Make a Donut 🍩 : Language-Guided Hierarchical EMD-Space Planning for Zero-Shot Deformable Object Manipulation

Yang You[1], Bokui Shen[1], Congyue Deng, Haoran Geng[2], He Wang[2], Leonidas Guibas[1]

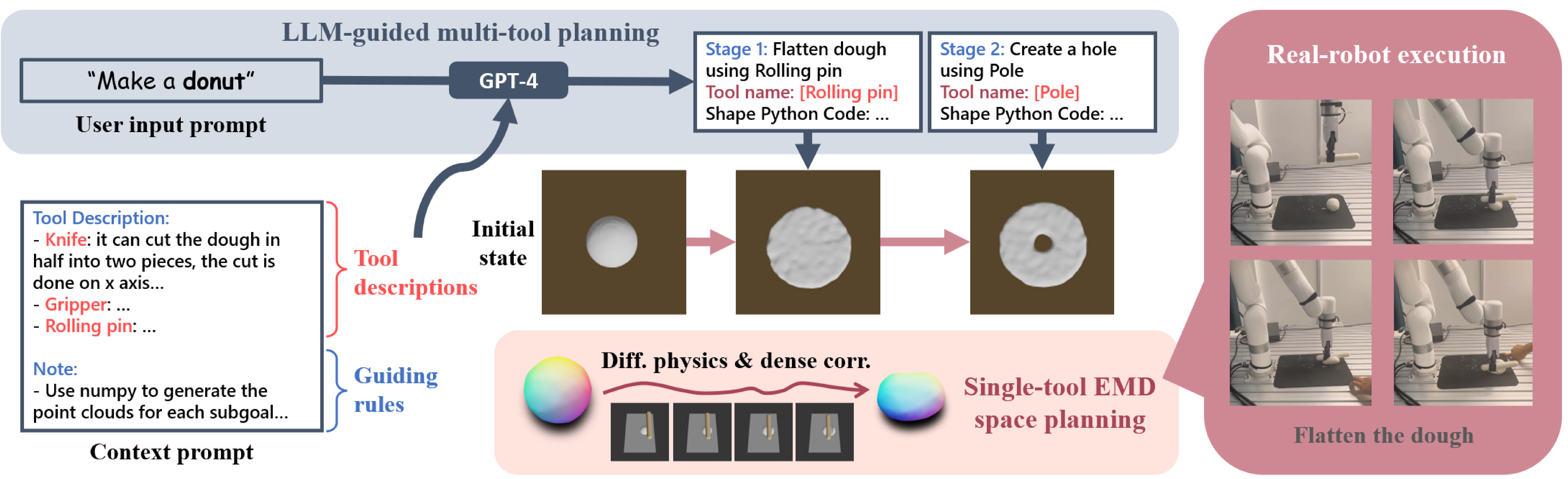
[1]Stanford University [2]Peking University

Problems with Deformable Manipulation

- How to generate good paths, with infinite degrees of freedom
- In general, long-horizon task planning is non-trivial

Check out our project website

Method Overview



Top level: Multi-Tool Selection with LLM Guidance

We ask the LLM to output the following items:

- A one-line explanation of what this step is doing
- Name of the tool to be used.
- Python code to generate target point clouds
- Variable names for the input and output.
- Location of each piece in a dictionary format with a variable name as the key
- Volume of each piece is also in a dictionary format with a variable name as the key

Additional guidelines in the prompt:

- Volume preserving
- Chain of reasoning

volume changes ↓

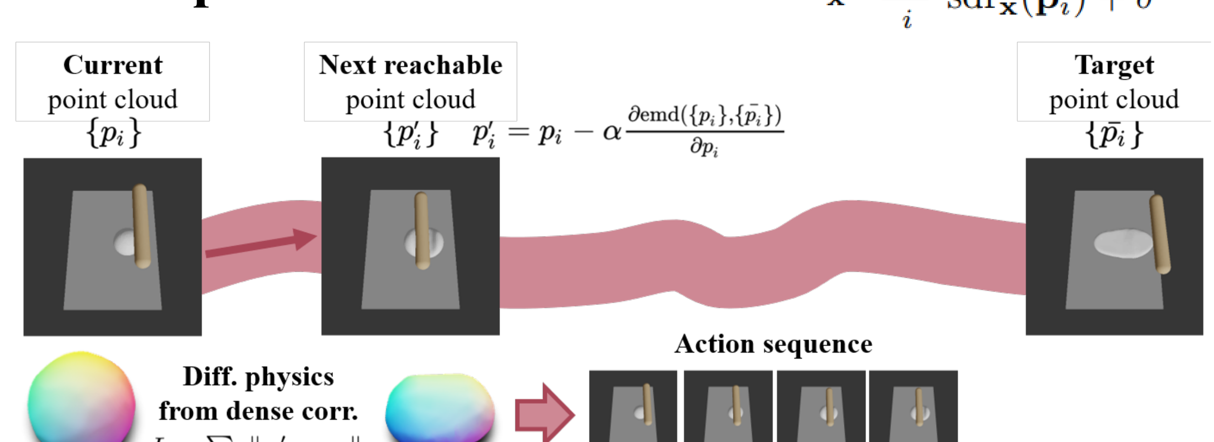
	Donut ↓	Baguette ↓	TwoPancakes ↓
w/o Volume Preserving and Chain Reasoning	73.9%	42.5%	65.0%
Ours	**9.8%**	**38.9%**	**0.0%**

Bottom level: Single-Tool Planning in the EMD Space

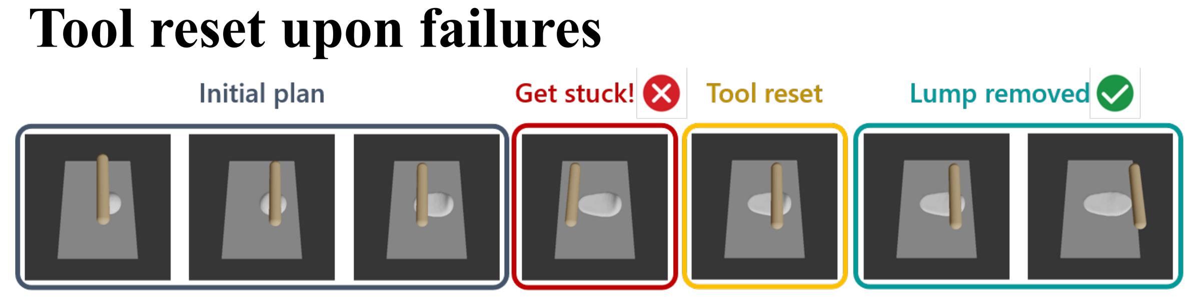
DiffPhysics-P2P: Decide the next subgoal with

$$\mathbf{p}_i' = \mathbf{p}_i - \alpha \cdot \frac{\partial \mathrm{emd}(\{\mathbf{p}_i\}, \{\bar{\mathbf{p}}_i\})}{\partial \mathbf{p}_i}$$

Initial position selection $\mathbf{x}^* = \arg\max_{\mathbf{x}} \sum_i \frac{\|\mathbf{p}_i' - \mathbf{p}_i\|_1}{\mathrm{sdf}_{\mathbf{x}}(\mathbf{p}_i) + \delta}$

