Learning Generalizable Task Policies for Embodied Agents in Domains with Rich Inter-object Interactions

Mustafa Chasmai, Shreshth Tuli, Mausam, Rohan Paul
Department of Computer Science and Engineering
Indian Institute of Technology Delhi

Abstract: The ability to plan in large-scale settings with several objects and high-level tasks is a challenging problem. Unlike myopic planning, handling complex application domains requires an agent to reason about the salient aspects of the environment conditioned on the task specification. Recent approaches leverage dense object embeddings using commonsense knowledgebases and neural inference to predict plans for a given goal specification and environment state. We present a neural model, namely TANGO, to learn abstractions from human demonstrations for scaling to domains with complex inter-object interactions. TANGO encodes the world state using a graph neural network with ConceptNet embeddings, and applies goal-conditioned attention to decode symbolic actions to be executed by an embodied agent. A combination of teacher-forced and end-to-end training, enables TANGO to outperform state-of-the-art baseline in both small and large scale settings, increasing the goal reaching rate by 2.2-5.8 times.

Keywords: Long-Horizon Planning, Large-Scale Settings, Imitation Learning.

1 Introduction

As robots become more capable and efficient in planning and manipulation, existing works still face limitations in large-scale settings requiring long-horizon reasoning complex task specifications. To *scale up* planning in complex scenarios, we leverage recent advances in autonomy, which have enabled robots to enter human-centric domains such as homes and factories. General purpose tasks such as transport, assembly, and clearing require a robot to interact with objects, often using them as *tools*, such as using a *tray* for carrying items and facilitates long-horizon reasoning in large-scale domains with several objects.

Learning to predict task-directed tool interactions poses several challenges. First, real environments are typically large where an expansive number of tool interactions may be possible (e.g., objects supporting others while transporting). This requires long-horizon reasoning about the world objects/tools conditioned on the salient aspects of tasks. Second, the robot may encounter new environments populated with novel objects not encountered during training. Hence, the agent's model must be able to *generalize* by reasoning about interactions with novel objects unseen during training.

We hypothesize that humans possess innate commonsense knowledge about contextual use of tools for an intended goal [1, 2]. For example, a human actor when asked to move objects is likely to use trays, boxes, or even improvise with a new object with a flat surface. Thus, in this work, we train a neural model that learns such abstractions by imitating human partners to infer goal reaching plans in unseen complex domains. We focus on generalizing planning in fully observable environment with unseen tools. We present Tool Interaction Prediction Network for Generalized Object environments (TANGO). It learns a dense representation of the object-centric graph of the environment which is augmented with word embeddings from a knowledge base, facilitating generalization to novel environments. Experiments in simulated environments and benchmark datasets with embod-

6th Conference on Robot Learning (CoRL 2022), Auckland, New Zealand.

ied agents show the effectiveness of TANGO in learning to plan multi-step plans even in domains with apriori unknown objects. This is an extension of an accepted work [3] where we take a step in the direction of the general problem of long-horizon planning in domains with complex symbolic interactions.

2 Related Work

Learning control policies for manipulating tools has received recent attention in robotics. Finn et al. [4] learn tool manipulation policies from human demonstrations. Holladay et al. [5] learn physics-based models and *effects* enabling compositional tool use. Toussaint et al. [6] present a planner to compose physics tool interactions using a logic-based symbolic planner. The aforementioned works focus on learning *how* to manipulate a tool, whereas we discuss how tool-use can facilitate planning.

Others address the problem of acquiring knowledge for completing high-level task specifications. Puig et al. [7] and Liao et al. [8] create a knowledge base of task decompositions as *action sketches* and learn to translate sketches to executable plans. These efforts rely on the causal knowledge of sequences on sub-steps required to achieve an activity which are then contextually grounded. Instead, this work learns compositional tool use required to achieve the task without any causal sequence as input. Huang et al. [9] learn task decompositions from human demonstration videos. However, the work does not explicitly model the physical constraints of the robot and does not generalize to new environments.

