damaging the zucchini. The second criterion is whether a continuous peel longer than 10 cm is produced during the peeling process.

d) Results: The results of the four methods are summarized in TABLE I, and the peeled skins are shown in Fig. 6. The Raw DP method achieved a motion success rate of 80% (16/20), with instances of failure detailed in Fig. 6 (b). Failures marked as ② involved excessive force during the peeling process, which resulted in damage to the zucchini. One instance even broke the bottom of the zucchini, as shown in the bottom of Fig. 6 (b). Failure marked as ④ resulted from no contact with the zucchini, hence no peeling occurred. In contrast, HybridIL demonstrated a 100% success rate (20/20), with all attempts resulting in successful contact and peeling, as illustrated in Fig. 6 (a). When the success criterion was increased to a continuous peeling length of more than 10 cm, both Raw DP and HybridIL experienced a decrease in success rates. Raw DP's success rate dropped to 55%, with additional failure cases marked as ① and ③. Case ① indicated peeling lengths shorter than 10 cm, and case ③ involved peeling breakage,

attributed to discontinuity between the output poses, which caused peeling interruptions. For HybridIL, the success rate decreased to 85%, with failures in cases ① and ③. These failures were due to the premature ending of the output force-position parameters, leading to an early switch from the hybrid force control primitives to IK-based joint position control primitives, which caused peeling discontinuities.

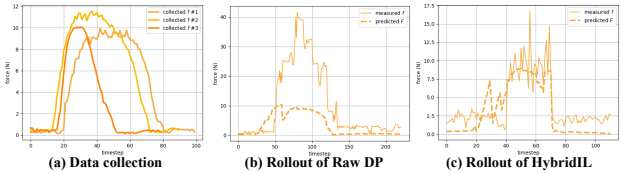
TABLE I: Quantitative results of zucchini peeling.

Methods	**Success rate (%)**	
	motion correct	peel length $>$ 10cm
Raw DP	80	55
Force DP	60	10
Force+Hybrid DP	80	20
HybridIL (proposed)	**100**	**85**



(a) HybridIL (b) Raw DP (c) Force DP vs Force+Hybrid DP

Fig. 6: Visualization of the peeled skins by different methods. Failure cases are numbered with circles.



(a) Data collection (b) Rollout of Raw DP (c) Rollout of HybridIL

Fig. 7: Examples of force curves during peeling a zucchini in different scenarios.

The models that incorporate force as an input, including Force DP and Force+Hybrid DP, performed poorly. While the initial motion leading to the contact with the zucchini was generally correct, once contact occurred, these models struggled to predict the correct pose and force, making successful peeling nearly impossible. This result is counterintuitive—— one would expect that adding force sensing would improve peeling performance, but instead, it worsened the outcome. The reasons for this can be understood from the interactive force curves of the Raw DP and HybridIL during the peeling process, as shown in Fig. 7. Although Raw DP successfully peeled the zucchini, the interactive forces were significantly higher, averaging around 20N and reaching over 40N in some areas. In contrast, the dataset from which these models were trained exhibited much lower interaction forces, around 10N. This mismatch between the input forces and the force distribution in the dataset made it difficult for the models to predict the correct actions. The inconsistency between the force interaction controller used during robot deployment and data collection might be a potential factor contributing to the issue. Addressing this discrepancy could improve the model's performance. Effectively utilizing the force data collected by ForceCapture as sensory input remains an open challenge and a promising direction for future research. Further exploration on how to better integrate force perception into control strategies could lead to significant advancements in improving task performance. Additionally, as shown in Fig. 6 (a) and Fig. 7 (c), the HybridIL method maintained an average interaction force of 9N, closely aligning with the model's predicted forces, resulting in evenly peeled sections with consistent thickness and width. However, a slight deviation from the force distribution in the dataset was still noticeable, which likely explains why the Force+Hybrid DP failed to yield good results. Nonetheless, Force+Hybrid DP demonstrated an improvement over the Force DP.

V. Conclusion and Discussions

We present ForceMimic, a system aimed at advancing force-centric robot learning. This system includes ForceCapture, a scalable on-site force-position data collection system, and HybridIL, a method based on force-interaction control primitives to fit force-position parameters in imitation learning tasks. The effectiveness of both the system and method is demonstrated in the zucchini peeling task. We hope that ForceMimic will pave the way for future research on force-centric perception and hybrid force-position decision-making models in imitation learning.

