A Contrastive Learning-based Planning and Control Framework for Symbolic Manipulation of Deformable Linear Objects

Shengzeng Huo[1], Jihong Zhu[2], Hesheng Wang[3], David Navarro-Alarcon[1]

***Abstract*— Most robotic deformable object manipulation strategies are based on the assumption that the environment is structured (i.e., pre-grasping without any obstacles) and the goal's details have been fully specified (e.g., the exact target shape). However, there are many tasks that involve spatial relations in human environments where these conditions may be hard to satisfy, e.g., bending and placing a cable inside an unknown container. To develop advanced robotic manipulation capabilities that avoid these assumptions, we propose a contrastive learning-based planning and control framework. Using simulation data collected by random actions, we learn an embedding model in a contrastive manner that encodes the spatio-temporal information from successful experiences, which facilitates the subgoal planning through clustering in the latent space. Based on the keypoint correspondence-based action parameterization, we design a leader-follower control scheme for the collaboration between dual arms. All models of our policy are automatically trained in simulation and can be directly transferred to real-world environments. To validate the proposed framework, we conduct a detailed experimental study on a complex scenario subject to environmental and reachability constraints in both simulation and real environments.**

I. INTRODUCTION

Deformable object manipulation (DOM) has many promising applications in growing fields, such as flexible cable arrangement [1], [2], clothes manipulation [3], [4], robot-assisted dressing [5], [6] and open bags [7]. Compared with rigid objects, manipulating deformable objects is more challenging due to their complex mechanical structure (i.e., variable morphology and the high number of degrees of freedom) [8], [9].

Although great success has been achieved in DOM (e.g. [10]–[16]), most of them assume a structured configuration (pre-grasping without any obstacles) and a fully specified goal (e.g. the exact target shape). However, these assumptions are hard to satisfy in some real-world scenarios. For example, in the case where a robot is commanded to pick a deformable cable from a cluttered environment and arrange it inside a box; The relative spatial relationship "inside the box" represents the desired goal rather than the specific target shape of the cable [17], which we consider it as a symbolic goal. In this paper, we provide a solution to this problem in the context of automatically rearranging a deformable linear object (DLO) in a planar setting to make it satisfy several geometric constraints simultaneously. Specifically, our goal is to enable dual arms to perform prehensile grasping about the corresponding ends of a DLO, which is practical since it can be considered a prerequisite for DOM tasks with fixed contacts [18]. There are several challenges in this setting: (1) Lack of a goal specification; (2) Nonlinear dynamics of the system in unstructured environments; (3) Long-horizon planning complexity; (4) High-dimensional continuous state-action spaces.

[1] The Hong Kong Polytechnic University, KLN, Hong Kong.
[2] The University of York, York, UK.
[3] Shanghai Jiaotong University, Shanghai, China.

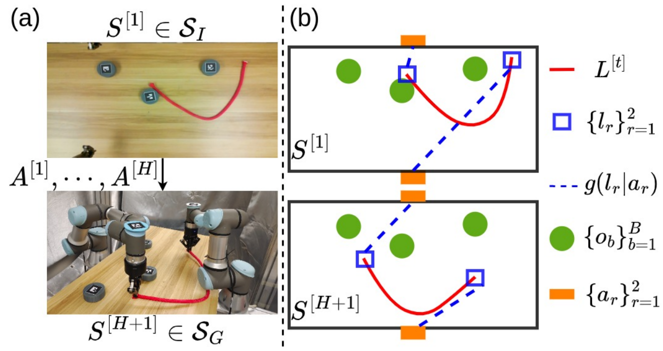

Fig. 1. Schematic diagram of our bimanual manipulation setting. (a) Illustration of the context. The goal of the task is to enable dual arms to grasp the corresponding ends of the DLO. (b) Graphical representation of the problem formulation. The blue dashed lines represent the correspondence between a robotic arm and a end of the DLO.