Recently, agents leverage pre-trained large-language models (LLMs) to ground instructions to actionable steps [10, 11, 12]. Without any human input on instruction following, these methods do not incorporate human preferences, which could possibly lead to sub-optimal plans. We adopt a framework in which only the natural demonstrations are provided to a learner having access to the full environment state in lieu of explicit knowledge of predicates corresponding to input instructions.

3 Problem Formulation

We consider a mobile agent with a single manipulator arm operating in a fully observable environment. Each object has one of possible symbolic states such as $\mathrm{Open/Closed}$, $\mathrm{On/Off}$ etc. We also consider *relations* between two objects. Let s denote the world state that maintains object states, class type and object relations as OnTop, Near, Inside and $\mathrm{ConnectedTo}$. We denote the state at the t-th timestep as s_t, with s_0 denoting the initial state. We denote the set of object instances $O = \mathcal{O}(s)$ populating state s. Let A denote the robot's symbolic action space. An action $a \in A$ is abstracted as $I(o^1, o^2)$, with an action type predicate $I \in \mathcal{I}$ that affects the states of objects $o^1 \in O$ and $o^2 \in O$, for instance, $\mathrm{Move}(\mathrm{fruit}_0, \mathrm{tray}_0)$.

We assume the presence of a simulator that can realize symbolic actions using a low-level planner. Let $\mathcal{T}(\cdot)$ denote the transition function. The successor state s_{t+1} upon taking the action a_t in state s_t is sampled from a physics simulator. Let $\eta_t = \{a_0, a_1, \ldots, a_{t-1}\}$ denote the *action history* till time t. The robot is instructed by providing a *declarative* goal g expressing the symbolic constraint between world objects. For example, the *declarative* goal, "place milk in fridge" can be expressed as a constraint $\mathrm{Inside}(\mathrm{milk}_0, \mathrm{fridge}_0)$ between specific object instances.

We aim at learning a policy $f_\theta(\cdot)$ that estimates the next action a_t conditioned on the the goal g and the initial state s (including the action history η_t) such that the robot's *goal-reaching* likelihood is maximized. See prior work [3, 13] for more details. Let $\mathcal{G}(s, g)$ denote the *goal check* function that determines if the intended goal g is achieved by a state s. Let $S_t^{f_\theta, s}$ be a random variable denoting the state resulting from executing actions from f_θ for start state s_0 for t time steps. Formally, the objective is formulated as

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & p(\mathcal{G}(S_k^{f_\theta, s_0}, g) = 1) \\ \text{s. t.} \quad & \forall\, t, a_t = f_\theta(s_t, g, \eta_t), \\ & \forall\, t, s_{t+1} = \mathcal{T}(s_t, a_t). \end{aligned} \tag{1}$$

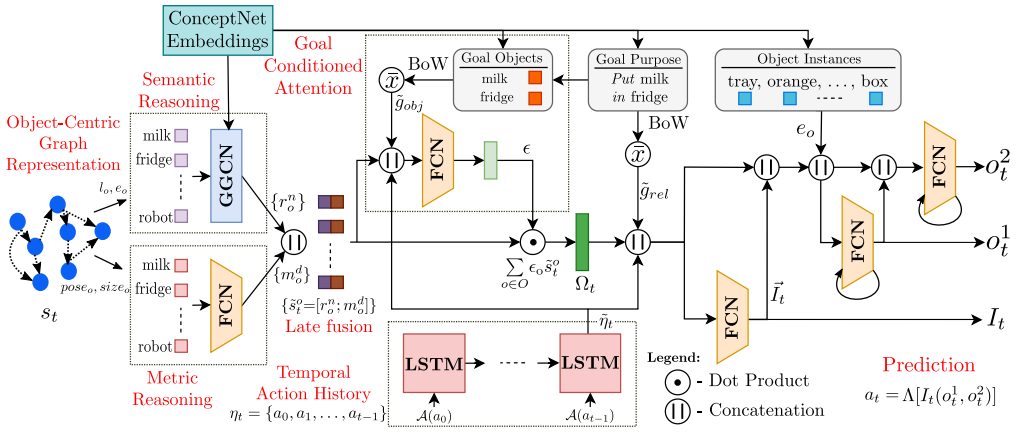


Figure 1: The TANGO model. It uses graph convolutions on an object-centric world representation and late fusion with metric encodings with goal-conditioned attention and action history to predict appropriate actions.

