ROBUSTNAV: Towards Benchmarking Robustness in Embodied Navigation

Prithvijit Chattopadhyay[1,2*] Judy Hoffman[1] Roozbeh Mottaghi[2,3] Aniruddha Kembhavi[2,3]

[1]Georgia Tech [2]PRIOR @ Allen Institute of AI [3]University of Washington

{prithvijit3,judy}@gatech.edu {roozbehm,anik}@allenai.org

prior.allenai.org/projects/robustnav

Abstract

As an attempt towards assessing the robustness of embodied navigation agents, we propose **ROBUSTNAV**, *a framework to quantify the performance of embodied navigation agents when exposed to a wide variety of visual – affecting RGB inputs – and dynamics – affecting transition dynamics – corruptions. Most recent efforts in visual navigation have typically focused on generalizing to novel target environments with similar appearance and dynamics characteristics. With* ROBUSTNAV, *we find that some standard embodied navigation agents significantly underperform (or fail) in the presence of visual or dynamics corruptions. We systematically analyze the kind of idiosyncrasies that emerge in the behavior of such agents when operating under corruptions. Finally, for visual corruptions in* ROBUSTNAV, *we show that while standard techniques to improve robustness such as data-augmentation and self-supervised adaptation offer some zero-shot resistance and improvements in navigation performance, there is still a long way to go in terms of recovering lost performance relative to clean "non-corrupt" settings, warranting more research in this direction. Our code is available at* https://github.com/allenai/robustnav.

1. Introduction

A longstanding goal of the artificial intelligence community has been to develop algorithms for embodied agents that are capable of reasoning about rich perceptual information and thereby accomplishing tasks by navigating in and interacting with their environments. In addition to being able to exhibit these capabilities, it is equally important that such embodied agents are able to do so in a robust and generalizable manner.

A major challenge in Embodied AI is to ensure that agents can generalize to environments with different *appearance statistics* and *motion dynamics* than the environment used for training those agents. For instance, an agent

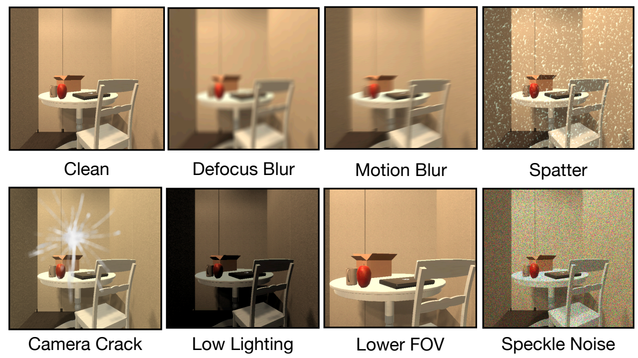

Figure 1. **Visual Corruptions.** Visual corruptions ROBUSTNAV supports in the unseen target environments. Top-left shows a clean RGB frame and rest show corrupted versions of the same. Defocus Blur, Motion Blur, Spatter, Low lighting and Speckle Noise are supported at 5 progressively increasing levels of severity.

that is trained to navigate in "sunny" weather should continue to operate in rain despite the drastic changes in the appearance, and an agent that is trained to move on carpet should decidedly navigate when on a hardwood floor despite the discrepancy in friction. While a potential solution may be to calibrate the agent for a specific target environment, it is not a scalable one since there can be enormous varieties of unseen environments and situations. A more robust, efficient and scalable solution is to equip agents with the ability to autonomously adapt to new situations by interaction without having to train for every possible target scenario. Despite the remarkable progress in Embodied AI, especially in embodied navigation [14, 11, 12, 13, 3], most efforts focus on generalizing trained agents to unseen environments, but critically assume similar appearance and dynamics attributes across train and test environments.

As a first step towards assessing general purpose robustness of embodied agents, we propose ROBUSTNAV, a framework to quantify the performance of embodied navigation agents when exposed to a wide variety of common visual (`vis`) and dynamics (`dyn`) corruptions – artifacts that affect the egocentric RGB observations (see Fig. 1; akin to [8] for object recognition) and transition dynamics (see

[*]Part of the work done when PC was a research intern at AI2.