We present a novel contrastive learning-based planning and control framework to deal with the challenges. As opposed to modeling the complex dynamics of the DLO, our method utilizes spatio-temporal information from previous successful experiences, which enables to transfer of the trained policy from simulation to the real world. Our original contributions are:

- A contrastive learning-based subgoal planner for long-horizon sparse reward tasks without a goal specification.
- A leader-follower control scheme for goal-conditioned collaborative manipulation under geometric constraints.
- A detailed experimental study that evaluates the proposed method in both simulation and real environments.

II. METHODS

A. Problem Formulation

We formulate the problem as a discrete-time episodic Markov Decision Process (MDP) represented by a tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, R, \mathcal{P})$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, R is the reward function, $\mathcal{P}(S^{[t+1]}|S^{[t]}, A^{[t]})$ is the transition function. The objective of this context is to reach the goal space $\mathcal{S}_G$, a subset of the state space $\mathcal{S}_G \subseteq \mathcal{S}$ that satisfies geometric conditions. Thus, robots need to

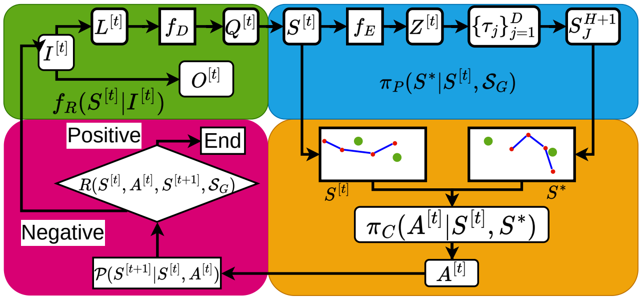

Fig. 2. Overview of the proposed planning and control framework.

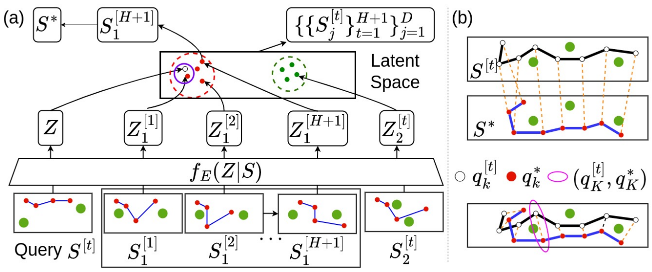

Fig. 3. (a) Conceptual representation of our contrastive subgoal planning model. (b) Visualization of the leader-follower control scheme.

perform correct actions continuously and finally complete the task to obtain the sparse positive reward. A typical example is shown in Fig. 1, in which dual arms $\{a_r\}_{r=1}^2$ can not grasp the corresponding ends $\{l_r\}_{r=1}^2$ of the DLO $L^{[t]}$ initially ($g(l_r|a_r)$ denotes the relationship between the end and the individual arm). Our goal is to change the state of the DLO through H steps of manipulation, reaching the goal space $S^{[H+1]} \in \mathcal{S}_G$. Specifically, it means the geometric conditions (robotic reachability and collision avoidance) are satisfied (a real scenario shown in Fig. 1(a)).

B. Solution

Due to sampling inefficiency in this long-horizon sparse reward task, we factorize the policy $\pi(A^{[t]}|S^{[t]}, \mathcal{S}_G)$ into global subgoal planning $\pi_P(S^*|S^{[t]}, \mathcal{S}_G)$ and local goal-conditioned control $\pi_C(A^{[t]}|S^{[t]}, S^*)$, as shown in Fig. 2.

We describe a DLO with a link-joint structure and designate the joints as representative sequential keypoints [19], denoted as $Q^{[t]} = \{q_k^{[t]}\}_{k=1}^M$. In addition, the coordinates of the obstacles are also included in the state representations $S^{[t]}$. The action sequence of an arm $\mathcal{T}_r^{[t]} = \{P_{pick}, P_{place}\}$ is defined as correspondence-based manipulation from the present state $S^{[t]}$ to the intended state S^*. The picking and placement locations are specifically chosen inside the keypoints of the present state $P_{pick} \leftarrow q_k^{[t]} \in Q^{[t]}$ and the intended state $P_{place} \leftarrow q_k^* \in Q^*$, respectively.