4 Technical Approach

TANGO learns to predict the next robot action a_t, given the world state s_t, the goal g and the action history η_t. TANGO is realized as a neural network model f_θ as follows:

$$a_t = f_\theta\left(s_t, g, \eta_t\right) = f_\theta^{act}\left(f_\theta^{goal}\left(f_\theta^{state}\left(s_t\right), g, f_\theta^{hist}\left(\eta_t\right)\right)\right)$$

Graph Structured World Encoding. We consider an object centric world encoded as a graph with objects as nodes and relations as edges. For an input state s_t, each object o has l_o that represents the one-hot encoding of the discrete object state, $\mathcal{C}(o)$ is the object embedding (we use ConceptNet vectors as per prior work [3]). We then use Gated Graph Convolution Network (GGCN) [8, 7], which gives

$$\{r_o^n\}_{o\in\mathcal{O}(s_t)} = \text{GGCN}(s_t), \tag{2}$$

where n is a hyperparameter that denotes the number of convolutions performed by the GGCN network. We then fuse the metric information of objects (if available) using a Fully Connected Network (FCN) with a Parameterized ReLU (PReLU) [14] activation as:

$$\{m_o^d\}_{o\in\mathcal{O}(s_t)} = \text{FCN}(s_t), \tilde{s}_t = \{\tilde{s}_t^o = [r_o^n; m_o^d]\}_{o\in\mathcal{O}(s_t)} = f_\theta^{state}(s_t). \tag{3}$$

Encoding Action History. The sequence of previously executed actions facilitate in providing an *intent* for predicting the next action. We define action encoding $\mathcal{A}(a_{t-1})$ of $a_{t-1} = I_t(o_{t-1}^1, o_{t-1}^2)$ as $[\vec{I}_{t-1}; \mathcal{C}(o_{t-1}^1); \mathcal{C}(o_{t-1}^2)]$, where $\vec{I}_{t-1}$ is a one-hot vector over possible interaction types $\mathcal{I}$. We use an LSTM to encode the action history η_t.

$$\tilde{\eta}_t = \text{LSTM}(\mathcal{A}(a_{t-1}), \tilde{\eta}_{t-1}) = f_\theta^{hist}(\eta_t). \tag{4}$$

Learning Goal-Conditioned Contexts. An input goal g is partitioned as relations g_{rel} and the object instances specified in the goal g_{obj}. The resulting encodings are denoted as $\tilde{g}_{rel}$ and $\tilde{g}_{obj}$:

$$\tilde{g}_{rel} = \frac{1}{|g_{rel}|}\sum_{j\in g_{rel}} \mathcal{C}(j) \;\text{ and }\; \tilde{g}_{obj} = \frac{1}{|g_{obj}|}\sum_{o\in g_{obj}} \mathcal{C}(o). \tag{5}$$

We then learn attention weights over objects in the environment [15]. This results in the attended scene encoding Ω_t as:

$$\Omega_t = \sum_{o\in O} \epsilon_o \tilde{s}_t^o = f_\theta^{goal}(\tilde{s}_t, g, \tilde{\eta}_t) \text{ where, } \epsilon_o = \text{softmax}\left(W_g[\tilde{s}_t^o; \tilde{g}_{obj}; \tilde{\eta}_t] + b_g\right). \tag{6}$$

The attention mechanism aligns the goal information with the scene for long-horizon reasoning in large-scale environments.

