# SafeMimic: Towards Safe and Autonomous Human-to-Robot Imitation for Mobile Manipulation

Arpit Bahety, Arnav Balaji, Ben Abbatematteo, Roberto Martín-Martín
The University of Texas at Austin

***Abstract*—For robots to become efficient helpers in the home, they must learn to perform new mobile manipulation tasks simply by watching humans perform them. Learning from a single video demonstration from a human is challenging as the robot needs to first extract from the demo *what* needs to be done and *how*, translate the strategy from a third to a first-person perspective, and then adapt it to be successful with its own morphology. Furthermore, to mitigate the dependency on costly human monitoring, this learning process should be performed in a safe and autonomous manner. We present SAFEMIMIC, a framework to learn new mobile manipulation skills safely and autonomously from a single third-person human video. Given an initial human video demonstration of a multi-step mobile manipulation task, SAFEMIMIC first parses the video into segments, inferring both the semantic changes caused and the motions the human executed to achieve them and translating them to an egocentric reference. Then, it adapts the behavior to the robot's own morphology by sampling candidate actions around the human ones, and verifying them for safety before execution in a receding horizon fashion using an ensemble of safety Q-functions trained in simulation. When safe forward progression is not possible, SAFEMIMIC *backtracks* to previous states and attempts a different sequence of actions, adapting both the trajectory and the grasping modes when required for its morphology. As a result, SAFEMIMIC yields a strategy that succeeds in the demonstrated behavior and learns task-specific actions that reduce exploration in future attempts. Our experiments show that our method allows robots to safely and efficiently learn multi-step mobile manipulation behaviors from a single human demonstration, from different users, and in different environments, with improvements over state-of-the-art baselines across seven tasks. For more information and video results, https://robin-lab.cs.utexas.edu/SafeMimic/**

## I. INTRODUCTION

For decades, we have dreamed to teach robots a new task in the same way we would teach it to another human: by demonstrating it in from of them. This would bypass the need for costly teleoperated data collection [1, 2, 3], which is significantly complex and time-consuming for multi-step tasks and those combining navigation and manipulation into mobile manipulation. Recent advances in human motion perception and parsing [4, 5, 6, 7] have brought us closer to this dream: they have enabled new approaches that extract information from human videos and use it to "seed" an exploratory process in which the robot adapts the strategy to its own embodiment [8, 9, 10, 11]. However, these techniques are restricted to **short horizon** skills and require tedious **human supervision**, tasked with ensuring that the robot exploration is **safe**, **resetting** the task constantly and **detecting success**. What is necessary to enable robots to learn multi-step mobile manipulation tasks from a single human demonstration **safely** and **autonomously**?

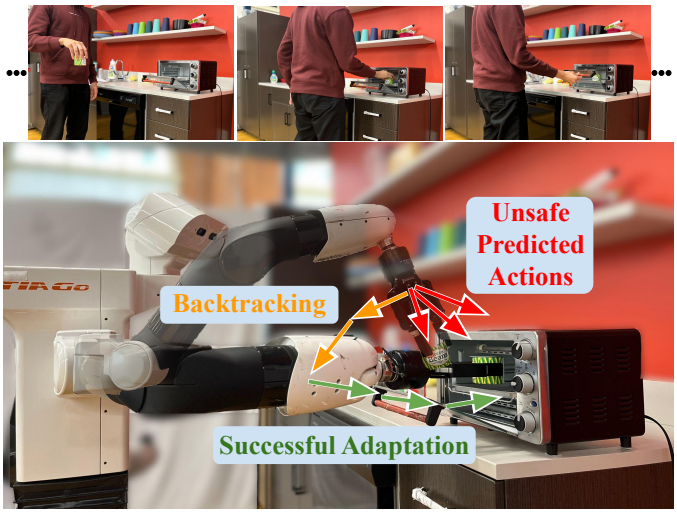


Fig. 1. Robot imitating a single video of a human-demonstrated mobile manipulation task safely and autonomously with SAFEMIMIC. From a video of a multi-step mobile manipulation task (*top*), SAFEMIMIC extracts an initial motion strategy and adapts it to its own embodiment by exploring new actions in a safe manner. It combines an ensemble of safety Q-functions that predict future unsafe motions (*red arrows*, motion will collide) and a backtracking mechanism (*orange arrows*) that enables autonomous exploration until a successful action sequence is found (*green arrows*).

Learning multi-step tasks from a human video in a safe and self-supervised manner presents multiple technical challenges. First, it requires for the robot to understand both the high-level semantics of the task and the associated low-level motion. This implies extracting from the video both the human's motion as well as the semantic changes they caused in the environment. This information needs to be translated from a third to a first person view so that the robot can execute the motion and monitor the semantic changes with its onboard sensors. When executed, the initial translated motion may fail due to morphological differences or noise in the video parsing; while finding small adaptations have been shown feasible through trial-and-error learning in the real world [8, 9, 10, 11], it is still unclear how to explore for longer multi-step tasks that require a larger adaptation, e.g., for different grasping strategies. Finally, such a real-world exploration quickly becomes unsafe due to potential damage to the robot or the environment, requiring another agent (human) to monitor and reset the scene for further exploration. Safe and autonomous exploration calls for a new method that allows the robot to predict when something may go wrong before it happens and backtrack to previous states to keep trying new strategies.

In this work, we introduce SAFEMIMIC, a framework for safely and autonomously learning mobile manipulation behaviors from a single third-person human video. SAFEMIMIC overcomes all aforementioned challenges: first, it parses the third-person demonstration combining a human motion tracker and a vision-language model (VLM), obtaining an initial task plan composed of sequences of distinct human motion and easy-to-detect semantic changes executable from a first-person perspective. Then, SAFEMIMIC adapts the initial plan by refining each segment using a safe exploration procedure in the real world. At the core of our safe exploration sits an ensemble of **safety Q-functions** pretrained in simulation that enable SAFEMIMIC to attempt risk-free actions both for continuous motion as well as discrete grasps, overcoming large differences in morphology and manipulation capabilities. General safety Q-functions have the potential to generalize across tasks as they capture generic risks common to manipulation –such as collisions, excessive forces, or grasp losses– and, as we observe empirically, they require less precise sim-to-real alignment than direct policy transfer. When SAFEMIMIC detects no safe actions to progress in the exploration, it backtracks to a previous state and tries a different strategy, including changing the grasping mode if necessary. Finally, when a safe and successful adaptation is found, SAFEMIMIC associates it to the geometry of the task so that exploration is reduced in future attempts.

In summary, SAFEMIMIC introduces several novel contributions:

- A comprehensive framework to parse a single multi-step mobile manipulation video demonstration from a human and adapt it to the robot's capabilities in a safe and self-supervised manner.
- A mechanism to parse human videos into sequences of motions with unique and easy-to-perceive semantic effects combining human pose tracking and VLM detections.
- A safe exploration in the real world with a predictive control strategy informed by an ensemble of safety Q-functions pretrained in simulation.
- A self-supervised mechanism to detect success and to backtrack to previous states to try new strategies, including different grasp modes.
- A mechanism to learn from previous experiences to reduce the amount of exploration necessary in future attempts.

We demonstrate the performance of SAFEMIMIC in seven challenging, multi-step mobile manipulation tasks in different environments with different human teachers, and observe experimentally that our framework enables the robot to successfully acquire the desired behaviors safely and more efficiently than direct sim-to-real imitation learning approaches, previous human-to-robot methods, and variants without safety Q-functions.

## II. RELATED WORK

SAFEMIMIC is a novel framework for safely and autonomously learning multi-step mobile manipulation tasks from human demonstrations. In this section, we contrast SAFEMIMIC to prior efforts on learning from human video as well as safe imitation and reinforcement learning.

**Learning from Human Video** directly has received increasing attention as strategy to learn manipulation skills. Some works have explored leveraging large collections of human activity data [12, 13, 14] in order to learn cost functions from video and language data [15, 16, 17, 18]. Other works have explored human video modeling as a pretraining objective [19, 20]. Most closely related to SAFEMIMIC, several works imitate human actions directly by tracking the human pose and extracting actions using pose tracking before finetuning with gradient-free RL [8, 21, 9, 10, 22, 11]. These works demonstrate impressive behaviors, yet typically refine the initial policy obtained from the human in a naive exploratory fashion. This has the potential to damage the robot or environment, as actions are not checked for collision or other potentially dangerous failures. Other works rely on morphological similarity between humans and humanoid robots to retarget motions directly [23, 24, 25, 26]. In contrast, we focus on imitating and refining the demonstration safely when the robot and human embodiments do not necessarily match. Moreover, most existing works on learning from human video focus on short-horizon tasks (e.g., opening a drawer) rather than multi-step mobile manipulation behaviors.