The dataset for training the data-driven model is collected in simulation. Since our pick-and-place sequence $\mathcal{T}_r^{[t]} = (P_{pick}, P_{place})$ is determined based on the current state $S^{[t]}$ and a desired goal S^*, we firstly record G states within the goal space $\mathcal{S}_G$ to form a dataset $\{S_g^* \in \mathcal{S}_G\}_{g=1}^G$ through manipulating the DLO with arbitrary actions and then implement the correspondence-based action randomly. The procedures include choosing a goal S_g^* within the dataset $\{S_g^*\}_{g=1}^G$ and sample feasible actions from the parameterization space to execute. Finally, we obtain a dataset $\mathcal{D}$ automatically with D successful episodes $\mathcal{D} = \{\tau_j\}_{j=1}^D$, where an episode is $\tau_j = \{[S_j^{[1]}, A_j^{[1]}, S_j^{[2]}, \cdots, A_j^{[H]}, S_j^{[H+1]}]\}$.

The aim of the subgoal planner $\pi_P(S^*|S^{[t]}, \mathcal{S}_G)$ is to point out a promising direction toward the goal space $\mathcal{S}_G$ for the query state $S^{[t]}$. We consider the subgoal planning problem as searching for a suitable state from previous exploration $S \in \mathcal{D}$. The motivation of our search-based subgoal planner is the ultimate accomplished state $S_j^{[H+1]} \in \mathcal{S}_G$ of a successful episode τ_j is a desirable and feasible goal for the states within this episode $\{S_j^{[t]}\}_{t=1}^H \in \tau_j$. The concept of the training and prediction of the contrastive learning-based subgoal planner is illustrated in Fig. 3(a). For a state in the dataset $S_j^{[t]} \in \{\{S_j^{[t]}\}_{t=1}^{H+1}\}_{j=1}^D$, its positive samples are other states belong to the same episode $\{\{S_j^{[t]}\}_{t=1}^{H+1} \setminus S_j^{[t]}\} \in \tau_j$ while its negative samples are other states belong to different episodes in the dataset $\mathcal{D} \setminus \{S_j^{[t]}\}_{t=1}^{H+1}$. With these pairs, we leverage InfoNCE loss [20] to train the encoder $f_E(Z^{[t]}|S^{[t]})$. Within the embedding space, this results in the states belong to the same episode being placed together but the negative samples pushed further apart.

The local controller $\pi_C(A^{[t]}|S^{[t]}, S^*)$ is responsible for refining the configuration of the DLO $S^{[t]}$ based on the subgoal S^* supplied by the planner model $\pi_P(S^*|S^{[t]}, \mathcal{S}_G)$. We decouple the roles of twin arms as a leader and a follower [21]. Both the leader and the follower determine the pick-and-place sequence according to the correspondence of keypoints between $q_k^{[t]} \in S^{[t]}$ and $q_k^* \in S^*$, as shown in Fig. 3(b). Without fixed contacts, dual arms adjust their picking points at each time step and their individual roles can be switched. Both pick-and-place sequences of them are acquired through optimization subject to constraints.

The complete policy implementation process incorporates global subgoal planning and local goal-conditioned control and iterates until the task is completed.

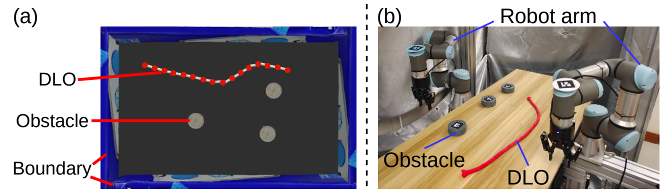

Fig. 4. (a) A snapshot of the simulation in Pybullet. (b) Experimental setup on a planar workspace.