Action Prediction. TANGO takes the encoded information about the world state, goal and action history to auto-regressively decode the next symbolic action $a_t = I_t(o_t^1, o_t^2)$. The resulting factored likelihood allows the model to generalize to an *a-priori* unknown number and types of object instances:

$$I_t = \text{argmax}_{I \in \mathcal{I}} \left(\text{softmax}(W_I[\Omega_t; \tilde{g}_{rel}; \tilde{\eta}_t] + b_I)\right), \tag{7}$$

$$o_t^1 = \text{argmax}_{o \in O} \alpha_t^o = \text{argmax}_{o \in O} \left(\sigma(W_\alpha[\Omega_t; \tilde{g}_{rel}; \tilde{\eta}_t; e_o; \vec{I}_t] + b_\alpha)\right), \tag{8}$$

$$o_t^2 = \text{argmax}_{o \in O} \left(\sigma(W_\beta[\Omega_t; \tilde{g}_{rel}; \tilde{\eta}_t; e_o; \vec{I}_t; \alpha_t^o] + b_\beta)\right). \tag{9}$$

Here α_t^o denotes the likelihood prediction of the first object. Finally, we impose grammar constraints (denoted as Λ) at inference time based on the number of arguments that the predicted interaction I_t accepts. Thus, predicted action, $a_t = f_\theta^{act}(\Omega_t, \tilde{g}_{rel}, \tilde{\eta}_t) = \Lambda[I_t(o_t^1, o_t^2)]$, is then executed by the robot in simulation.

Dataset and Training. We adopt imitation learning approach and learn the function $f_\theta(.)$ from demonstrations by human teachers. We denote a dataset of N goal-reaching plans by

$$\mathcal{D}_{\text{Train}} = \{(s_0^i, g^i, \{s_j^i, a_j^i\}) \mid i \in \{1, N\}, j \in \{0, t_i - 1\}\},$$

where the i^{th} datum consists of the initial state, the goal and a state-action sequence. The model is trained using the Binary Cross-Entropy loss in a teacher forced manner [16]. We use a datapoint with 0.8 probability of teacher-forcing and 0.2 for simulator based next state input. Online, the robot uses the learned model to sequentially predict actions and execute in the simulation environment till the goal state or a large preset plan length is attained.

5 Experimental Results

Our experiments investigate the ability of the model to learn goal reaching policies in two domains with increasing length of plans evaluated against a large number of metrics. Further, we evaluate generalization to settings with unseen objects, tools and goals.

Datasets. We use two datasets of human demonstrations. First, by Tuli et al. [3], which has been collected on a PyBullet based physics simulator [17] with a Universal Robotics 5 arm mounted on a Husky mobile base. This dataset consists of both object states and pose information. We use 75:25 split for training and testing with 10% of the former as a validation set. This dataset also consists of generalization test sets that have position perturbations, tool and object replacements. Second, by Puig et al. [7] that has been collected on the Virtual-Home symbolic simulator with human-like agents having two arms for manipulation. We use 80:10:10 train, validation and test split. We also generate splits to test generalization where we split the data based on objects and goals.

Metrics. The performance of the models are evaluated as the ability of the agent to reach goal states using their plans. Specifically, we compare the relations that have been established in the final state s_T by the model and s_{t_i-1} ground-truth, compared to the initial state s_0. Following She and Chai [18], we use the following metrics:

- *SJI (State Jaccard Index)* is the overlap between the established and ground-truth relation sets.
- *IED (Instruction Edit Distance):* similarity between the generated and ground-truth action sequences.
- *GRR (Goal Reaching Rate):* if the ground-truth relations are present in the established ones.
- *P, R, F1:* are the average of the precision, recall and F1 scores between the established relations compared to the ground-truth sets.

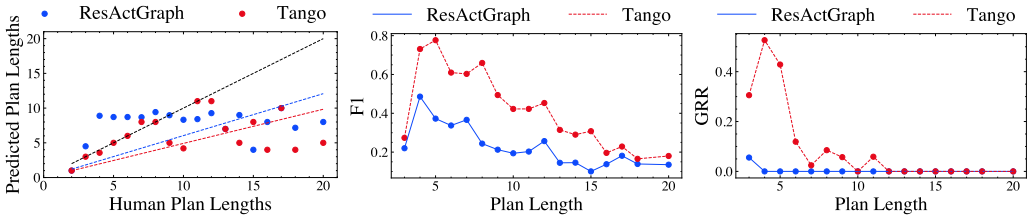
Results. We compare TANGO against the RESACTGRAPH model [8], augmented with FastText embeddings [19]. Table 1 demonstrates the results of baseline and TANGO on the dataset by Tuli et al. [3]. TANGO improves GRR, IED, SJI and F1 by at least 3, 3, 2 and 3 points, respectively. Table 2 shows results on the dataset by Puig et al. [7]. TANGO improves the scores by at least 15, 41,

Table 1: Performance on the dataset by Tuli et al. [3]. We show results of random splits on (1) Home and (2) Factory domains, (3) perturbing object positions, (4) replacing most used tool, (5) replacing environment object, and (6) changing goal. The last four splits demonstrate the ability to generalize.