**Autonomous Real-World Learning** prior methods have examined learning to automatically reset the environment in order to enable learning without human supervision [27, 28, 29, 30, 31, 32]. These works address the requirement of autonomy, but generally sidestep the question of safety — a critical challenge when learning mobile manipulation in the real world. Further, these methods require extensive trial-and-error learning not suitable for efficiently learning from a single human video demonstration. SAFEMIMIC instead combines both autonomous and safe exploration when learning from human video and employs a simple but effective backtracking strategy when failures are predicted for all sampled actions in a given state.

**Safe Imitation and Reinforcement Learning** has received significant study in order to reduce risks in the real world. *Design-time* approaches [33, 34, 35, 36, 37, 38] operate during data collection or model training in order to ensure robustness to perturbation during execution. When learning from human video, design-time approaches are infeasible, as they are typically targeted for teleoperation when disturbances can be injected. *Deployment-time* approaches [39, 40, 41, 42, 43] filter actions and defer to a backup or recovery policy when constraints are violated. Deployment-time methods usually assume access to constraints in closed form, an unrealistic requirement in novel environments and tasks. SAFEMIMIC employs learned safety Q-functions that are pretrained from simulation data, allowing it to learn from human video demonstrations safely.

Safe RL methods [44, 45, 46, 47] provide a framework for safe policy learning, enabling the discover of complex behaviors from scratch while avoiding failures in a target task. These approaches typically jointly optimize task and

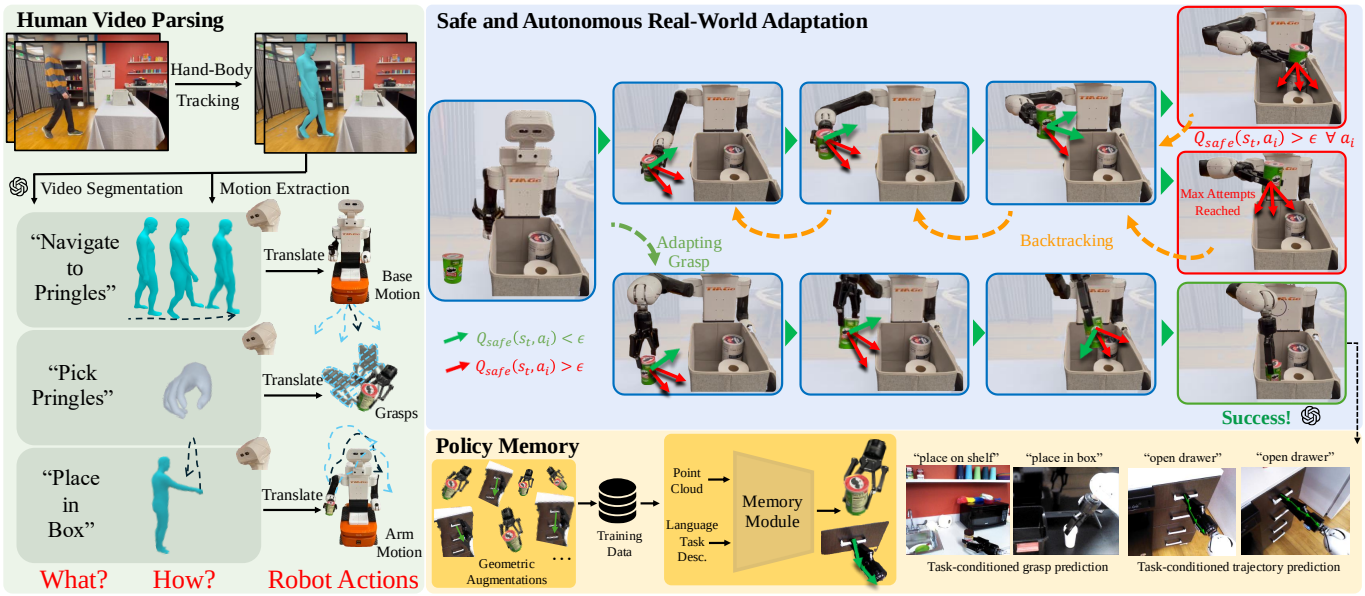


Fig. 2. **Overview of SAFEMIMIC.** From an RGB-D video of a human performing a multi-step mobile manipulation task acquired by the robot, SAFEMIMIC uses a combination of human pose tracking models [4, 5] and VLM prompting to perform coarse-to-fine segmentation obtaining semantic changes –"what?"– and human action trajectories –"how?"–, and translating them to the robot's point of view (*left*, Sec. III-A). SAFEMIMIC then refines and adapts each task segment safely by sampling and verifying actions before executing them thanks to an ensemble of safety Q-functions pretrained in simulation, $Q_{\text{safe}}$ (*top right*, Sec. III-B). If forward progress is not possible, SAFEMIMIC autonomously backtracks and tries different actions (*orange arrows*). When required to overcome large differences in morphology, SAFEMIMIC explores alternative grasp modes (*second row of samples*), adapting the grasp to enable successful execution. Successful attempts are detected by a VLM that verifies when the parsed semantic change of the segment is achieved. Successes are stored and used to train a policy memory module with geometric augmentations (*bottom right*, Sec. III-C) that predicts actions (grasps or action trajectories) in subsequent attempts, given a pointcloud and language task description, in order to reduce the need for exploration.

safety Q-functions during a pre-training phase, and do not address the case of imitation learning or learning from human video. In contrast, we employ task-agnostic data collection in simulation, and learn directly from human teachers in the real world. Constrained RL methods [48, 49, 50, 51] similarly allow for policy learning while obeying constraints, though typically require closed-form constraints available at runtime. In the real-world, such constraints are difficult to acquire in novel environments, much less in the presence of human teachers.

**Failure Prediction** ahead of execution has received considerable study in robotics, often with the goal of providing safety certificates on policies for deployment [52, 53, 54, 55]. Many works approach failure prediction through the lens of reachability analysis [56, 57] or design control barrier functions [58]. Recent work has sought to learn failure predictors with PAC guarantees [59] or conformal prediction [60]. However, these approaches assume access to a black box policy or dynamics model of the environment, both which are unknown in the case of learning a new task from human video in a novel environment. Similarly, motion planning methods [61, 62] enable collision-free motion generation for a given environment geometry but fail to capture other possible failure modes involving contact, such as force-torque limit violations or grasp loss. SAFEMIMIC provides a unified framework for failure prediction when learning mobile manipulation behaviors from human videos.

## III. SAFEMIMIC FRAMEWORK

Our goal is to enable a robot to safely and autonomously learn to adapt and imitate a multi-step mobile manipulation task demonstrated in a single third-person human video. Figure 2 provides an overview of our framework, SAFEMIMIC. In this section, we describe each of its components.

### A. Factorizing, Parsing and Translating Human Videos

The first step to imitate a human video demonstration is to extract information from it: *what* did the human do, and *how* did they do it? In SAFEMIMIC, we identify the *what* by detecting the semantic changes caused by the human in the environment, and the *how* by tracking the motion of the human that caused those changes. However, as we aim at imitating complex multi-step mobile manipulation tasks, extracting multiple semantic changes and intricate human motion for the entire video and learning to adapt it to the robot all at once may be unfeasible. SAFEMIMIC factorizes the original video into distinct segments where only a single semantic change happens. This naturally leads to segments with separate navigation and manipulation to achieve single semantic changes such as `navigate_to`, `reach_for_and_grasp`, `open`, ...(see full list in Appendix B) that can be adapted and optimized sequentially in a self-supervised manner by the robot.

To factorize and parse the multi-step video into single semantic change segments, SAFEMIMIC combines a body and hand visual tracker [4, 5], a contact detector [6], and a VLM [63]. An initial coarse segmentation is obtained based on the tracked human motion by detecting if the human is

navigating or manipulating by thresholding the amount of inter-frame body translation. Then, SAFEMIMIC annotates each coarse segment with its semantic change and possibly factories it further. Navigation segments are assumed to be at the lowest level of granularity; SAFEMIMIC just needs to extract the semantic goal, the object/location that the navigation segment tries to reach, obtained using a VLM. Manipulation segments are further factorized by SAFEMIMIC as necessary by combining information from a VLM and a contact detector that separates manipulation phases without contact from the ones with contact. This leads to a natural factorization into segments that begin with a grasping action or with a change in the contact interface, e.g., when a wiping segment begins. Each segment is then annotated with its semantic goal, again obtained with a VLM query.

The factorization and parsing process mentioned above leads to a sequence of segments with a individual semantic goals and motions of the body and/or arm of the human (Fig. 2, left). While the semantic change is invariant to the point of view, the motion is viewed from a third-person perspective and needs to be translated into the robot's reference frame to be executed and monitored with the onboard sensors. Similarly, the human grasps may need to be translated to the robot morphology. To that end, SAFEMIMIC assumes that in navigation segments, the robot starts from a location *close enough* to the human initial location (no calibration needed!) and translates the sequence of human navigation actions to relative changes in base pose between frames. Similarly, for manipulation segments, SAFEMIMIC first transforms the hand pose to be relative to the human body and then computes the relative motion of the hand between consecutive frames. To translate grasps, SAFEMIMIC will detect grasp candidates available to the robot [64] and match the one that is closest to the grasp demonstrated by the human. However, other grasp candidates will be explored as a result of our safe and autonomous human-to-robot adaptation, as we explain in the next section.

