

Learning to Ground Objects for Robot Task and Motion Planning

Yan Ding, Xiaohan Zhang, Xingyue Zhan, and Shiqi Zhang

Abstract—Task and motion planning (TAMP) algorithms have been developed to help robots plan behaviors in discrete and continuous spaces. Robots face complex real-world scenarios, where it is hardly possible to model all objects or their physical properties for robot planning (e.g., in kitchens or shopping centers). In this paper, we define a new object-centric TAMP problem, where the TAMP robot does not know object properties (e.g., size and weight of blocks). We then introduce Task-Motion Object-Centric planning (TMOC), a grounded TAMP algorithm that learns to ground objects and their physical properties with a physics engine. TMOC is particularly useful for those tasks that involve dynamic complex robot-multi-object interactions that can hardly be modeled beforehand. We have demonstrated and evaluated TMOC in simulation and using a real robot. Results show that TMOC outperforms competitive baselines from the literature in cumulative utility.

Index Terms—Task and motion planning, Integrated Planning and Learning, Robot Manipulation, Grounded Planning

I. INTRODUCTION

ROBOTS that operate in the real world need to plan at both task and motion levels. At the task level, the robot computes a sequence of symbolic actions in a discrete space to achieve long-term goals [1]. At the motion level, the symbolic actions are implemented in a continuous space, and the computed trajectories can be directly applied to the real world [2]. Planning at task and motion levels at the same time is challenging [3]–[8], resulting in the so-called integrated *task and motion planning* (TAMP) problem [9], [10], where the main challenge is to achieve task-level goals while maintaining motion-level feasibility.

Planning algorithms (at the task, motion, or both levels) frequently assume that a world model is provided beforehand, including how the world reacts to robot behaviors (i.e., world dynamics). However, many real-world scenarios are very complex, making it hardly possible to model all objects or their physical properties at planning time, e.g., kitchens and shopping centers. In such scenarios, modeling how a robot interacts with multiple objects is even more challenging, e.g., holding a stack of plates, cutting onions, and squeezing through a crowd. Fig. 1 shows a “stack-and-push” task as

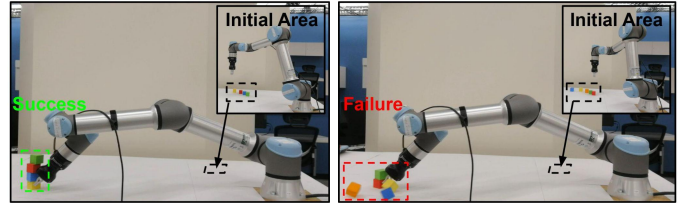


Fig. 1. A robot performing “stack-and-push” tasks by repeatedly building a stack of blocks and pushing the stack to a goal area. There is a trade-off between stability and efficiency (higher stacks are less stable) that the robot must learn from trial and error. It is a TAMP problem, e.g., the stacking behaviors require both task planning and motion planning.

an example scenario with complex multi-object interactions, where a robot repeatedly builds a block tower and then pushes it to a goal area. Aiming to move all blocks to the goal area, the robot needs to learn object properties (e.g., size and weight), how to grasp the blocks (which depends on the object properties), and a stack-and-push strategy (e.g., how many blocks to be stacked together).

The **first contribution** of this work is a new grounded, object-centric TAMP framework, called task and motion Planning with Physics-based Simulation (PPS). The uniqueness of PPS lies in the inclusion of a physics engine for grounding objects and their physical properties. Object grounding enables the robot to collect simulated interaction experiences for policy learning purposes. The PPS framework is general enough to accommodate different TAMP problems that vary in model completeness.

Our **second contribution** is an algorithm, called Task-Motion Object-Centric planning (TMOC), that addresses a challenging PPS problem where *object properties are not provided*. TMOC is particularly useful for those TAMP domains that involve complex multi-object dynamic interactions. Our TMOC robot can ground objects in a high-fidelity physics engine, learn object properties to facilitate the grounding, and improve its task-motion planning skills.

We have demonstrated and evaluated TMOC in simulation (where the robot uses a container to move objects), and using a real robot (where the robot conducts stack-and-push tasks). The simulation and real-world scenarios share the challenge of robot task-motion planning with unknown object properties. From the results, we see that TMOC outperforms a set of baselines from literature in cumulative utility. Finally, we demonstrate the learning process of a real robot that uses TMOC for task-motion planning.

Manuscript received: September 9, 2021; Revised December 29, 2021; Accepted January 30, 2022.

This paper was recommended for publication by Editor Hanna Kurniawati upon evaluation of the Associate Editor and Reviewers’ comments. This work has taken place at the Autonomous Intelligent Robotics (AIR) Group, SUNY Binghamton. AIR research is supported in part by grants from the NSF (NRI-1925044), Ford, OPPO, and SUNY RF.

The authors are with the Department of Computer Science, SUNY Binghamton, Binghamton NY 13902 yding25@binghamton.edu

Digital Object Identifier (DOI): see top of this page.

II. RELATED WORK

There is a long history of developing planning algorithms in robotics research. We summarize three research areas that are the most relevant to this research, namely TAMP, symbol learning for robot planning, and sim-to-real transfer.

A. Task and Motion Planning (TAMP)

Broadly, any robots that plan behaviors at a high level and operate in the real world would need algorithms for TAMP. However, it is not until recently that TAMP has been used as a term to refer to the algorithms that interleave the processes of task planning and motion planning [8]–[10]. Early examples include aSyMov [11] and Semantic Attachment [12]. Hierarchical planning in the now (HPN) [13] has been extended to model the uncertainty in action outcomes and observability [3]. FFRob directly conducts task planning over a set of samples generated in the configuration space [14]. Probabilistically complete TAMP was achieved using constraint satisfaction methods [6]. Wang et al. used a policy synthesis approach to account for uncontrollable agents (such as humans) [15]. Off-the-shelf task-motion planners can be integrated using a planner-independent interface [16]. Optimization methods have been applied to TAMP domains, where the goal is specified with a cost function [5]. In comparison to planning methods (including TAMP) that assume knowledge of complete world models, this work considers robots that compute world models (e.g., objects’ physical properties) through perception, and actively estimate the current world state, which renders “grounding” necessary.

