Taskography: *Evaluating robot task planning over large 3D scene graphs*

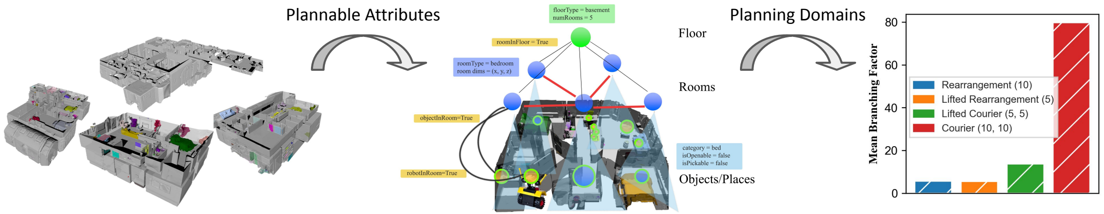
Christopher Agia [1*], Krishna Murthy Jatavallabhula [2*], Mohamed Khodeir [1]
Ondrej Miksik [3], Mustafa Mukadam [4], Vibhav Vineet [3], Liam Paull [2], Florian Shkurti [1, 5]

[1]University of Toronto, [2]Montreal Robotics and Embodied AI Lab, Mila, Université de Montréal, [3]Microsoft, [4]Facebook AI Research, [5]Vector Institute

- **Taskography** is a robotic task planning benchmark grounded in large-scale 3D scene graphs from real-world distributions
- The current **taxonomy of symbolic task planners** consists of 18 classical and learning-based methods - paradigms include:

`Optimal/satisficing propositional planners, SAT solvers, tree search, regression planning, relational RL policies, state space sparsifiers`

- By open-sourcing **Taskography-API**, we enable researchers to create novel planning domains and sample datasets of tasks
- We hope our works spurs ongoing research along the lines of multi-task robot planning and spatial scene representations



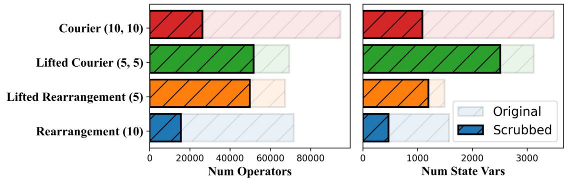
Taskography-API: Generate novel task planning domains over building-scale 3D scene graphs.

In Taskography-API, we provide interfaces for easy specification of customized symbolic planning domains and problems over 3D scene graphs. We also provide a python interface to access all classical and learning-based planners in the benchmark, enabling easy testing and integration of newer, learning-based planning solutions. For more details, please visit https://taskography.github.io

	Planner	Rearr(1) Tiny Len.	Rearr(1) Tiny Time	Rearr(1) Tiny Fail	Rearr(2) Tiny Len.	Rearr(2) Tiny Time	Rearr(2) Tiny Fail	Rearr(10) Medium Len.	Rearr(10) Medium Time	Rearr(10) Medium Fail	Cour(7, 10) Medium Len.	Cour(7, 10) Medium Time	Cour(7, 10) Medium Fail	Lifted Rearr(5) Tiny Len.	Lifted Rearr(5) Tiny Time	Lifted Rearr(5) Tiny Fail	Lifted Cour(5, 5) Tiny Len.	Lifted Cour(5, 5) Tiny Time	Lifted Cour(5, 5) Tiny Fail
optimal	**FD-seq-opt-lmcut**	15.77	24.81	0.04	**25.80**	104.47	0.55	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
	SatPlan	14.77	10.35	0.45	26.67	3.27	0.67	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
	Delfi	15.13	0.36	0.16	29.10	27.77	0.29	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
	DecStar-opt-fb	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
	MCTS	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
satisficing	**FF**	16.71	0.19	**0.00**	34.44	0.55	**0.00**	159.04	5.30	0.09	128.41	6.62	0.24	62.86	3.40	0.47	**57.74**	4.03	0.44
	FF-X	16.71	0.25	**0.00**	34.44	0.58	**0.00**	159.80	5.02	0.08	128.19	6.72	0.24	67.88	3.48	0.89	61.19	7.56	0.77
	FD-lama-first	15.19	2.96	0.33	38.47	3.25	0.18	208.28	6.35	0.49	156.34	4.92	0.29	66.81	3.20	0.49	61.13	3.34	0.56
	Cerberus-sat	**11.50**	12.00	0.85	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
	Cerberus-agl	14.77	5.13	0.45	33.00	7.30	0.49	176.60	8.91	0.72	**125.73**	12.99	0.83	60.50	7.62	0.60	59.19	7.05	0.77
	DecStar-agl-fb	14.72	2.62	0.55	34.96	2.58	0.58	211.16	7.20	0.82	132.60	4.50	0.58	66.30	3.02	0.71	58.75	4.46	0.71
	BFWS	15.56	0.90	0.22	32.16	0.37	0.18	**151.17**	0.41	0.23	152.71	1.13	0.21	**56.90**	0.94	0.41	61.92	2.30	0.43
	Regression-plan	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
learn	**Relational policy [40]**	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00	-	-	1.00
	PLOI [12]	16.45	**0.00***	**0.00**	37.04	**0.00***	**0.00**	213.43	**0.17**	**0.00**	161.90	**0.34**	**0.00**	78.68	**0.22**	**0.24**	71.71	**0.26**	**0.26**

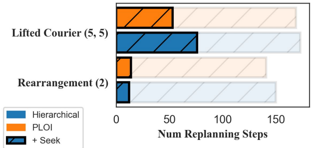
Taskography: Benchmark on selected grounded and lifted Rearrangement and Courier planning domains.

Rearrangement domains require the agent to deliver *k items* to *k target* locations; *Courier domains* further equip the agent with a knapsack to stow objects as it traverses the scene. (Red) Optimal planners work only the simplest of domains, failing to scale with increasing task complexity. (Orange) Satisficing planners degrade in domains requiring long-horizon reasoning. (Green) Learning-based planners that prune the state space excel on all domains.



SCRUB: A procedure for pruning 3D scene graphs via task-conditioning

SCRUB is a preprocessing algorithm tailored to 3D scene graphs that drastically reduces the state space of planning problems, while taking just a few milliseconds.



SEEK: Reachability of task relevant objects.

Performant learning-based task planners employ NNs to score task relevant objects prior to planning. Augmenting them with SEEK substantially reduces replanning iterations by ensuring that all *important* objects are reachable for the robot.

UNIVERSITY OF TORONTO · Université de Montréal · Microsoft Research · facebook artificial intelligence research · Mila · VECTOR INSTITUTE

Taskography: *Evaluating robot task planning over large 3D scene graphs*

*Christopher Agia**, *Krishna Murthy Jatavallabhula**, Mohamed Khodeir, Ondrej Miksik, Vibhav Vineet, Mustafa Mukadam, Liam Paull, Florian Shkurti

University of Toronto | Université de Montréal | Microsoft Research | facebook artificial intelligence research | Mila | Vector Institute

*Equal contribution