### B. Safe and Autonomous Real-World Adaptation

The factorizing, parsing, and translating process described above leads to an initial policy that the robot could directly execute. However, this initial policy will fail due to differences in embodiment and tracking inaccuracies (see Exp. IV). SAFEMIMIC implements an exploration and adaptation procedure in the real world to find the actions (close to the ones demonstrated by the human) that will lead to the same sequence of semantic changes and thus to success on the multi-step mobile manipulation task. To that end, SAFEMIMIC explores actions for each of the segments in turn, until the semantic goal of the segment is achieved in a process summarized in pseudocode in Alg. 1. In contrast to previous human-to-robot approaches, SAFEMIMIC will perform this real-world exploration *safely and autonomously* thanks to a combination of safety Q-functions and backtracking capabilities (see Fig. 2, right).

*Safe Exploration with Safety Q-functions:* We assume that, at each segment, the robot's objective is to achieve the segment's semantic goal while avoiding unsafe states. We adopt the standard MDP formalism and represent each segment by the tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, R, T, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $R$ is a sparse reward based on task success (e.g., semantic goal achievement), $T$ is the transition function, and $\gamma$ is a discount factor. Using the Safe RL framework proposed by Srinivasan et al. [45], we denote the set of unsafe states $\mathcal{S}_{\text{unsafe}} = s \in \mathcal{S} \mid \mathcal{I}(s) = 1$, where $\mathcal{I}(s)$ is an indicator boolean function that triggers for any unsafe states. This function can then be considered a composition of single failure mode indicators, $\mathcal{I}(s) = \max_i \mathcal{I}_i(s)$, where $\{\mathcal{I}_i\}$ a set of functions for distinct failure modes, e.g., exerting too much force, colliding with the environment, reaching joint limits, ...(for the complete list of unsafe state, see Appendix A). Given this function, the robot's objective is to find a policy that maps states to the actions that maximize the task reward while remaining safe, given formally by:

$$\max_{\pi} \sum_{t=0}^{T} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi} \left[R\left(s_t, a_t\right)\right] \text{ s.t. } \mathbb{E}_{s_t\sim\rho_\pi} \left[\mathcal{I}\left(s_t\right)\right] = 0,$$

where $\rho_\pi$ denotes the state-action distribution visited by the policy $\pi$.

Thus, to satisfy this objective, the robot requires a predictive model of the actions that will result in failure. To achieve this, we define a *Safety Q-function* that predicts the probability that an action $a_t$ taken in state $s_t$ will result in a failure:

$$Q_{\text{safe}}\left(s_t, a_t\right) = \mathcal{I}\left(s_t\right) + \left(1 - \mathcal{I}\left(s_t\right)\right) \left[\mathbb{E}_{s_{t'}\sim T(\cdot|s_t,a_t)} \mathcal{I}\left(s_{t'}\right)\right]$$

As with the unsafe state indicator function, we can factorize the Safety Q-function into an *ensemble of Safety Q-functions*, $\{Q_{\text{safe,i}}\}_{i=0}^{K}$, each one trained for each type of failure, $Q_{\text{safe}}\left(s_t, a_t\right) = \max_i Q_{\text{safe,i}}\left(s_t, a_t\right)$.

Training these Safety Q-functions in the real world would be dangerous, as the robot needs to experience and learn what actions may lead to unsafe states and when. Therefore, we pretrain an ensemble of Safety Q-functions in simulation, one for each type of unsafe transition. The ensemble of Safety Q-functions is pretrained in domain-randomized environments in the simulator OmniGibson [65] in different scenarios, including articulated object interaction, rigid-body pick-and-place, and base navigation by sampling random and noise-corrupted task-related actions as generated by a motion planner. The state representation consists of simulated pointclouds and robot proprioceptive information (for details of the network architecture, see Appendix A). Pointclouds provide a smaller sim2real gap than RGB images, allowing us to train our ensemble of Safety Q-functions in simulation and apply them zero-shot to safely explore in the real world.

During the safe adaptation and exploration procedure, SAFEMIMIC samples candidate actions around the human demonstrated motion. For the grasping action, SAFEMIMIC explores among a discrete set of options (3 grasps in our case) provided by a robot grasp generator [64], first prioritizing the grasp closest to the human demonstrated one. For

Fig. 3. **Multi-Step Mobile Manipulation Evaluation Tasks.** SAFEMIMIC is evaluated on seven complex multi-step tasks combining navigation and manipulation. From left to right: `boxing` an item, `shelving` an item, `store_in_drawer`, `erase_whiteboard`, `refrigerating` an item, `fill_pot`, `load_oven`. The tasks involve risky contact-rich phases, grasping steps and the manipulation of constrained mechanisms. SAFEMIMIC is able to adapt actions obtained from one single human demonstration safely and autonomously.

continuous motion, SAFEMIMIC explores action sequences from a Gaussian distribution with the mean given by the parsed human motion and variance adapted to encourage exploration. Sampled actions with the lowest $Q_{\text{safe}}$ are selected and executed at each step, after which SAFEMIMIC evaluates for segment completion based on the parsed semantic goal of the segment and a VLM. If the segment has not been completed, the process repeats with a new set of actions around the human demonstrated one (see pseudocode in Alg. 1).

*Autonomous Exploration with Backtracking Mechanisms:* Thanks to the previous process, SAFEMIMIC generates a safe exploratory behavior around the human-demonstrated motion that leads to adaptation. However, human monitoring would still be necessary to reset the robot and the environment and attempt new sequences of actions. We aim at reducing this dependency in SAFEMIMIC. To that end, we implement a simple but effective *backtracking* mechanism. During safe exploration, if no sampled actions are safe from the current state (i.e., $Q_{\text{safe}}(s, a_i) > \epsilon$ for all sampled $a_i$), the agent backtracks one step in the exploration: it executes the inverse to the last action taken, i.e., performs the opposite motion. When interacting with objects (e.g., opening a door or drawer, moving a grasped object), this backtracking mechanism will bring the environment to the previous state, allowing SAFEMIMIC to explore a different branch of the state-action space, verifying and executing actions until either the segment is complete or the maximum number of attempts is reached.

The safe and autonomous exploration from above allows the robot to find minor adaptations to the parsed human trajectories for each segment. However, often, adaptation needs to be larger than a small change in motion, especially when it comes to adapting grasping strategies from a human to a robot. When, after several iterations (50 actions in our case) of the single-step backtracking mechanism, SAFEMIMIC exhausts all safe motion alternatives, it backtracks until it reaches a grasping segment, and selects a different grasp mode to explore. We observe that this exploration of grasping modes combined with trajectory-level probing is critical to adapt multi-step human strategies into successful ones for the robot (see Fig. 5 for examples of grasping mode exploration).

### C. Learning from Previous Successful Exploration

Through our safe and autonomous exploratory process, SAFEMIMIC adapts the motion and sequence of semantic changes initially parsed from the human video. However, successful strategies need to be learned in order to prevent repeated exploration for the same task. To that end, SAFEMIMIC integrates a policy memory mechanism that biases exploration based on previous successes. The policy memory biases the exploration (of grasp modes and of motion actions) from the human demonstrated to one that demonstrates success for the robot. To that end, we then train an action prediction policy network that maps point clouds, $P$ and language description of the task, $l$ to actions, e.g., the grasping mode, $g \in SE(3)$, and sequence of post-grasp actions, $(a_0, \ldots, a_T)$, that led to success. The architecture for the action prediction policy network is composed by a PointNet [66] encoder for the visual information, and a SentenceTransformer [67] for the task description, combined with an MLP head, and trained with geometric augmentations (rotations and translations) of the data from successful trials. This model associates successful strategies to geometric and language information about the task, and allows SAFEMIMIC to predict them from different viewing angles, a critical capability to leverage successful exploration for new instances of the same multi-step mobile manipulation tasks (see Fig. 7). Furthermore, the model also helps reduce future explorations for the same task (see Exp. 3, Fig. 7).

## IV. EXPERIMENTAL EVALUATION

We evaluate SAFEMIMIC in 7 challenging multi-step mobile manipulation tasks demonstrated by humans. The tasks all consist of multiple stages and require navigation, rigid-body pick-and-place, articulated object interaction, and contact-rich control, all while avoiding common failure modes, including collisions, joint limit violations, force-torque limits, and grasp loss. We use these tasks to evaluate SAFEMIMIC's task performance, safety, generalization, and robustness to different users and environments. In the following, we briefly explain each task.