III. Results

The simulations in Pybullet [22] for collecting data and validating the algorithm are visualized in Fig. 4(a), in which a DLO and multiple cans serving as obstacles are included in the workspace. The manipulation with our leader-follower control scheme is rendered as a virtual force, which is

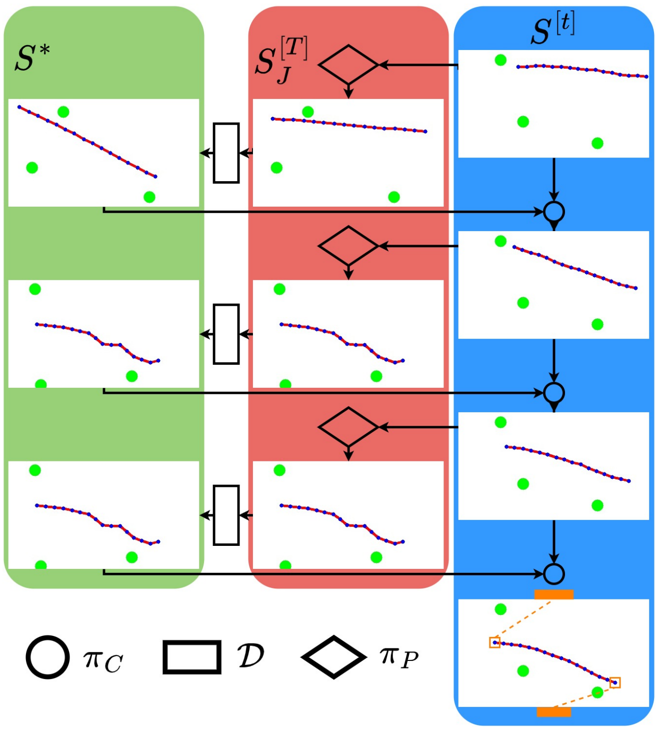

Fig. 5. A complete episode in simulation. The subgoal planner finds the most similar state in the dataset (red column) concerning the current state (blue column) and assigns the corresponding achieved goal of the corresponding episode as a subgoal (green column). Then, the controller determines the action conditioning on the current state and the subgoal.

parameterized by its position and magnitude and maintained until the intended displacement distance is reached.

A complete episode of our constrained bimanual manipulation in simulation is shown in Fig. 5. We retrieve the embedding $Z^{[t]}$ of the state $S^{[t]}$ with the encoder $f_E(Z^{[t]}|S^{[t]})$ and then locate the most comparable embedding $S_J^{[T]}$ in the dataset $\mathcal{D}$. Then, we assign the achieved goal $S_J^{[H+1]}$ of the $J-th$ episode in the dataset as a subgoal $S_J^{[H+1]} \rightarrow S^*$. At last, the local goal-conditioned controller $\pi_C(A^{[t]}|S^{[t]}, S^*)$ takes the current state $S^{[t]}$ and the planned subgoal S^* as input and output the correspondence-based action $A^{[t]}$. The entire planning and control framework iterates until the attached state belong to the goal space $S^{[t+1]} \in \mathcal{S}_G$.

Next, we show how well our suggested framework works to transfer from simulation to reality without any fine-tuning. The direct transfer is feasible since we do not require the accurate dynamic consistency between them. We contend that it is advantageous to interleave planning and control for these complicated manipulation tasks. Fig. 4(b) shows our physical robotic environments. Two ur3 manipulators equipped with 2-fingered Robotiq grippers are used for this constrained bimanual manipulation task. The obstacles in the environment are localized with markers and fixed during an episode. An Intel Realsense L515 camera is attached to sense the top-down perspective of the environment $I^{[t]}$. To analyze our framework in detail, we provide two typical examples in the trials, visualized in Fig. 6. Fig. 6(a) shows an