Method	GRR	IED	SJI	P	R	F1	GRR	IED	SJI	P	R	F1
	Home						Factory					
RESACTGRAPH	83.08	**38.21**	16.26	26.26	49.06	34.21	22.08	39.72	10.32	25.00	55.00	34.38
TANGO	**87.69**	35.65	**28.35**	**43.44**	**100.00**	**60.57**	**72.73**	**42.98**	**17.05**	**29.53**	**95.00**	**45.06**
	Position Generalization						Alternate Generalization					
RESACTGRAPH	69.76	-	37.30	45.14	51.92	45.72	61.02	-	**29.06**	**36.6**	**36.72**	**35.82**
TANGO	**91.76**	-	**51.72**	**53.17**	**63.36**	**57.06**	**76.06**	-	8.28	8.01	16.06	10.69
	Unseen Generalization						Goal Generalization					
RESACTGRAPH	61.02	-	29.06	36.6	36.72	35.82	38.88	-	10.64	11.81	25.25	16.09
TANGO	**80.56**	-	**37.58**	**36.81**	**48.88**	**40.94**	**58.11**	-	**19.22**	**19.84**	**30.86**	**24.15**

Table 2: Performance on Virtual Home dataset by Puig et al. [7]. We use four dataset splits. (1) Uniformly at random for testing, (2) based on the objects in the environment, (3) based on the objects in the goal specification, and (4) goals. The last three splits demonstrate the ability to generalize in unseen settings.

Method	GRR	IED	SJI	P	R	F1	GRR	IED	SJI	P	R	F1
	Random Split						Environment Objects					
RESACTGRAPH	8.31	12.63	30.81	**81.03**	24.36	36.92	4.35	15.36	25.97	78.11	22.67	34.72
TANGO	**27.87**	**64.65**	**46.92**	77.28	**41.82**	**53.89**	**19.83**	**56.17**	**39.96**	**82.54**	**34.04**	**46.12**
	Goal Objects						Instructions					
RESACTGRAPH	8.00	16.97	28.21	85.71	23.84	37.26	3.46	11.90	17.70	**83.87**	10.56	18.52
TANGO	**25.66**	**65.04**	**47.21**	**80.18**	**43.48**	**56.02**	**23.69**	**63.03**	**46.14**	82.77	**41.21**	**54.57**



(a) Comparison of plan lengths. (b) F1 score with plan length. (c) GRR with plan length.

Figure 2: Analysis of predicted plans. TANGO predicts shorter plans and outperforms RESACTGRAPH in long-horizon settings.

14 and 11 points, respectively. Note that the RESACTGRAPH model learns a scene representation assuming a fixed and known set of object types and hence can only generalize to new randomized scenes of known objects. In contrast, the TANGO model can not only generalize to randomized scenes with known object types (sharing the GGCN backbone with RESACTGRAPH) but can to novel scenes new object types (relying on dense semantic embeddings) and an a-priori unknown number of instances (enabled by a factored likelihood). The improvements in the scores in unseen settings demonstrate TANGO's ability to generalize. Specifically, in the home domain, if the *tray* is not present in the scene, the model is able to use a *box* instead to transport multiple objects (see Figure 3). Similarly, for Virtual-Home, in the case of having the instruction "drink water", the model uses a *water glass* when it has not seen the same at training time (see Figure 4). The RESACTGRAPH model is unable to adapt to novel worlds and obtains zero points in several generalization tests. Another example of generalization is the unseen task of "wash dishes" in the Virtual-Home domain, where TANGO predicts that it needs to use the *dishwasher* using the commonsense knowledge encapsulated in the object embeddings (see Figure 5).