- `boxing` an item: Pick the object from the table and place it in a box, requiring a top-down grasp by the robot to avoid collision.
- `shelving` an item: Pick an object from the table and store it on a shelf above the sink, or in a bookshelf, as demonstrated. This task requires differentiating the human's semantic goal, and avoiding collisions and adapting grasps for successful placement.
- `store_in_drawer`: Open the drawer, pick the object from the table, store the object in the drawer, and close the drawer. This task consists of several segments, including an articulated object interaction that typically necessitates

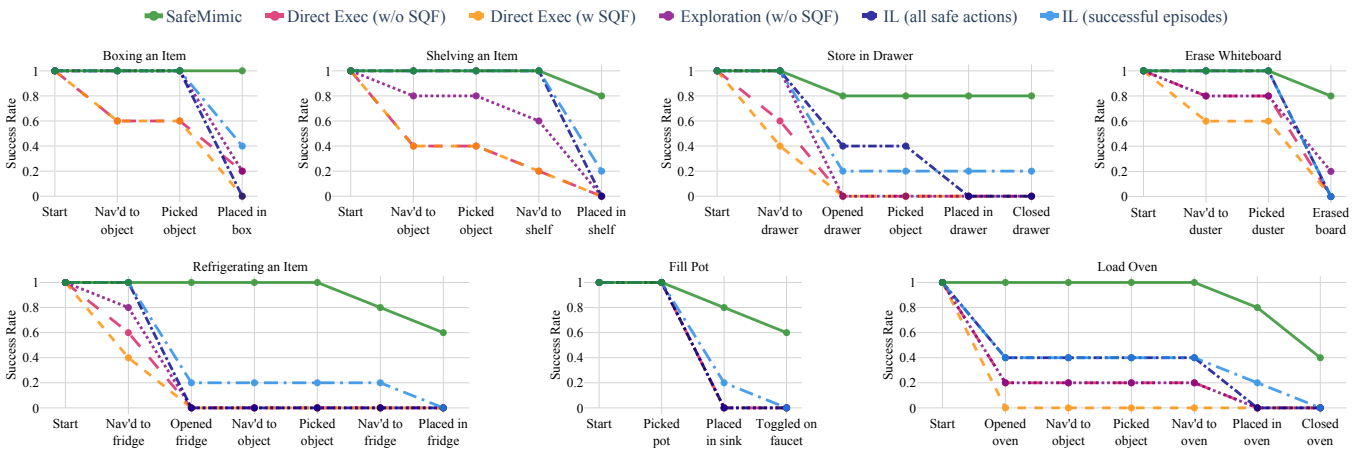


Fig. 4. **Accumulated Success on Multi-Step Tasks**. Accumulated success rate at each stage of each of the seven evaluated multi-step mobile manipulation tasks, indicating the percentage of the five trials each method completed up to and including that segment. We compare SAFEMIMIC's task performance to five baselines: direct execution without safety Q-functions (SQFs), which requires human supervision, direct execution with SQFs, exploration without SQF, Imitation Learning (IL) with all safe actions in simulation and IL only with the successful episodes. Note that the IL baselines were evaluated **only** on "Place", "Open", and "Close" segments and were provided successful solutions for navigation and pick segments. Note as well that some lines overlap at the same rate. Across all tasks, SAFEMIMIC significantly outperforms all baselines and achieves up to 100% success in exploratory adaptation, indicating a superior ability to safely and autonomously refine the demonstrated behaviors.

adapting human-demonstrated grasps to avoid joint limit violations.

- `erase_whiteboard`: Grasp an eraser and erase the writings on a whiteboard. This task requires contact-rich control, demanding that the robot avoid force-torque limits in particular.
- `refrigerating` an item: Open the refrigerator, pick the object from the table, store it in the refrigerator. This task requires adapting the grasp poses and demonstrated trajectories to open the large, heavy refrigerator door and place the object in a constrained space.
- `fill_pot`: Pick a saucepan, place it in the sink, and toggle on the faucet. This task necessitates adapting the commonly used human grasp to one that the robot can use to lower the pot into the sink.
- `load_oven`: Open the oven, pick the object, place it in the oven, and close the oven. This task consists of several segments, and requires a specific side-grasp to successfully place the food in the small oven.

In all the experiments, we use a PAL-Robotics Tiago++ mobile manipulator. We control one arm's end-effector pose using IK control [68] and the base using relative position and yaw commands. For perception, we use an Orbbec Astra S RGB-D camera mounted on Tiago++'s head both to observe humans and as input to the safety Q-functions, and an ATI mini45 force-torque sensor mounted on the wrist to detect and predict excessive force-torque violations. While SAFEMIMIC is generic and can include many possible failure modes, we consider the following in this work: arm collisions, base collisions, joint limit violations, force-torque limits, grasp loss, and dropping objects.

Throughout our experiments, we compare SAFEMIMIC against the following baselines: `Direct Execution (w/o SQF)` directly executes the actions obtained from the video parsing module and does not verify those actions using the Safety Q-Functions (SQFs). `Direct Execution (with SQF)` directly executes the human actions as well but verifies the actions using SQFs, avoiding unsafe robot actions. `Exploration (w/o SQF)` performs exploration starting from the human tracking actions but does not use SQFs. This baseline is SAFEMIMIC without the use of SQFs. We also evaluate if the data generated to train our safety Q-functions would suffice for training task policies: we include `Imitation Learning (IL)` baselines based on a BC-RNN Behavior Cloning policy with a PointNet encoder trained on two types of datasets: `IL (all safe actions)` in which the policy is trained on all simulation data where the state-action pairs did not lead to safety criteria being violated, and `IL (successful episodes)` which is trained only on the subset of successful task executions in the simulator. Since the data collection for navigation is task-agnostic (random base commands), we can not train a task-oriented IL policy on that data. Furthermore, since SafeMimic performs picking action based on a grasp generator, we also don't train an IL policy for picking. Instead, we use the successful navigation and picking segments from the SAFEMIMIC trials for this baseline. Hence, navigation to an object and picking always succeeds and we focus the comparison on the manipulation segments of "Place", "Open", and "Close" for this baseline. Note that we trained separate IL policies for each segment.

*Experiments and Results*

In our experiments, we aim to answer four questions:
**Q1) Does SAFEMIMIC enable a robot to successfully complete a multi-step mobile manipulation task from a third-person demonstration?** To evaluate SAFEMIMIC's task performance, we measure its ability to successfully imitate and adapt demonstrations of the aforementioned tasks. For each task, SAFEMIMIC and the baselines begin with the initial

Fig. 5. **Grasping mode adaptation.** Two examples (*top and bottom rows*) of SAFEMIMIC's grasping mode adaptation. *Left column:* human demonstrated grasp. *Middle column:* robot failing when attempting the task by matching the human grasp. *Right column:* robot succeeding in the task through SAFEMIMIC's grasp adaptation. In 6 out of 7 tasks, SAFEMIMIC's grasp adaptation is critical to overcome human-robot embodiment differences and successfully imitate the demonstration.

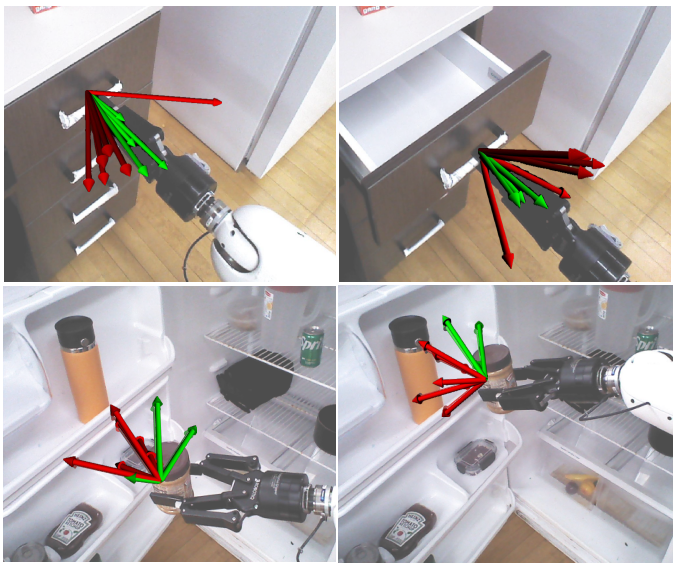

Fig. 6. **Safe exploration with Safety Q-function predictions.** Examples of predictions of the Safety Q-function (SQF) for two tasks: the opening drawer segment in `store_in_drawer` (top-row) and placing a bottle in the fridge segment in `refrigerating` (bottom-row). Red arrows indicate predicted unsafe actions ($Q_{\text{safe}} > \epsilon$) and green arrows show predicted safe actions. We observe that the simulation-trained SQFs reliably predict unsafe actions such as trying to open the drawer by moving in the wrong direction or trying to place an object in the fridge by colliding with the orange bottle. This enables SAFEMIMIC to safely explore and adapt demonstrations in the real world.

trajectory obtained by tracking the human pose, broken into semantic segments identified by the parsing module. Each task is attempted five times, and success rates are reported at each stage of each task. The success rate for each segment (e.g., "Navigated to object," "Picked object") reflects the proportion of trials successfully completed up to and including that segment. SAFEMIMIC's real-world fine-tuning takes on average five minutes per navigation segment and 15 minutes per manipulation segment, and success is confirmed by prompting a VLM with the observation and semantic goal.