Very recently, researchers developed a visually grounded TAMP approach, called GROF, for mobile manipulation [17]. Compared with GROF that uses computer vision techniques to learn a state mapping function, TMOC (ours) learns to ground objects and their physical properties for TAMP tasks.

B. Symbol Learning for Robot Planning

Researchers have developed symbol learning methods for robot planning. Konidaris et al. (2018) focused on learning symbols to construct representations that are provably capable of evaluating plans composed of sequences of those actions [18], assuming the robot was equipped with a collection of high-level actions. Their mobile manipulator learned a grounded symbol that indicates that a cupboard door is open, which is necessary for determining whether or not the robot can pick up a bottle from the cupboard. Gopalan et al. (2020) introduced natural language into the loop, and showed that the learned symbols can enable a robot to learn to map natural language to goal-based planning with only trajectories as supervision [19]. The above-mentioned methods assumed that low-level motion primitives are provided (as skills or demonstrations), and the robot needs to learn a symbolic representation that can be used for planning to accomplish different complex tasks. In this work, we assume those symbols are provided, but there are no models about the symbols’ physical meanings in the real world.

C. Sim-to-Real Transfer

This work requires a high-fidelity simulator, which is related to “sim-to-real” methods to enable agents to learn from simulation for operating in the real world [20]. Focusing on addressing the reality gap, there are at least two families of sim-to-real methods. One intentionally adds different forms of noise into the simulation environment to learn policies that are robust enough to work in the real world [21]–[25]; the other actively updates parameters of the simulator toward generating realistic experience for policy learning purposes [26]–[29]. This work falls into the second category of methods from the perspective of updating the simulator’s parameters using real-world data. The main difference is that our method is placed in the TAMP context. As a result, our proposed method faces two simultaneous, interdependent reality gaps at task and motion levels, respectively.

This paper enables robots to plan at both task and motion levels in object-centric domains where the object properties can be unknown. Our developed approach is particularly suitable for task-motion domains that involve complex multi-object dynamic interactions. Next, we define the grounded TAMP problem and then describe our algorithm.

III. FRAMEWORK AND PROBLEM STATEMENTS

In this section, we first define an object-centric TAMP framework, and then describe how the framework accommodates a few TAMP problems.

A. Framework

Planning with Physics-based Simulation (PPS) is a task-motion planning framework¹, and is defined as tuple

$$\langle Obj, D^t, D^m, Y \rangle,$$

where Obj is a finite set of objects, and the other components are described next.

Task-level Domain Description $D^t : \langle E, A, T \rangle$, where E is a finite set of symbolic (discrete) properties, and $E^{obj} \subseteq E$ includes symbolic properties that are applicable to $obj \in Obj$. Each property $e \in E^{obj}$ has a finite domain of values, referred to as V_e^{obj} . For instance, in a manipulation domain, object $box \in Obj$ has a property of $loaded \in E^{box}$ that has domain $V_{loaded}^{box} = \{true, false\}$. A state s is an assignment of each object’s applicable symbolic properties. S denotes the set of possible states in symbolic forms. We use $s^{init} \in S$ and $S^G \subseteq S$ to represent the initial state and the set of goal states, respectively.

A is a finite set of object-centric actions, and action $a \in A$ is defined by its preconditions and effects. State transition model $T(s, a, s')$ defines how an action leads (deterministic or probabilistic) transitions. When action costs are considered in task planning, a task planner is able to select a plan out of all the plans that satisfy the goal conditions toward minimizing

¹Researchers have developed TAMP frameworks, such as PETLON for navigation domains [8], and FFRob [14] for heuristics-based TAMP. The uniqueness of PPS, as an object-centric TAMP framework, is attributed to its inclusions of a state mapping function, and a high-fidelity simulator.

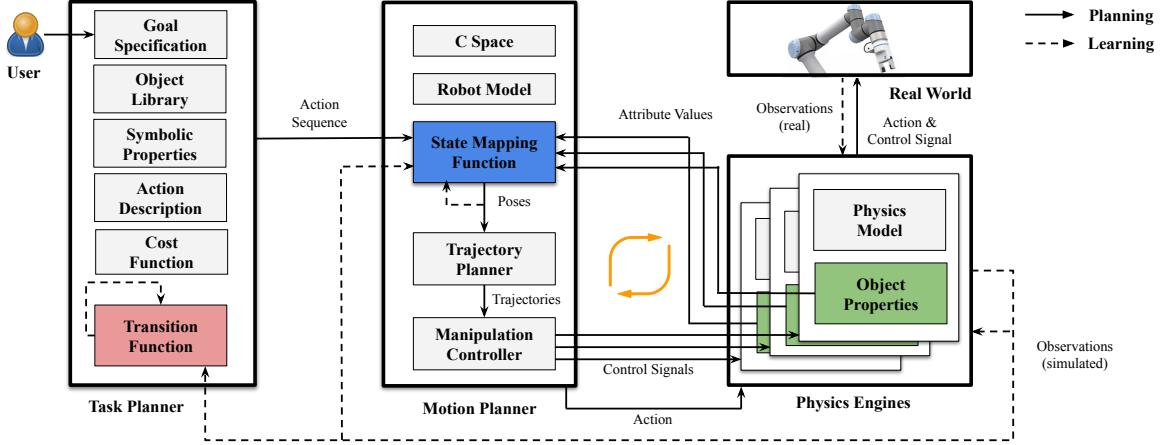


Fig. 2. An overview of the TMOC algorithm for our PPS problem with unknown object properties (L in green), state mapping function (Y in blue), and transition function (T in red). The main components of TMOC include a task planner for sequencing symbolic actions, a motion planner for computing motion trajectories, and a physics engine for simulating multiple grounded worlds. TMOC takes object library, task domain description, goal specification, cost function, motion domain description and state mapping function as input. TMOC aims to compute task-motion plans to achieve task-level goals while maximizing cumulative utilities. Using TMOC, a robot learns object properties (L) from task-completion experience, and leverages the learned properties to further improve its task-motion planning skills (Y and T) over time.

the overall cost. In this paper, at the task level, we are concerned with object-centric, cost-sensitive, goal-conditioned planning under uncertainty.