(a) Open Cupboard (b) Put fruits in box (c) Pick box (d) Put box in cupboard

Figure 3: Predicted plan for the task "put fruits in cupboard". Model had not seen *box* during training.

(a) Walk to dining-room (b) Find fridge (c) Open fridge (d) Grab waterglass

Figure 4: Predicted plan for the task "drink water". Model had not seen *water glass* during training.

(a) Walk to dining-room (b) Put back dish-soap (c) Close dishwasher (d) Switch-on washer

Figure 5: Predicted plan for the task "wash dishes". Model had not seen such tasks in training.

(a) Walk to dining-room (b) Walk to plate (c) Find dish-soap (d) Pour dish-soap on plate

Figure 6: Predicted plan for the task "wash dishes by hand". Execution fails since the plate is not reachable.

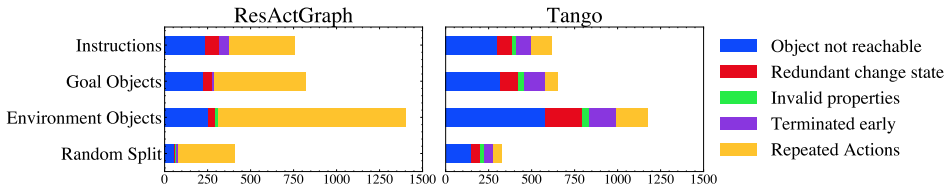


Figure 7: An analysis of the types of errors during plan execution

Analysis of Predicted Plans. Figure 2a compares the length of robot plans predicted by the learned model against human demonstrated plans for the Virtual-Home domain (for the other domain see prior work [3]). We observe that, on average, the predicted plan lengths are close to the human demonstrated ones for short-horizon plans. For long-horizon plans, however, TANGO is able to reach goals in fewer steps. Specifically, in 98% cases, the plans predicted by TANGO utilize tools satisfying the goal condition in fewer steps compared to the human demonstrated plan. This is primarily due to the ability of the model to extract patterns in plan fragments and use the one

that can reach the goal given the previous history. Figures 2b and 2c shows the performance of the model with the plan length. As expected, as the task becomes more complex, requiring manipulating several objects, *i.e.*, in long-horizon settings the performance drops. However, compared to the RESACTGRAPH, TANGO obtains higher scores. This is due to the goal-conditioned attention that allows RESACTGRAPH to focus on a small set of objects and not the entire domain, facilitating long-range reasoning.

Insights. The significant drop in performance as we increase plan length merits investigation into planning abstractions [20]. To identify possible future directions, we show the distribution of errors in Figure 7 for the Virtual-Home domain. In 80.84% cases RESACTGRAPH repeats the previous action and gets stuck in a loop. This is much lower for TANGO (15.76%). This shows the ability of TANGO to respect action pre-conditions and utilize history encoding to capture intent. A common problem in our plans is the model failing to learn certain pre-conditions, such as trying to pick up an object that is is not in reach (see Figure 6).

6 Conclusion

This paper proposes TANGO, that demonstrates accurate commonsense generalization to environments with novel object instances using the learned knowledge of shared spatial and semantic characteristics. We find that encoding the sequence of past actions enables the model to uncover correlated action sequences, such as moving close to an object before grabbing it. Further, we learn affordances like abstractions, such as correlating tools with tasks as well as goal-specific object contexts. Planning with such knowledge is likely to be crucial to scale to realistic environments (homes, factories, etc.) where the space of possible interactions may be large. Finally, we find evidence that learning dense representations for objects and scenes is key to generalisation to novel settings that the robot may not have encountered previously. Finally, we see further scope of improvement by exploring ways to recover causal relationships (PDDL-style rules) [21], discovering independent sub-goals in a task [22] and ways to estimate execution costs for tasks as mechanisms to further advance planning in such domains.