Fig. 4 depicts the results of our analysis. We observe that SAFEMIMIC achieves a minimum of 40% final success rate over the seven tasks, significantly outperforming all baselines. The `Direct Execution` baseline achieves 0% final success rate on all the seven tasks, demonstrating the need for exploration in order to effectively adapt the human demonstrations to the robot's morphology. Although `Direct Execution (w/o SQF)` and `Direct Execution (with SQF)` achieve similar success rates, `Direct Execution (w/o SQF)` required 82% more human interventions due to potentially unsafe actions as compared to `Direct Execution (with SQF)`, showing the importance of our learned SQFs. Interestingly, we observe that `Direct Execution (w/o SQF)` achieves higher success than `Direct Execution (with SQF)` in certain segments due to false positives of the SQFs. While the imitation-learning baselines demonstrate some successes, they fail to reliably perform the tasks, indicating that while the small amount of noisy data we generate in simulation is sufficient to train SAFEMIMIC's SQFs, training robust imitation learning policies has a higher data quantity and quality demands.

Grasp mode adaptation proved to be essential to SAFEMIMIC's success. For all but the `erase_whiteboard` task, SAFEMIMIC adapted the grasp mode for one or more segments in order to succeed. Fig. 5 depicts two such examples: for the `fill_pot` task, the robot is unable to place the pot in the sink with the human-like grasp, as the arm is predicted to collide with the edges of the sink as it attempts to place the pot. Hence, the robot backtracks to the start of the pick segment, samples and explores a top-down grasp, and successfully places the pot in the sink. Similarly, for the `store_in_drawer` task, the human-like grasp leads to joint limits being reached and so the robot explores and adapts its grasp to successfully open the drawer.

**Q2) How effectively does SAFEMIMIC reduce safety-critical failures?** To assess SAFEMIMIC's effectiveness for safe exploration, we compare the number of unsafe actions that happen during exploration with SAFEMIMIC to the baselines. A human monitors the execution of SAFEMIMIC and the baselines and flags unsafe actions. For each method, we measure the percentage of unsafe actions (`unsafe action rate`) as the ratio of all unsafe actions to the total number of actions the robot executed (over all trials for the seven tasks).

As expected, we observe that methods that do not use SQFs face the highest unsafe action rates. The `Direct Execution (without safety Q-functions)` generates 13.4% unsafe actions and incurs safety violations in nearly every task, commonly colliding during both navigation and manipulation segments or reaching force limits measured by the FT sensor. Exploration alone (`Exploration without SQF`) similarly results in 14.2% unsafe actions, demonstrating the critical need for safety during exploration

absent in current frameworks for learning from human video. Both the IL baselines, `IL (all safe actions)` and `IL (successful episodes)` observe slightly lower unsafe action rates at 10.8% and 9.5% respectively, but still not low enough for safe deployment and adaptation. The inclusion of the SQFs in (`Direct Execution with SQFs`) and SAFEMIMIC results in an unsafe action rate of 0.5 % and 0.6%, reducing the number of safety violations by 13.6%, demonstrating the efficacy of the SQFs. Fig. 6 includes examples of predictions of safe and unsafe action samples during `store_in_drawer` and `refrigerate_item`. It can be observed that actions that do not align with the drawer's kinematic constraints or that would lead to a collision during placing are correctly predicted to be unsafe by SAFEMIMIC's SQFs.

**Q3) Can SAFEMIMIC help robots learn task-specific actions to reduce exploration in future attempts?** SAFEMIMIC enables safe and autonomous learning of a demonstrated task, and the policy memory module (see Sec. III.D, Fig. 2, *bottom right*) allows the agent to associate solutions to the task for future attempts. To evaluate this capability, we applied SAFEMIMIC on several tasks and trained a policy memory module with successes. The memory module was trained to predict grasps for the pick segment of the tasks `boxing`, `shelving`, `refrigerate_item`, opening segments of tasks `refrigerate_item`, `store_in_drawer` and trained to predict the full action trajectory for the drawer opening segment of `store_in_drawer`. During evaluation, the objects to be picked and their poses were randomized, and due to the inherent stochasticity in the exploration phase of the navigation segments, the viewpoints are also variable during the evaluation (a common problem in mobile manipulation tasks). Three evaluation trials were performed for each task after initial refinement attempts during the first experiment. Fig. 2 *(bottom-right)* depicts two different viewpoints of the drawer (and the corresponding action prediction) that the robot sees after completing the navigate to drawer segment. For the segments that are not learned, we reuse the actions obtained from the human video parsing module. Fig. 7 depicts the results of our experiments. For the placing tasks, learning to predict successful grasping modes from previous adaptations dramatically reduces the number of waypoints explored during subsequent executions. In the case of `store_in_drawer`, the policy memory module enables the robot to manipulate the drawer successfully in future attempts, further reducing action exploration.

**Q4) Is SAFEMIMIC sensitive to demonstrations provided by different humans or in different environments?** To further validate the robustness of SAFEMIMIC, we conducted experiments with different users and in different environments. Three users provided demonstrations for the `shelving` task in the main environment used for previous experiments to assess whether SAFEMIMIC performed similarly across human demonstrators. Each human was free to perform the task as they naturally would, resulting in potentially different grasping strategies and trajectories, as depicted in Fig. 9. For all three demonstrators, SAFEMIMIC was able to successfully recover, parse and translate the demonstrated behavior. Interestingly, the robot had to adapt the grasping strategy for each of the three provided demonstrations, as all humans preferred a top-down grasp but such a grasp did not allow the robot to place the object in the shelf. We also tested one human demonstrator across three environments on the `shelving` task to ensure SAFEMIMIC is robust to different objects and room layouts. We performed 3 trials with the robot. SAFEMIMIC achieved 100%, 100% and 66% success in each of the three environments respectively, demonstrating the method's generalization capabilities.

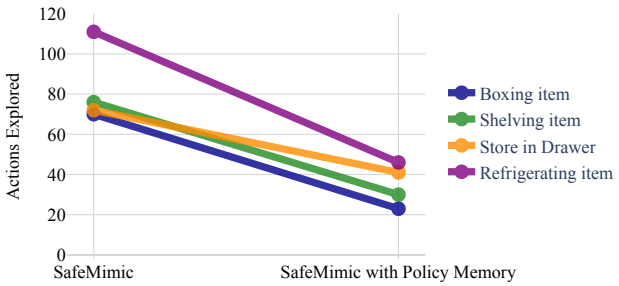


Fig. 7. **Exploration Reduction with Policy Memory.** Number of actions explored by SAFEMIMIC with (right) and without policy memory (left). Successful attempts from an initial exploration are recorded and use to train the policy memory. Object poses are varied relative to the robot before re-evaluating SAFEMIMIC with the learned policy memory on several task segments. The number of exploratory actions significantly reduces in all tasks, up to 67%, demonstrating that learning from prior successes is critical to increase efficiency in mobile manipulation tasks.

## V. LIMITATIONS AND FUTURE WORK

SAFEMIMIC enables safe, autonomous learning of multi-step mobile manipulation behaviors from third-person human video demonstrations. However, there are some limitations of the method that offer exciting avenues for future work. First, the simulated training data for the safety Q-functions must cover similar scenarios to those encountered during real-world exploration. Despite not requiring task-specific demonstrations or task successes in simulation, this does require some engineering of simulated tasks. This limitation could be alleviated by large-scale pretraining in simulation with a multitude of tasks and assets. Similarly, we study only a limited number of failure modes that we enumerate a priori. Scaling to other types of safety violations or task failures presents an opportunity for future work. Another limitation is that SAFEMIMIC relies on initial correspondence in pose between the human and robot such that actions represented as relative motions yield similar trajectories. We obtain this correspondence by estimating the initial human pose in the map frame while the robot observes, and navigating the robot there at the start of the episode. SAFEMIMIC also relies on grasp pose generation methods in order to obtain feasible grasps for the robot to explore — accurately inferring a suitable robot grasp from the demonstrated human grasp is

an open challenge. A third limitation lies in the backtracking strategy: backtracking is able to reset some aspects of the environment, but is unable to recover from irreversible events like dropping objects. Integrating autonomous resetting frameworks [27, 28, 31] is a natural opportunity for future work. Finally, an exciting avenue for future work is improving the safety Q-functions given real-world data. While SAFEMIMIC's safety Q-functions demonstrate high accuracy and a low false negative rate, learning from occasional observed failures in the real-world could increase the robustness in future learning attempts.