Motion-level Domain Description $D^m : \langle L, C \rangle$, where L is a finite set of continuous properties. L is the counterpart of E at the task level. $L^{obj} \subseteq L$ includes properties that are applicable to $obj \in Obj$, and properties $l \in L^{obj}$ is a real number. Example continuous properties include an object's size, position, and weight. The configuration space C is the set of all possible configurations. We use ξ to represent a motion trajectory, where the trajectory connects the robot's current configuration (x^{init}) to any configuration in the target space ($x^{goal} \in C$).

State Mapping Function Y maps a task-level state to a motion-level pose: $x \leftarrow Y(s)$, where s is a task-level state and $x \in C$ is a configuration. Y^{-1} is the inverse function of Y , and outputs a symbolic state s given a motion-level configuration. Here we assume Y^{-1} can be derived from Y , and is thus omitted from the definition of PPS.

PPS is a general-purpose TAMP framework that is able to accommodate different task-motion planning problems. A realization of a PPS framework requires a task planning system, a motion planning system, and a physics-based simulation (physics engine) system. One can leverage the physics engine of PPS to ground different components of the real world.

Next, we define a challenging PPS problem, and discuss how this problem connects to a few existing TAMP problems.

B. Problem Statements

In this paper, we aim to address the PPS problem with the following set of functions being unknown:

$$\{L, Y, T\}$$

Consider a “**stack-and-push**” domain: A robot does not know block size and weight (about L), where to place the

gripper for grasping (about Y), and how stable a stack of N blocks is (about T). This is a challenging problem, because planning at task and motion levels highly depends on the estimation of object properties, and the task-level strategy further depends on the robot's motion-level performance. We use the stack-and-push domain in real-robot experiments, where the robot aims to move blocks from an initial area to a goal area as quickly and stably as possible.

The PPS framework is general enough to accommodate different TAMP problems (in addition to the above-mentioned problem that is the focus of this paper). For instance, when all components of PPS are known, it corresponds to a standard TAMP problem, e.g., [8], [30]. When only transition function T is unknown, it corresponds to existing research, such as [6], [31]. Planning domains with unknown physical properties L [32], and planning domains with unknown state mapping function Y [33] can be modeled as PPS problems as well.

The input of a PPS algorithm includes a PPS domain in the form of $\langle Obj, D^t, D^m, Y \rangle$, and a set of task-level goal states S^G . Our utility function incorporates action costs, success bonus, and failure penalty. A PPS algorithm aims to compute task-motion plans to achieve task-level goals while maximizing cumulative utilities.

Next, we focus on our PPS problem with unknown object properties L (where the robot is motivated to learn Y and T accordingly), and develop algorithm TMOC to enable robots to learn to plan at task and motion levels.

IV. ALGORITHM

In this section, we present the main contribution of this research, **TMOC** (short for “task-motion object-centric” planning), a grounded task-motion planning algorithm for addressing the PPS problem described in Section III-B, where object properties (L), state mapping function (Y), and transition function (T) are unknown. TMOC includes three main components

of a task planner, a motion planner, and a physics engine, as illustrated in Fig. 2.

Algorithm 1 presents the control loops of TMOC. The input of TMOC includes a PPS domain in the form of $\langle Obj, D^t, D^m, Y \rangle$, and a set of task-level goal states S^G . TMOC learns L , Y , and T in each iteration, where the robot completes a task *once* in the real world and N *times* in simulation. In each iteration, each of L , Y , and T is learned under the current estimation of the other two, while a TMOC agent plans at task and motion levels.

A. TMOC algorithm

A realization of a PPS framework requires task planning system P^t , motion planning system P^m , and physics-based simulation system \mathcal{M} . P^t takes task domain D^t together with the current state s (which can be derived from D^t) as input and generates an action sequence. Specifically, P^m takes the initial configuration x^{init} , goal configuration set X^G , and D^t as input, and generates motion trajectory ξ . \mathcal{M} takes ξ , object set Obj , and their physics-relevant properties L as input, and generates L' denoting the resulting properties of Obj . From L' , one can infer motion planning domain D^m , which can be further used for computing symbolic state s using the function Y^{-1} .

Algorithm 1 TMOC algorithm

Require: P^t , P^m , and \mathcal{M}

Input: $\langle Obj, D^t, D^m, Y \rangle$, S^G and $Cost$

```

1: Initialize  $T$  optimistically
2: Initialize  $N$  simulated worlds and each one has  $L_i$  (particle) with
   a uniform weight  $w_i$ , where  $i = 0, 1, \dots, N-1$ 
3: Initialize an empty experience pool  $Pool$ 
4: while True do ▷ start working on a new trial
5:    $p \leftarrow P^t(D^t, s^{init}, S^G, Cost)$ , where  $D^t$  includes  $T$  ▷ this is
   task planning
6:   for each action  $\langle s, a \rangle$  in  $p$  do
7:     for  $i = 0, 1, \dots, N-1$  do
8:       Compute  $Y_i$  using  $L_i$  for its simulated world
9:       Generate a feasible pose  $x_i$  for  $\langle s, a \rangle$  using  $Y_i$ 
10:    end for
11:    Compute a pose  $x$  for the real world by weighted
       averaging poses  $\{x_0, x_1, \dots, x_{N-1}\}$ 
12:     $\xi \leftarrow P^m(x_{curr}, x, D^m)$  ▷ this is motion planning
13:    Execute  $\xi$  in the real world and then obtain a resulting
       state  $s'$ 
14:     $Pool = Pool \cup \{\langle s, a, s' \rangle\}$ 
15:    for  $i = 0, 1, \dots, N-1$  do
16:       $\xi_i \leftarrow P^m(x_{curr}, x_i, D^m)$ 
17:       $L'_i \leftarrow \mathcal{M}(\xi_i, Obj, L_i)$ , where  $L_i$  is in  $D^m$ 
18:      Infer a resulting state  $s'_i$  using  $L'_i$  and  $Y_i^{-1}$ , and then
        $L_i \leftarrow L'_i$ 
19:       $Pool = Pool \cup \{\langle s, a, s'_i \rangle\}$ 
20:      Compute weight  $w_i$  of  $L_i$  using  $s'$  and  $s'_i$ 
21:    end for
22:    Update  $L_i$  from the discrete distribution given by
        $\{w_0, w_1, \dots, w_{N-1}\}$ 
23:    Update  $T$  using  $Pool$ 
24:  end for
25: end while