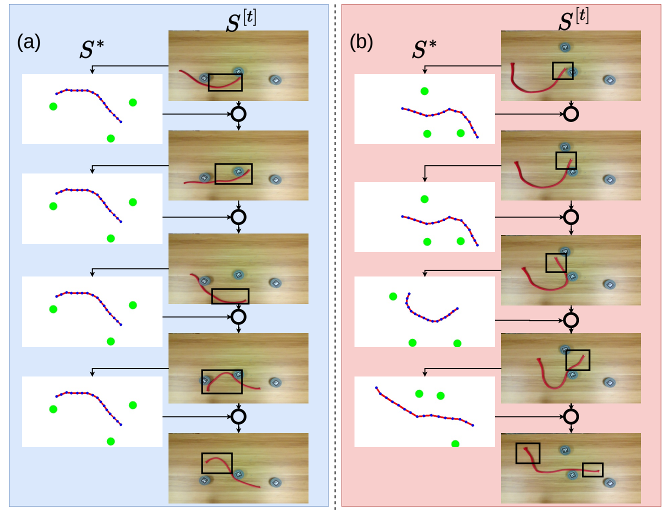

Fig. 6. Pictures of two typical examples in physical robot demonstrations. The black box highlights the manipulated region during action execution.

episode with a constant subgoal that is presented throughout the whole episode. The local goal-conditioned controller initially arranges the DLO to the center of the workspace, allowing dual arms to engage in the subsequent manipulation. The DLO is then adjusted with dual arms, namely rotating it around the obstacle. In order to bypass environmental restrictions, robots finally shift the DLO further from the obstruction. We acknowledge that the state after action implementation and the planned subgoal vary in certain ways. Actually, rather than requesting the controller to explicitly attain a particular state, the planner is used to indicate a promising way to approach the goal space. Owing to the replanning operation, the desired subgoal probably varies throughout the episode, as shown in Fig. 6(b). In the beginning, the controller attempts to maneuver the DLO through the barriers by moving it to the right of the workspace. A new subgoal S^* is included to promote shifting the right end of the DLO to the upper right corner as the state of the DLO changes. Then, both arms participate in distributing the DLO horizontally in the workspace based on a new subgoal S^*. This example illustrates that replanning is useful to adjust the approaching direction in the tasks without goal specifications.

IV. Conclusion

In this paper, we propose a novel contrastive learning-based planning and control framework for the manipulation of a DLO under environmental and reachability constraints. Removing the assumption of a structured world and a goal specification, our proposed methodology further enhances the dexterity of robotic manipulation. To deal with the sparse reward settings in long-horizon tasks, our policy model is factorized into global subgoal planning and local goal-conditioned control. All the models are trained in simulation and can be transferred to real environments without any fine-tuning. A detailed experimental study is reported to illustrate the effectiveness of the framework.

References

[1] H. Zhang, J. Ichnowski, D. Seita, J. Wang, H. Huang, and K. Goldberg, "Robots of the lost arc: Self-supervised learning to dynamically manipulate fixed-endpoint cables," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4560–4567, 2021.

[2] V. Lim, H. Huang, L. Y. Chen, J. Wang, J. Ichnowski, D. Seita, M. Laskey, and K. Goldberg, "Real2sim2real: Self-supervised learning of physical single-step dynamic actions for planar robot casting," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 8282–8289, 2022.

[3] S. Tirumala, T. Weng, D. Seita, O. Kroemer, Z. Temel, and D. Held, "Learning to singulate layers of cloth using tactile feedback," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7773–7780, IEEE, 2022.

[4] I. Garcia-Camacho, M. Lippi, M. C. Welle, H. Yin, R. Antonova, A. Varava, J. Borras, C. Torras, A. Marino, G. Alenya, *et al.*, "Benchmarking bimanual cloth manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1111–1118, 2020.