## VI. CONCLUSION

In this work, we present SAFEMIMIC, a framework for safe, autonomous human-to-robot imitation for mobile manipulation tasks. SAFEMIMIC effectively parses human demonstrations through a combination of pose tracking and VLM prompting, inferring both semantic segments and corresponding motions. The framework then safely refines the demonstrated behavior by sampling and verifying actions using safety Q-functions pre-trained in simulation, backtracking and exploring new actions and grasp modes when necessary. The result is a policy that adapts the demonstrated behavior to the robot's morphology, and can be associated with the task to reduce exploration in future attempts. We validate SAFEMIMIC on seven multi-step mobile manipulation tasks and find that it outperforms representative baselines in both task success and safety metrics, and is robust across different environments and human teachers. Through our experiments we show that safety Q-functions learned in simulation outperform IL policies trained in simulation due to the higher data quality and quantity demands of the latter. Taken together, our work is a promising step toward robots that can robustly learn new behaviors from human teachers in their own home environments.

## ACKNOWLEDGEMENTS

This work took place at the Robot Interactive Intelligence Lab (RobIn) at UT Austin. RobIn is supported in part by the College of Natural Sciences (CNS) Catalyst Grant (CAT-24-MartinMartin).

## REFERENCES

[1] Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, et al. Open x-embodiment: Robotic learning datasets and rt-x models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903. IEEE, 2024.

[2] Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, Lawrence Yunliang Chen, Kirsty Ellis, et al. Droid: A large-scale in-the-wild robot manipulation dataset. *arXiv preprint arXiv:2403.12945*, 2024.

[3] Frederik Ebert, Yanlai Yang, Karl Schmeckpeper, Bernadette Bucher, Georgios Georgakis, Kostas Daniilidis, Chelsea Finn, and Sergey Levine. Bridge data: Boosting generalization of robotic skills with cross-domain datasets. *arXiv preprint arXiv:2109.13396*, 2021.

[4] Vickie Ye, Georgios Pavlakos, Jitendra Malik, and Angjoo Kanazawa. Decoupling human and camera motion from videos in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023.

[5] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. Reconstructing hands in 3D with transformers. In *CVPR*, 2024.

[6] Dandan Shan, Jiaqi Geng, Michelle Shu, and David Fouhey. Understanding human hands in contact at internet scale. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[7] Yu Rong, Takaaki Shiratori, and Hanbyul Joo. Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1749–1759, 2021.

[8] Shikhar Bahl, Abhinav Gupta, and Deepak Pathak. Human-to-robot imitation in the wild. 2022.

[9] Aditya Kannan, Kenneth Shaw, Shikhar Bahl, Pragna Mannam, and Deepak Pathak. Deft: Dexterous fine-tuning for hand policies. In *Conference on Robot Learning*, pages 928–942, 2023.

[10] Kenneth Shaw, Shikhar Bahl, and Deepak Pathak. Videodex: Learning dexterity from internet videos. In *Conference on Robot Learning*, pages 654–665, 2023.

[11] Arpit Bahety, Priyanka Mandikal, Ben Abbatematteo, and Roberto Martín-Martín. ScrewMimic: Bimanual Imitation from Human Videos with Screw Space Projection. In *Robotics: Science and Systems*. Robotics: Science and Systems Foundation, 2024.

[12] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. Ego4d: Around the world in 3,000 hours of egocentric video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18995–19012, 2022.

[13] Aravind Sivakumar, Kenneth Shaw, and Deepak Pathak. Robotic telekinesis: Learning a robotic hand imitator by watching humans on youtube. In *Robotics: Science and Systems*, 2022.

[14] Haoyu Xiong, Quanzhou Li, Yun-Chun Chen, Homanga Bharadhwaj, Samarth Sinha, and Animesh Garg. Learning by watching: Physical imitation of manipulation skills from human videos. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7827–7834. IEEE, 2021.

[15] Annie S Chen, Suraj Nair, and Chelsea Finn. Learning generalizable robotic reward functions from" in-the-

wild" human videos. *arXiv preprint arXiv:2103.16817*, 2021.

[16] Lin Shao, Toki Migimatsu, Qiang Zhang, Karen Yang, and Jeannette Bohg. Concept2robot: Learning manipulation concepts from instructions and human demonstrations. *The International Journal of Robotics Research*, 40(12-14):1419–1434, 2021.

[17] Yecheng Jason Ma, William Liang, Vaidehi Som, Vikash Kumar, Amy Zhang, Osbert Bastani, and Dinesh Jayaraman. Liv: Language-image representations and rewards for robotic control. *arXiv preprint arXiv:2306.00958*, 2023.

[18] Yecheng Jason Ma, Shagun Sodhani, Dinesh Jayaraman, Osbert Bastani, Vikash Kumar, and Amy Zhang. Vip: Towards universal visual reward and representation via value-implicit pre-training. *arXiv preprint arXiv:2210.00030*, 2022.

[19] Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. R3m: A universal visual representation for robot manipulation. *arXiv preprint arXiv:2203.12601*, 2022.

[20] Ilija Radosavovic, Tete Xiao, Stephen James, Pieter Abbeel, Jitendra Malik, and Trevor Darrell. Real-world robot learning with masked visual pre-training. In *Conference on Robot Learning*, pages 416–426. PMLR, 2023.

[21] Homanga Bharadhwaj, Abhinav Gupta, Shubham Tulsiani, and Vikash Kumar. Zero-shot robot manipulation from passive human videos, 2023.

[22] Russell Mendonca, Shikhar Bahl, and Deepak Pathak. Structured world models from human videos. 2023.

[23] Jinhan Li, Yifeng Zhu, Yuqi Xie, Zhenyu Jiang, Mingyo Seo, Georgios Pavlakos, and Yuke Zhu. Okami: Teaching humanoid robots manipulation skills through single video imitation. In *Conference on Robot Learning*, 2024.

[24] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. *arXiv preprint arXiv:2402.16796*, 2024.

[25] Sungjoon Choi, Matt Pan, and Joohyung Kim. Nonparametric motion retargeting for humanoid robots on shared latent space. In *16th Robotics: Science and Systems, RSS 2020*. MIT Press Journals, 2020.

[26] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Learning human-to-humanoid real-time whole-body teleoperation. *arXiv preprint arXiv:2403.04436*, 2024.

[27] Henry Zhu, Justin Yu, Abhishek Gupta, Dhruv Shah, Kristian Hartikainen, Avi Singh, Vikash Kumar, and Sergey Levine. The ingredients of real-world robotic reinforcement learning. *arXiv preprint arXiv:2004.12570*, 2020.

[28] Russell Mendonca and Deepak Pathak. Continuously improving mobile manipulation with autonomous real-world rl. In *RSS 2024 Workshop: Data Generation for Robotics*.

[29] Archit Sharma, Ahmed M Ahmed, Rehaan Ahmad, and Chelsea Finn. Self-improving robots: End-to-end autonomous visuomotor reinforcement learning. In *Conference on Robot Learning*. PMLR, 2023.

[30] Hoai-An Nguyen and Ching-An Cheng. Provable reset-free reinforcement learning by no-regret reduction. In *International Conference on Machine Learning*, pages 25939–25955. PMLR, 2023.

[31] Homer Rich Walke, Jonathan Heewon Yang, Albert Yu, Aviral Kumar, Jedrzej Orbik, Avi Singh, and Sergey Levine. Don't start from scratch: Leveraging prior data to automate robotic reinforcement learning. In *Conference on Robot Learning*, pages 1652–1662. PMLR, 2023.

[32] Chen Tang, Ben Abbatematteo, Jiaheng Hu, Rohan Chandra, Roberto Martín-Martín, and Peter Stone. Deep reinforcement learning for robotics: A survey of real-world successes. *Annual Review of Control, Robotics, and Autonomous Systems*, 8, 2024.

[33] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

[34] Michael Laskey, Jonathan Lee, Roy Fox, Anca Dragan, and Ken Goldberg. Dart: Noise injection for robust imitation learning. In *Conference on robot learning*, pages 143–156. PMLR, 2017.

[35] Yusuf Umut Ciftci, Zeyuan Feng, and Somil Bansal. Safe-gil: Safety guided imitation learning. *arXiv preprint arXiv:2404.05249*, 2024.

[36] Huihan Liu, Shivin Dass, Roberto Martín-Martín, and Yuke Zhu. Model-based runtime monitoring with interactive imitation learning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4154–4161. IEEE, 2024.

[37] Chen Xu, Tony Khuong Nguyen, Patrick Miller, Robert Lee, Paarth Shah, Rares Andrei Ambrus, Haruki Nishimura, and Masha Itkina. Uncertainty-aware failure detection for imitation learning robot policies. In *CoRL Workshop on Safe and Robust Robot Learning for Operation in the Real World*.

[38] Cem Gokmen, Daniel Ho, and Mohi Khansari. Asking for help: Failure prediction in behavioral cloning through value approximation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5821–5828. IEEE, 2023.