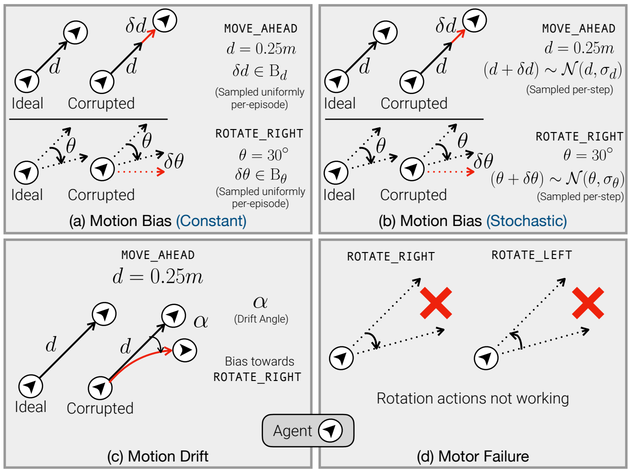

Figure 2. **Dynamics Corruptions.** We show the kinds of dynamics corruptions supported in ROBUSTNAV. Motion Bias (C & S) are modeled to mimic friction. Motion Drift models a setting where translation actions have a slight bias towards rotating right (or left). In Motor Failure, the one of the rotation actions fail.

				PointNav			
				RGB		RGB-D	
#	Corruption ↓	V	D	SR ↑	SPL ↑	SR ↑	SPL ↑
1	Clean			98.82	83.13	98.54	84.60
2	Low Lighting	✓		94.36	75.15	99.45	84.97
3	Motion Blur	✓		95.72	73.37	99.36	85.36
4	Camera Crack	✓		82.07	63.83	95.72	81.21
5	Defocus Blur	✓		75.89	53.55	99.09	85.54
6	Speckle Noise	✓		67.42	48.57	98.73	84.66
7	Lower-FOV	✓		42.49	31.73	89.08	73.59
8	Spatter	✓		33.58	24.72	98.91	84.81
9	Motion Bias (C)		✓	92.81	77.83	93.36	79.46
10	Motion Bias (S)		✓	94.72	76.95	96.72	79.08
11	Motion Drift		✓	95.72	76.19	93.36	75.08
12	PyRobot [10] (ILQR) Mul. = 1.0		✓	96.00	67.79	95.45	69.27
13	Motor Failure		✓	20.56	17.63	20.56	17.62
14	Defocus Blur + Motion Bias (S)	✓	✓	76.52	51.08	97.18	79.46
15	Speckle Noise + Motion Bias (S)	✓	✓	62.69	43.31	95.81	78.27
16	Spatter + Motion Bias (S)	✓	✓	33.30	23.33	95.81	78.85
17	Defocus Blur + Motion Drift	✓	✓	74.25	50.99	95.54	76.66
18	Speckle Noise + Motion Drift	✓	✓	64.42	44.73	94.36	75.23
19	Spatter + Motion Drift	✓	✓	32.94	23.44	95.45	76.61

Table 1. **POINTNAV Performance.** Degradation in task performance of pretrained POINTNAV (trained for $\sim$ 75M frames) agents when evaluated under `vis` and `dyn` corruptions present in ROBUSTNAV. For `vis` corruptions with controllable severity levels, we report results with severity set to 5 (worst). Rows are sorted based on SPL values for RGB POINTNAV agents. Success and SPL values are reported as percentages. (V = `vis`, D = `dyn`)

Fig. 2), respectively. We envision ROBUSTNAV as a testbed for adapting agent behavior across different perception and actuation properties. While assessing robustness to changes (stochastic or otherwise) in environments has been investigated in the robotics community [9, 4, 5, 6], the simulated nature of ROBUSTNAV enables practitioners to explore robustness against a rich and very diverse set of changes, while inheriting the advantages of working in simulation – speed, safety, low cost and reproducibility.

ROBUSTNAV consists of two widely studied embodied navigation tasks, Point-Goal Navigation (POINTNAV) [1] and Object-Goal Navigation (OBJECTNAV) [2] – the tasks of navigating to a goal-coordinate in a global reference frame or an instance of a specified object, respectively. Following the standard protocol, agents learn using a set of training scenes and are evaluated within a set of held out test scenes, but differently, ROBUSTNAV test scenes are subject to a variety of *visual* (see examples in Fig. 1) and *dynamics* (see examples in Fig. 2) corruptions.