[5] F. Zhang, A. Cully, and Y. Demiris, "Probabilistic real-time user posture tracking for personalized robot-assisted dressing," *IEEE Transactions on Robotics*, vol. 35, pp. 873–888, 2019.

[6] Z. M. Erickson, H. M. Clever, G. Turk, C. K. Liu, and C. C. Kemp, "Deep haptic model predictive control for robot-assisted dressing," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–8, 2017.

[7] L. Y. Chen, B. Shi, D. Seita, R. Cheng, T. Kollar, D. Held, and K. Goldberg, "Autobag: Learning to open plastic bags and insert objects," *arXiv preprint arXiv:2210.17217*, 2022.

[8] H. Yin, A. Varava, and D. Kragic, "Modeling, learning, perception, and control methods for deformable object manipulation," *Science Robotics*, vol. 6, no. 54, 2021.

[9] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, J. Kober, X. Li, J. Pan, W. Yuan, and M. Gienger, "Challenges and outlook in robotic manipulation of deformable objects," *IEEE Robotics and Automation Magazine*, vol. 29, no. 3, pp. 67–77, 2022.

[10] M. Lippi, P. Poklukar, M. C. Welle, A. Varava, H. Yin, A. Marino, and D. Kragic, "Latent space roadmap for visual action planning of deformable and rigid object manipulation," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5619–5626, 2020.

[11] C. Wang, Y. Zhang, X. Zhang, Z. Wu, X. Zhu, S. Jin, T. Tang, and M. Tomizuka, "Offline-online learning of deformation model for cable manipulation with graph neural networks," *IEEE Robotics and Automation Letters*, vol. 7, pp. 5544–5551, 2022.

[12] J. Zhu, B. Navarro, P. Fraisse, A. Crosnier, and A. Cherubini, "Dual-arm robotic manipulation of flexible cables," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 479–484, IEEE, 2018.

[13] S. Chen, Y. Liu, S. W. Yao, J. Li, T. Fan, and J. Pan, "Diffsrl: Learning dynamical state representation for deformable object manipulation with differentiable simulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9533–9540, 2022.

[14] A. Longhini, M. Moletta, A. Reichlin, M. C. Welle, D. Held, Z. M. Erickson, and D. Kragic, "Edo-net: Learning elastic properties of deformable objects from graph dynamics," *ArXiv*, vol. abs/2209.08996, 2022.

[15] R. Lee, M. Hamaya, T. Murooka, Y. Ijiri, and P. Corke, "Sample-efficient learning of deformable linear object manipulation in the real world through self-supervision," *IEEE Robotics and Automation Letters*, vol. 7, pp. 573–580, 2022.

[16] R. Lee, J. Abou-Chakra, F. Zhang, and P. Corke, "Learning fabric manipulation in the real world with human videos," *ArXiv*, vol. abs/2211.02832, 2022.

[17] Y. Zhu, J. Tremblay, S. Birchfield, and Y. Zhu, "Hierarchical planning for long-horizon manipulation with geometric and symbolic scene graphs," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6541–6548, 2020.

[18] P. Mitrano, D. Mcconachie, and D. Berenson, "Learning where to trust unreliable models in an unstructured world for deformable object manipulation," *Science Robotics*, vol. 6, 2021.

[19] S. Huo, A. Duan, C. Li, P. Zhou, W. Ma, H. Wang, and D. Navarro-Alarcon, "Keypoint-based planar bimanual shaping of deformable linear objects under environmental constraints with hierarchical action framework," *IEEE Robotics and Automation Letters*, vol. 7, pp. 5222–5229, 2022.

[20] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *ArXiv*, vol. abs/1807.03748, 2018.

[21] J. Liu, Y. Chen, Z. Dong, S. Wang, S. Calinon, M. Li, and F. Chen, "Robot cooking with stir-fry: Bimanual non-prehensile manipulation of semi-fluid objects," *IEEE Robotics and Automation Letters*, vol. 7, pp. 5159–5166, 2022.

[22] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," 2016.