Acknowledgments

Mausam is supported by an IBM SUR award, grants by Google, Bloomberg and 1MG, Jai Gupta chair fellowship, and a Visvesvaraya faculty award by Govt. of India. Rohan Paul acknowledges support from Pankaj Gupta Faculty Fellowship. We thank the IIT Delhi HPC facility and Prof. Prem Kalra and Mr. Anil Sharma at the CSE VR Lab for compute resources.

References

[1] K. R. Allen, K. A. Smith, and J. B. Tenenbaum. The tools challenge: Rapid trial-and-error learning in physical problem solving. *arXiv preprint arXiv:1907.09620*, 2019.

[2] S. Tuli, R. Bansal, R. Paul, and Mausam. Tooltango: Common sense generalization in predicting sequential tool interactions for robot plan synthesis. *JAIR*, 2022.

[3] S. Tuli, R. Bansal, R. Paul, and Mausam. Tango: Commonsense generalization in predicting tool interactions for mobile manipulators. *International Joint Conference on Artificial Intelligence*, 2021.

[4] C. Finn, T. Yu, T. Zhang, P. Abbeel, and S. Levine. One-shot visual imitation learning via meta-learning. In *Conference on robot learning*, pages 357–368. PMLR, 2017.

[5] R. Holladay, T. Lozano-Pérez, and A. Rodriguez. Force-and-motion constrained planning for tool use. In *IROS*, 2019.

[6] M. Toussaint, K. Allen, K. Smith, and J. Tenenbaum. Differentiable physics and stable modes for tool-use and manipulation planning. In *RSS*, 2018.

[7] X. Puig, K. Ra, M. Boben, J. Li, T. Wang, S. Fidler, and A. Torralba. Virtualhome: Simulating household activities via programs. In *CVPR*, 2018.

[8] Y.-H. Liao, X. Puig, M. Boben, A. Torralba, and S. Fidler. Synthesizing environment-aware activities via activity sketches. In *CVPR*, pages 6291–6299, 2019.

[9] D.-A. Huang, S. Nair, D. Xu, Y. Zhu, A. Garg, L. Fei-Fei, S. Savarese, and J. C. Niebles. Neural task graphs: Generalizing to unseen tasks from a single video demonstration. In *CVPR*, pages 8565–8574, 2019.

[10] W. Huang, P. Abbeel, D. Pathak, and I. Mordatch. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. *International Conference on Machine Learning*, 2022.

[11] W. Huang, F. Xia, T. Xiao, H. Chan, J. Liang, P. Florence, A. Zeng, J. Tompson, I. Mordatch, Y. Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. In *Conference on Robot Learning*. PMLR, 2022.

[12] I. Singh, V. Blukis, A. Mousavian, A. Goyal, D. Xu, J. Tremblay, D. Fox, J. Thomason, and A. Garg. Progprompt: Generating situated robot task plans using large language models. In *Second Workshop on Language and Reinforcement Learning*.

[13] Mausam and A. Kolobov. Planning with Markov decision processes: An AI perspective. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–210, 2012.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[15] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[16] N. B. Toomarian and J. Barhen. Learning a trajectory using adjoint functions and teacher forcing. *Neural networks*, 5(3):473–484, 1992.

[17] E. Coumans and Y. Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning. *GitHub repo*, 2016.

[18] L. She and J. Chai. Incremental acquisition of verb hypothesis space towards physical world interaction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 108–117, 2016.

[19] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin. Advances in pre-training distributed word representations. In *LREC*, 2018.

[20] W. Vega-Brown and N. Roy. Asymptotically optimal planning under piecewise-analytic constraints. In *Algorithmic Foundations of Robotics XII*, pages 528–543. Springer, 2020.

[21] S. Nair, Y. Zhu, S. Savarese, and L. Fei-Fei. Causal induction from visual observations for goal directed tasks. *arXiv:1910.01751*, 2019.

[22] S. Sharma, J. Gupta, S. Tuli, R. Paul, and Mausam. Goalnet: Inferring conjunctive goal predicates from human plan demonstrations for robot instruction following. *International Conference on Automated Planning and Scheduling (ICAPS) - Planning and Reinforcement Learning Workshop*, 2022.