As zero shot adaptation to test time corruptions may be out of reach for our current algorithms, we provide agents with a fixed "calibration budget" (number of interactions) within the target world for unsupervised adaptation. This mimics a real-world analog where a shipped robot is allowed to adapt to changes in the environment by executing a reasonable number of unsupervised interactions. Post calibration, agents are evaluated on the two tasks in the corrupted test environments using standard navigation metrics.

Our extensive analysis reveals that both POINTNAV and OBJECTNAV agents experience significant degradation in performance across the range of corruptions, particularly when multiple corruptions are applied together (POINTNAV results in Table. 1). We show that this degradation reduces in the presence of a clean depth sensor suggesting the advantages of incorporating multiple sensing modalities to improve robustness. We find that data augmentation and self-supervised adaptation strategies (PAD [7]) offer some zero-shot resistance and improvement over degraded performance, but are unable to fully recover this gap in performance. Interestingly, we also note that visual corruptions affect embodied tasks differently from static tasks like object recognition – suggesting that visual robustness should be explored within an embodied task. Finally, we analyze several interesting behaviors our agents exhibit in the presence of corruptions – such as increase in the number of collisions and inability to terminate episodes successfully.

In summary, our contributions include: (1) We present ROBUSTNAV– a framework for benchmarking and assessing the robustness of embodied navigation agents to visual and dynamics corruptions. (2) Our findings show that present day navigation agents trained in simulation underperform severely when evaluated in corrupt target environments. (3) We systematically analyze the kinds of mistakes embodied navigation agents make when operating under such corruptions. (4) We find that although standard data-augmentation techniques and self-supervised adaptation strategies offer some improvement, much remains to be done in terms of fully recovering lost performance.

ROBUSTNAV provides a fast framework to develop and test robust embodied policies, before they can be deployed onto real robots. While ROBUSTNAV currently supports navigation heavy tasks, the supported corruptions can be easily extended to more tasks, as they get popular within the Embodied AI community.

References

[1] Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, et al. On evaluation of embodied navigation agents. *arXiv*, 2018. 2

[2] Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. Objectnav revisited: On evaluation of embodied agents navigating to objects. *arXiv*, 2020. 2

[3] Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Russ R Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. In *NeurIPS*, 2020. 1

[4] Andrea Del Prete and Nicolas Mansard. Addressing constraint robustness to torque errors in task-space inverse dynamics. In *RSS*, 2015. 2

[5] Andrea Del Prete and Nicolas Mansard. Robustness to joint-torque-tracking errors in task-space inverse dynamics. *IEEE Trans. on Robotics*, 2016. 2

[6] Nirmal Giftsun, Andrea Del Prete, and Florent Lamiraux. Robustness to inertial parameter errors for legged robots balancing on level ground. In *International Conference on Informatics in Control, Automation and Robotics*, 2017. 2

[7] Nicklas Hansen, Yu Sun, Pieter Abbeel, Alexei A Efros, Lerrel Pinto, and Xiaolong Wang. Self-supervised policy adaptation during deployment. In *ICLR*, 2021. 2

[8] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *ICLR*, 2019. 1

[9] Sebastian Höfer, Kostas Bekris, Ankur Handa, Juan Camilo Gamboa, Florian Golemo, Melissa Mozifian, Chris Atkeson, Dieter Fox, Ken Goldberg, John Leonard, et al. Perspectives on sim2real transfer for robotics: A summary of the r:ss 2020 workshop. *arXiv*, 2020. 2

[10] Adithyavairavan Murali, Tao Chen, Kalyan Vasudev Alwala, Dhiraj Gandhi, Lerrel Pinto, Saurabh Gupta, and Abhinav Gupta. Pyrobot: An open-source robotics framework for research and benchmarking. *arXiv*, 2019. 2

[11] Nikolay Savinov, Alexey Dosovitskiy, and Vladlen Koltun. Semi-parametric topological memory for navigation. In *ICLR*, 2018. 1

[12] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. Habitat: A platform for embodied ai research. In *ICCV*, 2019. 1

[13] Mitchell Wortsman, Kiana Ehsani, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. Learning to learn how to learn: Self-adaptive visual navigation using meta-learning. In *CVPR*, 2019. 1

[14] Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J. Lim, Abhinav Gupta, Li Fei-Fei, and Ali Farhadi. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *ICRA*, 2017. 1