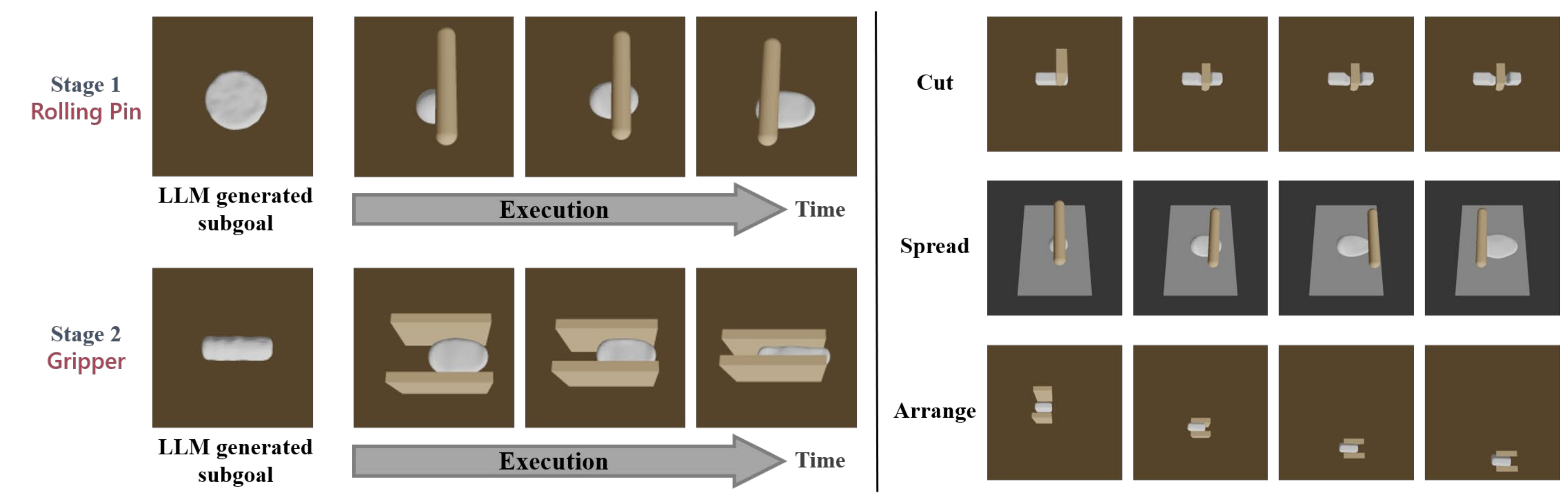


Tool reset upon failures

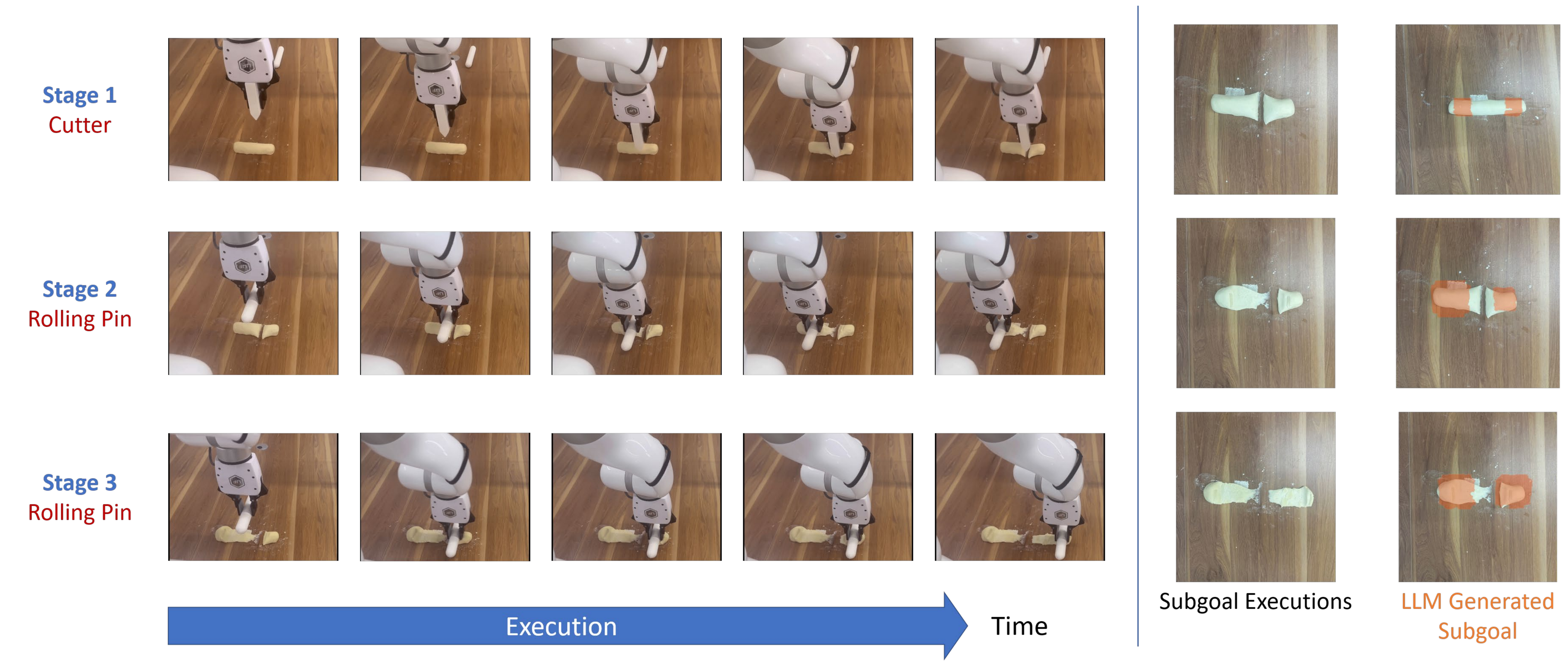


Experimental Results

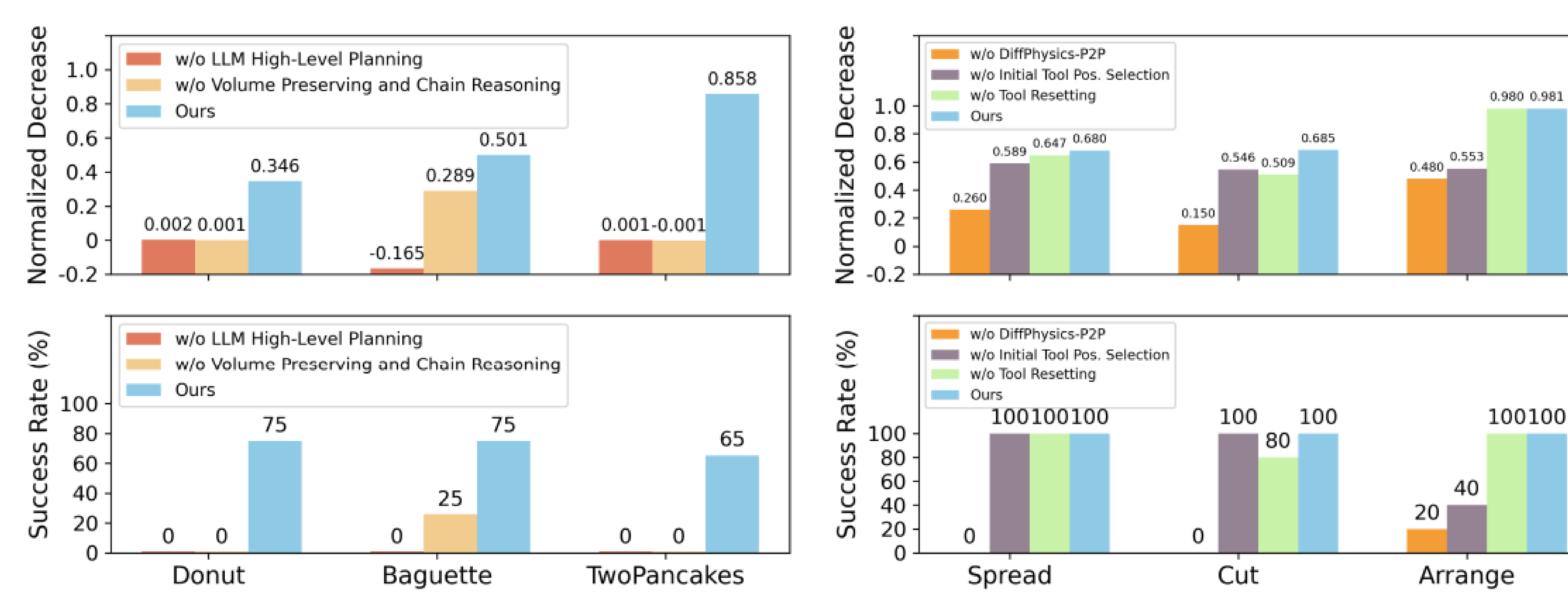
Simulation Results



Real-robot Results



Ablation Studies



(a) Ablation results for multiple-tool experiments. (b) Ablation results for single-tool experiments.