```

Structure of Algorithm TMOC: Lines 1~3 are for initializing data structures. Lines 5~24 form a complete iteration.

In each iteration, Line 5 computes a task-level plan, and the functions of Y , L , and T are updated from trial and error in Lines 8, 22, and 23, respectively. Lines 7~14 are for executing an action in the real world; and Lines 15~21 are for executing the same action in simulation. TMOC is a life-long learning algorithm, and does not have termination conditions. Next, we look into individual lines of TMOC.

Line 1 initializes transition function T “optimistically”. A transition function (T) indicates how reliable an action is. Initializing T optimistically means that the robot believes it always gets the desired results after performing an action, e.g., manipulation and push actions are always successful. This initialization strategy is inspired by the R-max algorithm [34], and encourages exploration in the early learning phase. TMOC initializes N simulated worlds, where each is specified by a set of physics-relevant properties L_i , e.g., length, width, and density of blocks (Line 2). Each simulated world is associated with a weight w_i , and the weights are uniformly initialized. An empty experience pool $Pool$ is initialized to store observations (i.e., actions and their resulting states) after action executions (Line 3).

Line 5 computes an optimal action sequence p using the task planning system P^t given the utility function, which takes into account transition function T (e.g., how reliable grasps are) and cost function $Cost$ (e.g., how long it takes a robot to grasp an object).

Lines 7~14 describe how action $\langle s, a \rangle$ is implemented in the **real world**. From L_i , we can get the configuration space, based on which Y_i can be computed using motion planning methods (Line 8). Thus, a feasible pose x_i for action $\langle s, a \rangle$ can be generated using Y_i (Line 9). A new pose x for the real world is computed by weighted averaging poses x_i , where $i = 0, 1, \dots, N-1$ (Line 11). A trajectory ξ is computed using the motion planner P^m , which takes x_{curr} , computed x , and D^m as input, where x_{curr} refers to the current pose of the robot (Line 12). ξ is executed in the real world, and thus a resulting state s' of action $\langle s, a \rangle$ is obtained (Line 13). Finally, tuple $\langle s, a, s' \rangle$ is added to experience pool $Pool$ (Line 14).

Lines 15~21 explain how action $\langle s, a \rangle$ is implemented in all **simulated worlds** (one iteration for each world). A trajectory ξ_i is computed using motion planning system P^m (Line 16). Then, a set of resulting properties L'_i is obtained using physical-based simulation system \mathcal{M} (Line 17). Thus, a resulting state s'_i can be inferred using L'_i and the reverse function of Y_i , and L_i is also updated by L'_i (Line 18). Finally, tuple $\langle s, a, s'_i \rangle$ is added to $Pool$ (Line 19).

Lines 20, 22 ~ 23 update L and T using action completion experience. Specifically, each weight w_i of L_i can be computed based on the resulting state from the real and simulated worlds (Line 20). Similar to the particle filter [35], a new L_i is sampled from the discrete distribution given by w_i , where $i = 0, 1, \dots, N-1$ (Line 22). Besides, T is updated using the experience pool $Pool$ (Line 23).

TMOC enables robots to plan at both task and motion levels in object-centric domains where the objects’ physical properties (L) are unknown (and thus T and Y are unknown). TMOC is particularly useful for those tasks that involve dynamic complex robot-multi-object interactions that can hardly

TABLE I
ACTION KNOWLEDGE IN OUR SYSTEM, ORGANIZED INTO PRECONDITIONS
AND EFFECTS.

Action	Precondition	Effect
$pickup(R, B, L)$	$is_in(B, L)$ $hand_empty(R)$	$is_holding(R, B)$ $\neg hand_empty(R)$
$stack(R, B, C)$	$is_holding(R, B)$	$\neg is_holding(R, B)$ $hand_empty(R)$ $at_container(B, C)$
$push(R, C, L1, L2)$	$hand_empty(R)$ $is_in(C, L1)$	$hand_empty(R)$ $is_in(C, L2)$

be modeled beforehand. Next, we focus on the evaluation of TMOC in both simulation and the real world.

B. Algorithm Instantiation

Task Planner: Our task planner P^t is implemented using Answer Set Programming (ASP), which is a popular declarative language for knowledge representation and reasoning. ASP has been applied to task planning [8], [32], [36], [37]. In our domain, predicate $is_holding(R1, B1)$ is used to specify block $B1$ being in the robot hand $R1$. We model three manipulation actions, including `pickup`, `stack`, and `push`. For instance, action `pickup` is used to help the robot arm pick up the target block from an initial location, where constraints, such as “`stack` is allowed only if a target block is in the robot hand”, have been modeled as well. An example goal specification can be “all blocks are in the container, and containers are at the goal location”. Table I defines three actions by their preconditions and effects, where R , B , C , and L stand for robot, block, container, and location, respectively.

Motion Planner: At the motion level, given a configuration space, a roadmap is firstly built for the robot. A trajectory planner then generates a desired continuous and collision-free trajectory with minimal traveling distance using optimization algorithms, RRT* in our case [38]. The trajectory includes a set of poses, and the trajectory is delivered to the manipulation controller, along with the robot’s current pose.