We provide an initial exploration of learning human force-position skills based on ForceCapture. However, there is still room for improvement. 1) Our model uses a simple MLP to represent point clouds, robot pose, and force. In the future, we could explore more advanced multimodal representations that combine visual, force, and robot state data to improve the model's generalization to diverse skills. 2) HybridIL only employs two control primitives to fit the model's predicted force-position parameters. Future research could involve exploring more control primitives to better align with model outputs, and the model itself could potentially predict the most suitable primitive and corresponding parameters in advance. 3) ForceMimic has so far demonstrated success with a single peeling skill. In the future, the system could be extended to more force-oriented tasks.

Acknowledgements

This work was supported by the Shanghai Commitee of Science and Technology (No. 24511103200), the National Key Research and Development Project of China (No. 2022ZD0160102), XPLORER PRIZE grants of Shanghai Artificial Intelligence Laboratory, and the National Natural Science Foundation of China (No. 52305030). We would like to thank Flexiv for the hardware of F/T sensor. We are deeply grateful to Shuhan Li, Chen Wang, Wenbo Tang, Jin Liu, Hongjie Fang, Qiaojun Yu, and Qi Wu for their invaluable insights and constructive discussions throughout this research endeavor.

REFERENCES

[1] J. R. Flanagan and A. M. Wing, "The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads," *Journal of Neuroscience*, vol. 17, no. 4, pp. 1519–1528, 1997.

[2] R. S. Johansson and G. Westling, "Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects," *Experimental brain research*, vol. 56, pp. 550–564, 1984.

[3] Y. Zheng, X. Yang, B. Mo, Y. Qi, Y. Yang, C. Lin, S. Han, N. Wang, C. Guang, and W. Liu, "Evaluation of the hand motion and peeling force in inner limiting membrane peeling," *Translational Vision Science & Technology*, vol. 12, no. 3, pp. 32–32, 2023.

[4] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay, *et al.*, "Roboturk: A crowdsourcing platform for robotic skill learning through imitation," in *Conference on Robot Learning*. PMLR, 2018, pp. 879–893.

[5] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[6] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, "DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation," in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[7] H.-S. Fang, H. Fang, Z. Tang, J. Liu, C. Wang, J. Wang, H. Zhu, and C. Lu, "Rh20t: A comprehensive robotic dataset for learning diverse skills in one-shot," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 653–660.

[8] A. Toedtheide, X. Chen, H. Sadeghian, A. Naceri, and S. Haddadin, "A force-sensitive exoskeleton for teleoperation: An application in elderly care robotics," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 624–12 630.

[9] H. Fang, H.-S. Fang, Y. Wang, J. Ren, J. Chen, R. Zhang, W. Wang, and C. Lu, "Airexo: Low-cost exoskeletons for learning whole-arm manipulation in the wild," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 15 031–15 038.

[10] S. Yang, M. Liu, Y. Qin, R. Ding, J. Li, X. Cheng, R. Yang, S. Yi, and X. Wang, "Ace: A cross-platform and visual-exoskeletons system for low-cost dexterous teleoperation," in *8th Annual Conference on Robot Learning*, 2024.

[11] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel, "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 5628–5635.

[12] A. Iyer, Z. Peng, Y. Dai, I. Guzey, S. Haldar, S. Chintala, and L. Pinto, "Open teach: A versatile teleoperation system for robotic manipulation," in *CoRL Workshop on Learning Robot Fine and Dexterous Manipulation: Perception and Control*, 2024.

[13] R. Ding, Y. Qin, J. Zhu, C. Jia, S. Yang, R. Yang, X. Qi, and X. Wang, "Bunny-visionpro: Real-time bimanual dexterous teleoperation for imitation learning," *arXiv preprint arXiv:2407.03162*, 2024.

[14] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang, "Open-television: Teleoperation with immersive active visual feedback," in *8th Annual Conference on Robot Learning*, 2024.

[15] H. Kim, Y. Ohmura, A. Nagakubo, and Y. Kuniyoshi, "Training robots without robots: deep imitation learning for master-to-robot policy transfer," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2906–2913, 2023.

[16] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, "Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 12 156–12 163.

[17] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[18] J. Aldaco, T. Armstrong, R. Baruch, J. Bingham, S. Chan, K. Draper, D. Dwibedi, C. Finn, P. Florence, S. Goodrich, *et al.*, "Aloha 2: An enhanced low-cost hardware for bimanual teleoperation," *arXiv preprint arXiv:2405.02292*, 2024.

[19] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," in *8th Annual Conference on Robot Learning*, 2024.