[39] Alfredo Reichlin, Giovanni Luca Marchetti, Hang Yin, Ali Ghadirzadeh, and Danica Kragic. Back to the manifold: Recovering from out-of-distribution states. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[40] Kim P Wabersich, Andrew J Taylor, Jason J Choi, Koushil Sreenath, Claire J Tomlin, Aaron D Ames, and Melanie N Zeilinger. Data-driven safety filters: Hamilton-jacobi reachability, control barrier functions,

and predictive methods for uncertain systems. *IEEE Control Systems Magazine*, 43(5):137–177, 2023.

[41] Kai-Chieh Hsu, Haimin Hu, and Jaime F Fisac. The safety filter: A unified view of safety-critical control in autonomous systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 7, 2023.

[42] Yue Yang, Letian Chen, Zulfiqar Zaidi, Sanne van Waveren, Arjun Krishna, and Matthew Gombolay. Enhancing safety in learning from demonstration algorithms via control barrier function shielding. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 820–829, 2024.

[43] Josiah Wong, Albert Tung, Andrey Kurenkov, Ajay Mandlekar, Li Fei-Fei, Silvio Savarese, and Roberto Martín-Martín. Error-aware imitation learning from teleoperation data for mobile manipulation. In *Conference on Robot Learning*, pages 1367–1378. PMLR, 2022.

[44] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[45] Krishnan Srinivasan, Benjamin Eysenbach, Sehoon Ha, Jie Tan, and Chelsea Finn. Learning to be safe: Deep rl with a safety critic. *arXiv preprint arXiv:2010.14603*, 2020.

[46] Brijen Thananjeyan, Ashwin Balakrishna, Suraj Nair, Michael Luo, Krishnan Srinivasan, Minho Hwang, Joseph E Gonzalez, Julian Ibarz, Chelsea Finn, and Ken Goldberg. Recovery rl: Safe reinforcement learning with learned recovery zones. *IEEE Robotics and Automation Letters*, 6(3):4915–4922, 2021.

[47] Tsung-Yen Yang, Tingnan Zhang, Linda Luu, Sehoon Ha, Jie Tan, and Wenhao Yu. Safe reinforcement learning for legged locomotion. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2454–2461. IEEE, 2022.

[48] Puze Liu, Davide Tateo, Haitham Bou Ammar, and Jan Peters. Robot reinforcement learning on the constraint manifold. In *Conference on Robot Learning*, pages 1357–1366. PMLR, 2022.

[49] Jean-Baptiste Bouvier, Kartik Nagpal, and Negar Mehr. POLICEd RL: Learning Closed-Loop Robot Control Policies with Provable Satisfaction of Hard Constraints. In *Robotics: Science and Systems*. Robotics: Science and Systems Foundation, 2024.

[50] Gal Dalal, Krishnamurthy Dj Dvijotham, Matej Vecerík, Todd Hester, Cosmin Paduraru, and Yuval Tassa. Safe exploration in continuous action spaces. *ArXiv*, abs/1801.08757, 2018.

[51] Dongjie Yu, Haitong Ma, Shengbo Li, and Jianyu Chen. Reachability constrained reinforcement learning. In *International conference on machine learning*, pages 25636–25655. PMLR, 2022.

[52] Homanga Bharadhwaj. Auditing robot learning for safety and compliance during deployment. In *Conference on Robot Learning*, pages 1801–1806. PMLR, 2022.

[53] Bhekisipho Twala. Robot execution failure prediction using incomplete data. In *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1518–1523. IEEE, 2009.

[54] Aneseh Alvanpour, Sumit Kumar Das, Christopher Kevin Robinson, Olfa Nasraoui, and Dan Popa. Robot failure mode prediction with explainable machine learning. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pages 61–66. IEEE, 2020.

[55] Maximilian Diehl and Karinne Ramirez-Amaro. A causal-based approach to explain, predict and prevent failures in robotic tasks. *Robotics and Autonomous Systems*, 162:104376, 2023. Publisher: Elsevier.

[56] Anayo K Akametalu, Jaime F Fisac, Jeremy H Gillula, Shahab Kaynama, Melanie N Zeilinger, and Claire J Tomlin. Reachability-based safe learning with gaussian processes. In *53rd IEEE conference on decision and control*, pages 1424–1431. IEEE, 2014.

[57] Ian M Mitchell, Alexandre M Bayen, and Claire J Tomlin. A time-dependent hamilton-jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on automatic control*, 50(7):947–957, 2005.

[58] Aaron D Ames, Xiangru Xu, Jessy W Grizzle, and Paulo Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, 62(8):3861–3876, 2016.

[59] Alec Farid, David Snyder, Allen Z Ren, and Anirudha Majumdar. Failure prediction with statistical guarantees for vision-based robot control. In *Robotics: Science and Systems (RSS)*, 2022.

[60] Rachel Luo, Shengjia Zhao, Jonathan Kuck, Boris Ivanovic, Silvio Savarese, Edward Schmerling, and Marco Pavone. Sample-efficient safety assurances using conformal prediction. *The International Journal of Robotics Research*, 43(9):1409–1424, 2024.

[61] Steven M La Valle. Motion planning. *IEEE Robotics & Automation Magazine*, 18(2):108–118, 2011.

[62] Sachin Chitta, Ioan Sucan, and Steve Cousins. Moveit![ros topics]. *IEEE robotics & automation magazine*, 19(1):18–19, 2012.

[63] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[64] Andreas Ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. Grasp pose detection in point clouds. *The International Journal of Robotics Research*, 36(13-14): 1455–1473, 2017.

[65] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Michael Lingelbach, Jiankai Sun, et al. Behavior-1k: A benchmark for embodied ai with 1,000 everyday activities and realistic simulation. In

*Conference on Robot Learning*, pages 80–93. PMLR, 2023.

[66] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[67] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.

[68] Patrick Beeson and Barrett Ames. Trac-ik: An open-source library for improved solving of generic inverse kinematics. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 928–935. IEEE, 2015.

## APPENDIX

### A. Model Details

The architecture of the SQFs is a PointNet++ [66] encoder for processing point cloud input, 4-layer MLP encoder for the action and 4-layer MLP encoder for the proprioception (end-effector pose and the FT value). We categorize the continuous FT value into 6 bins for better sim-to-real transfer. Features from all the encoders are concatenated and input to an MLP head. The action space is cartesian delta positions of the end-effector with the delta actions in the range $[0.0, 0.15]$.

We define several possible task failures: arm collision, base collision, joint limit violation, force/torque threshold violation, grasp loss during articulated object manipulation, and object dropping.

During data collection, trajectories are recorded along with the safety label for each state-action pair encountered (see Fig. 10). The Q-functions are optimized using a cross-entropy loss.

TABLE I
HYPERPARAMETERS FOR SAFE EXPLORATION, SQF TRAINING, BASELINES, AND VIDEO PARSING.

| **Hyperparameter** | **Q-function** |
|---|---|
| SQF threshold $\epsilon$ | 0.7 |
| Exploration Standard Deviation $\sigma$ | 0.05 |
| Grasp Mode Clustering? | K-Mediods |
| Batch Size | 16 |
| Optimizer | Adam) |
| Learning Rate | 1e-3 |
| SQF Dataset Size | 9000 |
| IL (all safe actions) Dataset Size | 23471 |
| IL (successful episodes) Dataset Size | 4305 |
| Nav/Stationary Pos Threshold | 0.07 |
| Nav/Stationary Ori Threshold | 0.2 |

### B. Human Video Parsing

SAFEMIMIC's video parsing module extracts semantic task segments, including the semantic goal of each segment and the human actions to be adapted to the robot. The process is depicted in Fig. 8. First, SAFEMIMIC extracts human motion from the RGB-D video using the combination of body Pavlakos et al. [5] and hand Ye et al. [4] tracking. Then, SAFEMIMIC performs a coarse segmentation into navigation and stationary manipulation segments by thresholding the human hip motion.

The stationary segments are refined into segments with hand contact and without hand contact using a coarse-to-fine second segmentation. First, SAFEMIMIC queries a VLM at a low frame rate (3 fps) to determine whether the human hand is in contact with an object. We use OpenAI's GPT-4o [63] for segmentation and later labeling. To obtain a more precise initial contact frame estimation, we refine the segmentation using a visual contact detector Shan et al. [6] around the frames where the VLM detects a change in contact state.