V. EXPERIMENTS

We have conducted experiments both in simulation and using a real robot. In simulation experiments, we focus on statistically comparing the performances of TMOC and a set of competitive baselines using results collected from large numbers of trials. The comparisons are based on cumulative utility, which is a combined measurement that incorporates action costs, success bonus, and failure penalty.² In real-world experiments, we illustrate a complete learning process of a real robot arm performing “stack and push” tasks.

A. Experiments in Simulation

We used an open-source 2D physics-based simulator called Pybox2D [39] to evaluate the performance of TMOC. We

²The action cost of `pickup`, `stack`, and `push` are 15, 10 and 30, respectively, where a cost technically corresponds to a negative reward. The bonus of successfully stacking a block, and pushing a container to a goal location is 40 and 80, respectively. The failure penalty is 50.

decided to use a 2D simulator instead of a 3D one because 3D simulators (e.g., Gazebo [40] and PyBullet [41]) are generally computationally less efficient. Our agent needs to learn to interact with objects (with unknown properties) at both task and motion levels, which requires considerable interaction experience, so we selected a 2D simulator that allows running experiments extensively.

Fig. 3 illustrates how a robot gripper uses a container to move eight blocks to a goal area³. The robot knows that all blocks share the same density, but must estimate their sizes for task-motion planning. The robot might fail in loading and unloading a block, and in blocks falling off from the container. Here we assume a “helper” helps move blocks out of the container in the goal area before the gripper and container are moved back. We use this task to capture complex contact-based interactions among multiple objects, whereas the interactions in “stack-and-push” scenarios (used for real-robot experiments) do not go beyond two objects.

Baselines: Three baseline methods are utilized in this research, and they are selected from the literature, referred to as **TMP-RL** [32], **GP** [31], and **TOEP** [33].

- **TMP-RL** is a task-motion planning algorithm that learns from trial and error. TMP-RL is not object-centric, and hence does not learn object properties (L) over time.
- **GP** is a task-motion planning algorithm that learns primitive skills from trial and error, but cannot improve its task-level planning strategies. In our implementation of GP, the agent randomly selects one of the satisficing plans to achieve high-level goals.
- **TOEP** does not learn the state mapping function, and the gripper is placed (on a block) in a grasping position that is randomly selected in a reasonable range. It should be noted that our “TOEP” agent is equipped with task-planning capability, whereas the original TOEP work used predefined task-level behaviors.

TMOC vs. Three Baselines: For every approach, we conducted 15 runs with 5000 episodes in each run. Fig. 4 (Left) illustrates the learning curves for utility value, averaged over the 15 runs with the shaded regions representing one standard deviation from the mean. From the results, we get the important observation that TMOC⁴ performs better than the baselines in terms of cumulative utility and learning rate.

One interesting observation from Fig. 4 (Left) is that TOEP fell behind the other three approaches (including our TMOC) at the beginning, and then later reached a cumulative utility level that is comparable to that of TMOC. This is because TOEP does not learn the state mapping function, and hence cannot improve its skills of “grasping” and “pushing” from trial and error. This disadvantage affects the learning rate of TOEP at the beginning. At the late learning phase (after about 2000 episodes), the TOEP agent was able to catch up, because the task planner learned to adapt to the motion planner (that is suboptimal on grasping and pushing). As a result, the ultimate

³In this paper, we assume blocks and the container are not too heavy to be manipulated by the robot. Besides, the blocks are small enough to be graspable by the robot arm.

⁴The number of grounded worlds in TMOC is 200 (i.e., $N = 200$).

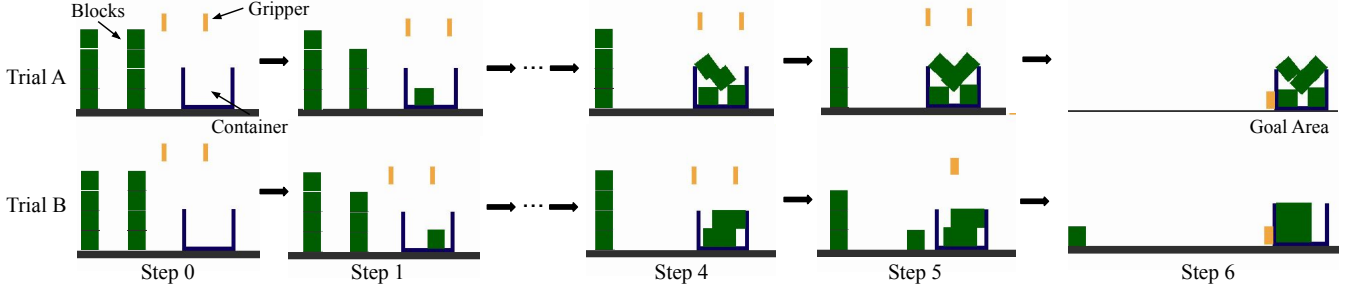


Fig. 3. A robot needs to stack a set of blocks (Steps 0 ~ 6 of Trial A and B) and then push them to the goal area with a container. The more blocks in the container, the less stable it is, resulting in a trade-off between work efficiency and success rate. At the task level, the robot is concerned about how many blocks (of initially unknown sizes and weights) can be placed in the container. Step 4 of trials A and B demonstrate two such situations, where the robot needs to decide if another block can be added to the container. Making this decision is difficult due to the complex multi-object interactions. The two trials give different results as shown in Step 5 of trials A and B.