[20] S. Song, A. Zeng, J. Lee, and T. Funkhouser, "Grasping in the wild: Learning 6dof closed-loop grasping from low-cost demonstrations," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4978–4985, 2020.

[21] S. Young, D. Gandhi, S. Tulsiani, A. Gupta, P. Abbeel, and L. Pinto, "Visual imitation made easy," in *Conference on Robot Learning*. PMLR, 2021, pp. 1992–2005.

[22] H. Ha, Y. Gao, Z. Fu, J. Tan, and S. Song, "UMI on legs: Making manipulation policies mobile with manipulation-centric whole-body controllers," in *8th Annual Conference on Robot Learning*, 2024.

[23] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," *Advances in neural information processing systems*, vol. 1, 1988.

[24] N. M. Shafiullah, Z. Cui, A. A. Altanzaya, and L. Pinto, "Behavior transformers: Cloning k modes with one stone," *Advances in neural information processing systems*, vol. 35, pp. 22 955–22 968, 2022.

[25] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and

J. Tompson, "Implicit behavioral cloning," in *Conference on Robot Learning*. PMLR, 2022, pp. 158–168.
[26] J. Pari, N. M. Shafiullah, S. P. Arunachalam, and L. Pinto, "The surprising effectiveness of representation learning for visual imitation," in *Proceedings of Robotics: Science and Systems (RSS)*, 2022.
[27] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion Policy: Visuomotor Policy Learning via Action Diffusion," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.
[28] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, "3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations," in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.
[29] C. Wang, H. Fang, H.-S. Fang, and C. Lu, "Rise: 3d perception makes real-world robot imitation simple and effective," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 2870–2877.
[30] R. S. Johansson, J. R. Flanagan, and R. S. Johansson, "Sensory control of object manipulation," *Sensorimotor control of grasping: Physiology and pathophysiology*, pp. 141–160, 2009.
[31] M. Edmonds, F. Gao, X. Xie, H. Liu, S. Qi, Y. Zhu, B. Rothrock, and S.-C. Zhu, "Feeling the force: Integrating force and pose for fluent discovery through imitation learning to open medicine bottles," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 3530–3537.
[32] F. Li, Q. Jiang, W. Quan, S. Cai, R. Song, and Y. Li, "Manipulation skill acquisition for robotic assembly based on multi-modal information description," *IEEE Access*, vol. 8, pp. 6282–6294, 2019.
[33] N. Fazeli, M. Oller, J. Wu, Z. Wu, J. B. Tenenbaum, and A. Rodriguez, "See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion," *Science Robotics*, vol. 4, no. 26, p. eaav3123, 2019.
[34] T. Yang, Y. Jing, H. Wu, J. Xu, K. Sima, G. Chen, Q. Sima, and T. Kong, "Moma-force: Visual-force imitation for real-world mobile manipulation," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 6847–6852.
[35] N. Hogan, "Stable execution of contact tasks using impedance control," in *Proceedings. 1987 IEEE International Conference on Robotics and Automation*, vol. 4. IEEE, 1987, pp. 1047–1054.
[36] C. Ott, R. Mukherjee, and Y. Nakamura, "Unified impedance and admittance control," in *2010 IEEE international conference on robotics and automation*. IEEE, 2010, pp. 554–561.
[37] J. A. Collins, C. Houff, Y. L. Tan, and C. C. Kemp, "Forcesight: Text-guided mobile manipulation with visual-force goals," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 10 874–10 880.
[38] C. Dong, L. Yu, M. Takizawa, S. Kudoh, and T. Suehiro, "Food peeling method for dual-arm cooking robot," in *2021 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2021, pp. 801–806.
[39] R. Ye, Y. Hu, Y. A. Bian, L. Kulm, and T. Bhattacharjee, "Morpheus: a multimodal one-armed robot-assisted peeling system with human users in-the-loop," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 9540–9547.
[40] A. Straižys, M. Burke, and S. Ramamoorthy, "Surfing on an uncertain edge: Precision cutting of soft tissue using torque-based medium classification," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4623–4629.
[41] H. Kim, Y. Ohmura, and Y. Kuniyoshi, "Goal-conditioned dual-action imitation learning for dexterous dual-arm robot manipulation," *IEEE Transactions on Robotics*, 2024.
[42] J. Carpentier, J. Mirabel, N. Mansard, and the Pinocchio Development Team, "Pinocchio: fast forward and inverse dynamics for poly-articulated systems," https://stack-of-tasks.github.io/pinocchio, 2015–2018.