For all segments obtained (navigation and stationary manipulation, with and without contact), SAFEMIMIC identifies and labels the semantic goal that will be used to enable

**Algorithm 1** SafeMimic Execution Pseudocode

**Require:** Segment videos $V_{1:N}$, video parser $f_{\text{parse}}$, standard deviation $\sigma$, num action samples $m$, safety Q-functions $Q_{\text{safe}}^{1:K}$, policy memory $\pi_{\text{mem}}$

**for** segment $i = 1...N$ **do**
  $\tau_{human}$, segment_type, $f_{\text{success}} \leftarrow f_{parse}(V_i)$
  **if** segment_type is grasping **then**
    $G \leftarrow$ `sample_grasp_modes()`
    **if** $\pi_{\text{mem}}$ exists for this segment **then**
      $g_{\text{robot}} \leftarrow \pi_{\text{mem}}(s_t)$
    **else**
      $g_{\text{human}} \leftarrow \tau_{\text{human}}[-1]$
      $g_{\text{robot}} \leftarrow \texttt{closest}(G, g_{\text{human}})$
    **end if**
    `execute`$(g_{\text{robot}})$
    $G$.`pop`$(g_{\text{robot}})$
  **else** ▷ Segment is nav or manipulation
    **if** $\pi_{\text{mem}}$ exists for this segment **then**
      Sample $\tau_{1:m} \sim \pi_{\text{mem}}(s_t)$
    **else**
      Sample $\tau_{1:m} \sim \mathcal{N}(\tau_{\text{human}}, \sigma)$
    **end if**
    segment_success $\leftarrow$ False
    **while** True **do**
      segment_success $\leftarrow$ `safe_explore`$(\tau_{1:m})$
      **if** segment_success **then**
        **break** ▷ Go to next segment
      **end if**
      **if** $G$ is not empty **then**
        $g_{robot} \sim G$ ▷ Explore new grasp
        `execute`$(g_{robot})$
        $G$.`pop`$(g_{robot})$
      **end if**
    **end while**
  **end if**
  **if** not segment_success **then return** False
  **end if**
**end for**
**return** segment_success
**procedure** SAFE_EXPLORE($\tau_{1:m}$)
  **if** len$(\tau_{1:m}) == 0$ **then** ▷ End of trajs
    **return** $f_{\text{success}}()$
  **end if**
  Filter $\tau_{1:m}$ if first action $Q_{\text{safe}}^{k}(s_t, a_0) < \epsilon \ \forall \ k$
  Sort $\tau_{1:m}$ by first action $\max_k Q_{\text{safe}}^{1:K}(s_t, a_0)$ ascending
  **for** each $a_0$ in $\tau_{1:m}$ **do**
    `execute`$(a_0)$
    success $\leftarrow$ `safe_explore`(`next`$(\tau_{1:m})$)
    **if** success **then return** True
    **end if**
    `execute`$(-a_0)$ ▷ Backtrack one step
  **end for**
  **return** False
**end procedure**

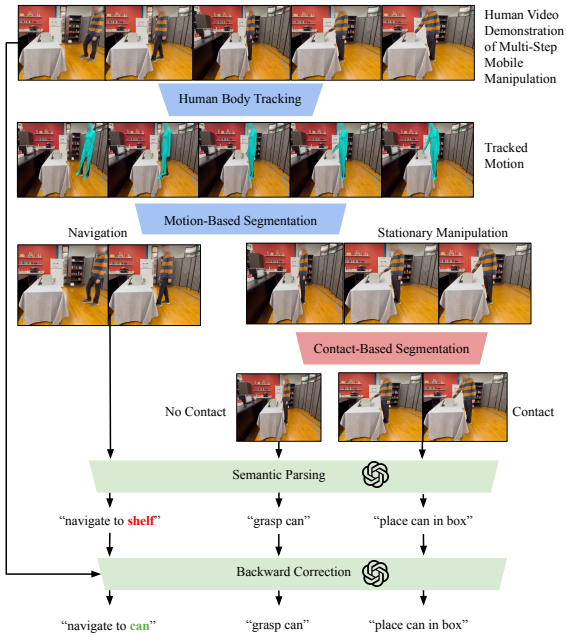


Fig. 8. Human video parsing by SAFEMIMIC. An initial RGB-D video demonstration is processed by SAFEMIMIC using a body tracking solution to obtain segments where the human is either navigating or performing stationary manipulation. Initial stationary manipulation segments are further segmented based on changes in the hand-object contact state. For all segments, the navigation and the manipulation ones, SAFEMIMIC prompts a VLM that provide a semantic description of the actions. SAFEMIMIC refines the semantic labeling results by reprompting the VLM for backward consistency between labels. The segmentation (motion and semantic labels) are used by SAFEMIMIC to refine and adapt the actions in a safe and autonomous manner.

autonomous success detection. Labeling the segments is performed with VLM queries that classify the segments into a set of possible actions and the objects the actions apply to, e.g., *navigate to a can*. Table II includes the list of actions the VLM classifies the segments on.

TABLE II
SEGMENT TYPES AND POSSIBLE ACTIONS

| Segment Type | Possible Actions |
|---|---|
| Navigation | Navigating to [object]<br>Navigating with [object$_1$] to [object$_2$] |
| Grasping | Reach for and grasp [object] |
| Manipulation | Pick [object]<br>Place [object$_1$] in [object$_2$]<br>Open [object]<br>Close [object]<br>Wipe [object$_1$] with [object$_2$] |

Since the segments are labeled independently, we observe some temporal inconsistencies in the results. We removed this effect through a backward correction method for the semantic descriptions by prompting again the VLM with all the descriptions in order, as well as frames from the entire video, requesting it to correct any inconsistencies between consecutive actions and objects. This helps significantly in situations where objects are inconsistent across descriptions, e.g., in the task of `store_in_drawer`, semantic segments

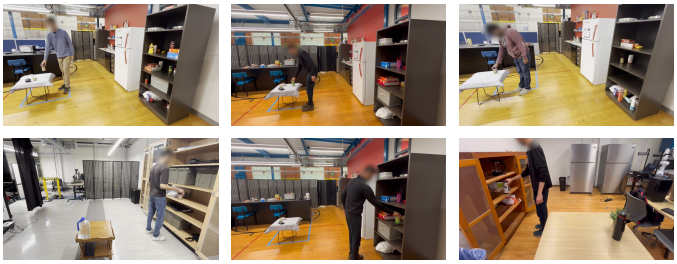

Fig. 9. Snapshots of three different human demonstrations of the same task (top), and demonstrations of the task in three different environments (bottom). We evaluate SAFEMIMIC's abilities to parse initial demonstrations of different humans, and in different environments. The performance of SAFEMIMIC is not affected (Sec. IV), indicating that our method is robust to variations in these dimensions and is able to explore and adapt actions in different conditions.


Fig. 10. Simulation training domains for the safety Q-function functions of SAFEMIMIC. We generate random and targeted actions in multiple simulated scenes in OmniGibson [65]. To increase generalization, the scenes are domain randomized using different objects and locations. We collect 43200 state-action pairs. Point clouds and robot proprioceptive signals are collected during data generation, as well as ground truth for different failure types, and used to train the ensemble of safety Q-function functions that SAFEMIMIC uses for real-world exploration. The process is cheap and scalable and does not require solving the tasks in simulation.

include *navigating to drawer* and *opening drawer*. However, the initial labeling may result in a first semantic label *"navigating to sink"* due to a sink being next to the drawer. By having access to the entire video and future labels, the VLM detects and corrects the wrong initial description.

The segments obtained in this process, including both the human motion and the semantic subgoal label, are transformed into the robot's initial motion for SAFEMIMIC to refine and adapt through safe and autonomous exploration.

### C. Robustness of the Human Video Parsing Module

We have already shown in Fig. 9 that SAFEMIMIC's video parsing module is robust to videos taken in different environments with different human demonstrators. However, to further demonstrate the robustness of the module, we conducted the following tests in challenging environments: We recorded videos of humans performing the *shelving* task from different camera angles, different lighting conditions and with "few" (1-


Fig. 11. Backtracking actions in SAFEMIMIC. Example sequence of backtracking behavior in `Shelving` task. When no safe samples are available, or when a segment exploration ends without success, SAFEMIMIC backtracks to previous states by undoing the last actions, possibly stepping back to explore a different grasping mode. This simple but effective mechanism enables autonomous exploration and adaptation of human motion for multi-step mobile manipulation tasks with SAFEMIMIC.

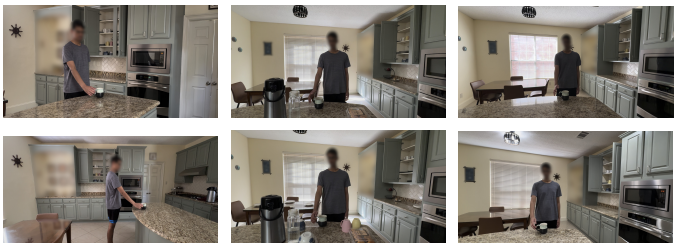

Fig. 12. Snapshots of the videos we evaluate SAFEMIMIC's video parsing module on in C. We test the module on videos with different camera angles (left column), different levels of clutter (middle column), and different lighting conditions (right column). Our results indicate that SAFEMIMIC is robust to videos in varying and challenging environments.

3), "several" (3-5), and "many" (>5) distractor objects, and measured success rate as the percent of correctly identified segments in the video. We observe **86%** success across different camera angles, **100%** across different lighting conditions and **88%** across different levels of clutter, indicating that our activity segmentation is a robust first module for SAFEMIMIC.