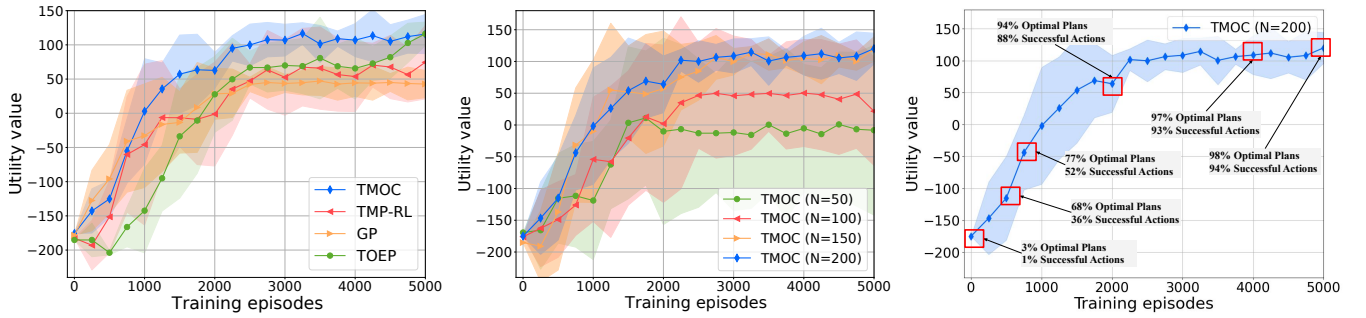


Fig. 4. **Left:** Comparisons between TMOC and three baselines of TMP-RL, GP and TOEP in the task of a gripper moving objects using a container; **Middle:** Performance of TMOC under different numbers of simulated worlds, where $N = 50, 100, 150$, and 200 ; **Right:** Milestones of TMOC, where we present the percentages of optimal plans and successful actions. In all subfigures, the y -axis refers to the cumulative utility obtained by the real robot.

performances of TOEP and TMOC (ours) are comparable. TMP-RL does not learn object properties (L), which affected its learning rate and cumulative utility level compared with TMOC (ours), because an inaccurate L is detrimental to the learning of both Y and T .

Number of Grounded Worlds: The performance of TMOC highly depends on the number of simulated worlds (or “particles,” in the terminology of Particle Filters), which is referred to in Line 2 of Algorithm 1. Each curve was based on 12 runs with 5000 episodes in each run. Sufficient particles are necessary to ensure the quality of a Particle filter, and to represent the distributions being estimated. We are interested in answering this question: *How many particles are adequate for our experiments?* Thus, we evaluated TMOC under different numbers of particles (i.e., $50 \leq N \leq 200$).

Fig. 4 (Middle) shows no significant improvement in learning rate or cumulative utility when we increased the particle number from $N = 150$ to $N = 200$. Therefore, we believe $N = 150$ is sufficient to the robot in our domain. This observation can serve as a reference to TMOC practitioners.

Milestones of TMOC: We are also interested in how the robot makes progress in learning L , Y , and T while running TMOC. Thus, we calculated the percentages of optimal plans and successful actions at some selected training episodes, where $N = 200$ in this experiment.

Fig. 4 (Right) shows that both percentages of optimal plans

and successful actions are close to 0% at the beginning, because of the highly inaccurate functions of L , Y , and T . They gradually increased and finally got close to 100% after 5000 episodes. At the same time, we see that L , Y , and T have different learning rates. For instance, when the robot completed about 2000 episodes, T converged well because most of the plans (about 94%) are optimal, while there is still room to further learn functions L and Y because only 88% of actions are successful. We attribute the growth of utility value after 2000 episodes to the learning of L and Y .

Task Variations of TMOC: The performance of TMOC is affected by task variations, such as blocks of different sizes. Thus, we changed the goal specification to let the robot stack and move one big block and seven small ones. The side length of the big one is 20% bigger than the small ones. The robot knows that all blocks share the same density, but must estimate their sizes for task-motion planning. Introducing different sizes makes the stack-and-push task too difficult for the robot. For evaluation purposes, we provided the robot with guidance that the big block should be put near the bottom (in one of the first two steps).

Fig. 6 shows that under the “different sizes” domain variation, TMOC needs more episodes to reach a utility level that was achieved when all blocks share the same shape. Also, under the “different sizes” variation, TMOC’s convergence level was lower than that under the “same size” variation. We

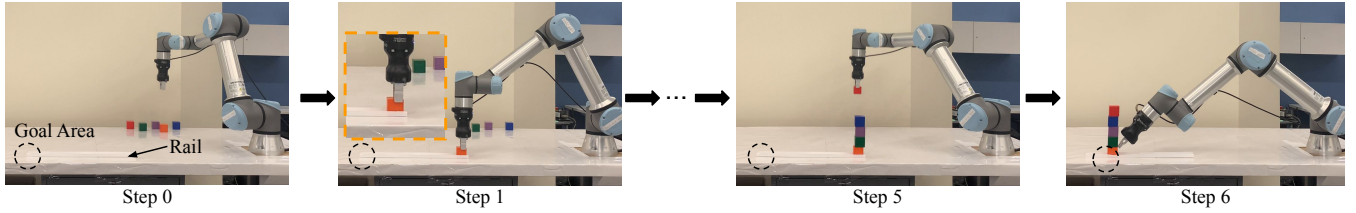


Fig. 5. A robot arm is tasked to move five blocks from an initial area. To perform “2D” experiments in the 3D real world, we attached a rail onto a table, where blocks can be only moved back and forth. There is a trade-off between stability and efficiency (higher stacks are less stable) that the robot must learn from trial and error. In this case, five blocks are stacked and then moved to the goal area. The robot was lucky enough to be successful in this trial, but might fail in other trials of performing the same plan.

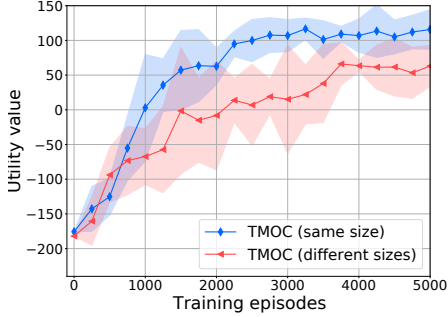


Fig. 6. Performance of TMOC under a task variation (i.e., size). **Same Size:** the robot is tasked to stack and move eight small blocks. **Different Sizes:** the robot is tasked to stack and move one big block and seven small ones. In both curves, N equals 200.

TABLE II
UTILITY VALUE OF TMOC UNDER DIFFERENT EPISODES IN THE
STACK-AND-PUSH TASK USING REAL ROBOT ARM UR5E

Episode	0	4	8	12	16
Utility	-23(18)	60(128)	-10(32)	4(67)	183(80)
Episode	20	24	28	32	36
Utility	70(98)	86(113)	105(74)	218(11)	235(0)

observed that grasps become more unreliable when the robot faces a bigger block. Incidentally, we found no significant difference in the performance of TMOC with and without block weight variations.

B. Experiments in the Real World

TMOC was evaluated using a UR5e robotic arm. In the experiment setting, the robot stacked blocks and then pushed them to a goal area, as illustrated in Fig. 5. We provided the robot with accurate object properties (block size in this case) to control the difficulty of learning task-motion behaviors to a reasonable level. As a result, the robot only needed to learn a state mapping function (e.g., where to place the gripper for grasping) and a transition function (e.g., a stack-and-push strategy, including how many blocks should be stacked together). We conducted three runs, each including 44 episodes, that together took more than 12 hours.

Table II shows the mean utility values and the corresponding standard variances of the TMOC algorithm after different numbers of episodes. At the beginning, the mean value is low while the variance is high, because the robot does not know how to grasp blocks or what a good action sequence is

like. With more trials, our TMOC robot learned a good state mapping function Y and transition function T , which resulted in a high utility value and a low variance. More specifically, the robot learned reasonably good Y and T functions after 15 and 32 episodes, respectively. The result demonstrates that learning Y and T enables the robot to complete the stack-and-push task efficiently and reliably, which is overall evaluated using the utility values.

VI. CONCLUSION AND FUTURE WORK

In this paper, we develop a grounded task-motion planning algorithm TMOC that can ground objects (including their physical properties) for robot planning in mobile manipulation tasks that involve complex multi-object dynamic interactions (e.g., building blocks). We assume the availability of a high-fidelity simulator, and the unavailability of accurate properties of those objects (e.g., size, weight, and shape). The robot needs to plan high-level behaviors to fulfill complex goals that require picking, moving, and placing the objects while minimizing overall action costs. Thus, we develop the TMOC algorithm to learn the object properties, state mapping function and transition function. Results from both the simulation and the real world demonstrate that TMOC improves the task-completion effectiveness and efficiency in terms of utility value and learning rate.

The paper assumes that all objects are relevant to the current task, and hence TMOC grounds all of them in physics-based simulation. In the future, we plan to enable the robot to learn what objects should be grounded and when. Objects have many different physical properties, and estimating their values requires different perception methods. This paper considered size and weight, and other properties can be estimated in future work. Another direction for future research is to further increase the number of particles (each corresponding to a grounded world) for more accurately estimating object properties, where if needed, distributed computational platforms can be used. We are also interested in applying TMOC to domains beyond “stack and push” in the future. Further, our real-robot experiment can be improved by including other everyday objects with diverse properties. Finally, it is important to look into the theoretical properties of TMOC, such as its completeness and scalability. Such theoretical analysis can be difficult due to the stochastic nature of TMOC (e.g., many grounded worlds) and the iterative learning process, and we leave it to future work.

REFERENCES

- [1] M. Ghallab, D. Nau, and P. Traverso, *Automated Planning: theory and practice*. Elsevier, 2004.
- [2] H. M. Choset, S. Hutchinson, K. M. Lynch, G. Kantor, W. Burgard, L. E. Kavraki, S. Thrun, and R. C. Arkin, *Principles of robot motion: theory, algorithms, and implementation*. MIT Press, 2005.
- [3] L. P. Kaelbling and T. Lozano-Pérez, “Integrated task and motion planning in belief space,” *The International Journal of Robotics Research*, vol. 32, no. 9-10, pp. 1194–1227, 2013.
- [4] C. R. Garrett, T. Lozano-Pérez, and L. P. Kaelbling, “Sampling-based methods for factored task and motion planning,” *The International Journal of Robotics Research*, vol. 37, no. 13-14, pp. 1796–1825, 2018.
- [5] M. Toussaint, “Logic-geometric programming: an optimization-based approach to combined task and motion planning,” in *International Conference on Artificial Intelligence (IJCAI)*, 2015.
- [6] N. T. Dantam, Z. K. Kingston, S. Chaudhuri, and L. E. Kavraki, “An incremental constraint-based framework for task and motion planning,” *The International Journal of Robotics Research*, vol. 37, no. 10, pp. 1134–1151, 2018.
- [7] R. Chitnis, D. Hadfield-Menell, A. Gupta, S. Srivastava, E. Groshev, C. Lin, and P. Abbeel, “Guided search for task and motion plans using learned heuristics,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 447–454.
- [8] S.-Y. Lo, S. Zhang, and P. Stone, “The petion algorithm to plan efficiently for task-level-optimal navigation,” *Journal of Artificial Intelligence Research*, vol. 69, pp. 471–500, 2020.
- [9] C. R. Garrett, R. Chitnis, R. Holladay, B. Kim, T. Silver, L. P. Kaelbling, and T. Lozano-Pérez, “Integrated task and motion planning,” *Annual review of control, robotics, and autonomous systems*, vol. 4, pp. 265–293, 2021.
- [10] F. Lagriffoul, N. T. Dantam, C. Garrett, A. Akbari, S. Srivastava, and L. E. Kavraki, “Platform-independent benchmarks for task and motion planning,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3765–3772, 2018.
- [11] S. Cambon, R. Alami, and F. Gravot, “A hybrid approach to intricate motion, manipulation and task planning,” *The International Journal of Robotics Research*, vol. 28, no. 1, pp. 104–126, 2009.
- [12] C. Dornhege, P. Eyerich, T. Keller, S. Trüg, M. Brenner, and B. Nebel, “Semantic attachments for domain-independent planning systems,” in *International Conference on Automated Planning and Scheduling*, 2009.
- [13] L. P. Kaelbling and T. Lozano-Pérez, “Hierarchical planning in the now,” in *Workshops at the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.
- [14] C. R. Garrett, T. Lozano-Pérez, and L. P. Kaelbling, “Ffrob: Leveraging symbolic planning for efficient task and motion planning,” *The International Journal of Robotics Research*, vol. 37, no. 1, pp. 104–136, 2018.
- [15] Y. Wang, N. Dantam, S. Chaudhuri, and L. Kavraki, “Task and motion policy synthesis as liveness games,” in *Proceedings of the International Conference on Automated Planning and Scheduling*, 2016.
- [16] S. Srivastava, E. Fang, L. Riano, R. Chitnis, S. Russell, and P. Abbeel, “Combined task and motion planning through an extensible planner-independent interface layer,” in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 639–646.
- [17] X. Zhang, Y. Zhu, Y. Ding, Y. Zhu, P. Stone, and S. Zhang, “Visually grounded task and motion planning for mobile manipulation,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.
- [18] G. Konidaris, L. P. Kaelbling, and T. Lozano-Pérez, “From skills to symbols: Learning symbolic representations for abstract high-level planning,” *Journal of Artificial Intelligence Research*, vol. 61, pp. 215–289, 2018.
- [19] N. Gopalan, E. Rosen, G. Konidaris, and S. Tellex, “Simultaneously learning transferable symbols and language groundings from perceptual data for instruction following,” *Robotics: Science and Systems XVI*, 2020.
- [20] N. Jakobi, P. Husbands, and I. Harvey, “Noise and the reality gap: The use of simulation in evolutionary robotics,” in *European Conference on Artificial Life*. Springer, 1995, pp. 704–720.
- [21] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, “Sim-to-real: Learning agile locomotion for quadruped robots,” *arXiv preprint arXiv:1804.10332*, 2018.
- [22] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, “Sim-to-real transfer of robotic control with dynamics randomization,” in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 3803–3810.
- [23] A. Molchanov, T. Chen, W. Hönig, J. A. Preiss, N. Ayanian, and G. S. Sukhatme, “Sim-to-(multi)-real: Transfer of low-level robust control policies to multiple quadrotors,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 59–66.
- [24] W. Yu, J. Tan, C. K. Liu, and G. Turk, “Preparing for the unknown: Learning a universal policy with online system identification,” *arXiv preprint arXiv:1702.02453*, 2017.
- [25] L. Pinto, J. Davidson, R. Sukthankar, and A. Gupta, “Robust adversarial reinforcement learning,” in *International Conference on Machine Learning*. PMLR, 2017, pp. 2817–2826.
- [26] A. Farchy, S. Barrett, P. MacAlpine, and P. Stone, “Humanoid robots learning to walk faster: From the real world to simulation and back,” in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, 2013, pp. 39–46.
- [27] Y. Chebotar, A. Handa, V. Makoviychuk, M. Macklin, J. Issac, N. Ratliff, and D. Fox, “Closing the sim-to-real loop: Adapting simulation randomization with real world experience,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8973–8979.
- [28] S. Zhu, A. Kimmel, K. E. Bekris, and A. Boularias, “Fast model identification via physics engines for data-efficient policy search,” *arXiv preprint arXiv:1710.08893*, 2017.
- [29] R. Jeong, Y. Aytar, D. Khosid, Y. Zhou, J. Kay, T. Lampe, K. Bousmalis, and F. Nori, “Self-supervised sim-to-real adaptation for visual robotic manipulation,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2718–2724.
- [30] Y. Ding, X. Zhang, X. Zhan, and S. Zhang, “Task-motion planning for safe and efficient urban driving,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [31] Z. Wang, C. R. Garrett, L. P. Kaelbling, and T. Lozano-Pérez, “Learning compositional models of robot skills for task and motion planning,” *The International Journal of Robotics Research*, vol. 40, no. 6-7, pp. 866–894, 2021.
- [32] Y.-q. Jiang, S.-q. Zhang, P. Khandelwal, and P. Stone, “Task planning in robotics: an empirical comparison of pddl-and asp-based systems,” *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 3, pp. 363–373, 2019.
- [33] J. Liang, S. Saxena, and O. Kroemer, “Learning active task-oriented exploration policies for bridging the sim-to-real gap,” *arXiv preprint arXiv:2006.01952*, 2020.
- [34] R. I. Brafman and M. Tennenholtz, “R-max-a general polynomial time algorithm for near-optimal reinforcement learning,” *Journal of Machine Learning Research*, vol. 3, no. Oct, pp. 213–231, 2002.
- [35] N. J. Gordon, D. J. Salmond, and A. F. Smith, “Novel approach to nonlinear/non-gaussian bayesian state estimation,” in *IEE Proceedings F-radar and signal processing*, vol. 140, no. 2, 1993, pp. 107–113.
- [36] V. Lifschitz, “Answer set programming and plan generation,” *Artificial Intelligence*, vol. 138, no. 1-2, pp. 39–54, 2002.
- [37] S. Amiri, S. Bajracharya, C. Goktolgal, J. Thomason, and S. Zhang, “Augmenting knowledge through statistical, goal-oriented human-robot dialog,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 744–750.
- [38] S. Karaman and E. Frazzoli, “Sampling-based algorithms for optimal motion planning,” *The international journal of robotics research*, vol. 30, no. 7, pp. 846–894, 2011.
- [39] J. Hunt, “Introduction to games programming,” in *Advanced Guide to Python 3 Programming*. Springer, 2019, pp. 121–123.
- [40] N. Koenig and A. Howard, “Design and use paradigms for gazebo, an open-source multi-robot simulator,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.
- [41] E. Coumans and Y. Bai, “Pybullet, a python module for physics simulation for games, robotics and machine learning,” 2016.