[Extended Abstract] Socially CompliAnt Navigation Dataset (SCAND): A Large-Scale Dataset Of Demonstrations For Social Navigation

Haresh Karnan[1], Anirudh Nair[2], Xuesu Xiao[2], Garrett Warnell[2,5], Sören Pirk[3], Alexander Toshev[3], Justin Hart[2], Joydeep Biswas[2] and Peter Stone[2,4]

***Abstract*— Social navigation is the capability of an autonomous agent, such as a robot, to navigate in a "socially compliant" manner in the presence of other intelligent agents such as humans. With the emergence of autonomously navigating mobile robots in human-populated environments (*e.g.,* domestic service robots in homes and restaurants and food delivery robots on public sidewalks), incorporating socially compliant navigation behaviors on these robots becomes critical to ensuring safe and comfortable human-robot coexistence. To address this challenge, imitation learning is a promising framework, since it is easier for humans to demonstrate the task of social navigation rather than to formulate reward functions that accurately capture the complex multi-objective setting of social navigation. The use of imitation learning and inverse reinforcement learning to social navigation for mobile robots, however, is currently hindered by a lack of large-scale datasets that capture socially compliant robot navigation demonstrations in the wild. To fill this gap, we introduce Socially CompliAnt Navigation Dataset (SCAND)—a large-scale, first-person-view dataset of socially compliant navigation demonstrations. Our dataset contains 8.7 hours, 138 trajectories, 25 miles of socially compliant, human-teleoperated driving demonstrations that comprises multi-modal data streams including 3D lidar, joystick commands, odometry, visual and inertial information, collected on two morphologically different mobile robots—a Boston Dynamics Spot and a Clearpath Jackal—by four different human demonstrators in both indoor and outdoor environments. We additionally perform preliminary analysis and validation through real-world robot experiments and show that navigation policies learned by imitation learning on SCAND generate socially compliant behaviors.** [1]

I. INTRODUCTION

Social navigation is the capability of an autonomous agent to navigate in a socially compliant manner such that it recognizes and reacts to the objectives of other navigating agents, at least somewhat adjusting its own path in response, while also projecting signals that can help the other agents reciprocate. Enabling mobile robots to navigate in a socially compliant manner has been a subject of great interest recently in the robotics and learning communities [1]–[5]. Towards enabling this capability, demonstration data of socially compliant navigation for mobile robots, such as the ones shown in Fig. 1, can be a valuable resource. For instance, such demonstration information can be used to learn socially compliant robot navigation using the paradigm of Learning from Demonstrations (LfD) [6], [7] or understanding human navigation in the presence of autonomous robots [8].

Fig. 1: A human demonstrator teleoperates the two robots, following a socially compliant strategy (left- moving with traffic, right- sticking to the right of the road) around human crowds.

Datasets for social navigation, generally used for learning and benchmarking, include data collected both in the real-world [9] and in simulated environments [10], [11]. While such datasets provide basic trajectories of the robots and humans, they either contain limited interactions in constrained, orchestrated environments or restrict themselves to indoor-only navigation scenarios. When collecting data in such controlled settings [9], naturally occurring social interactions including—but not limited to—following lane rules of a country, yielding to pedestrians and vehicles, walking with and against a crowd of people, and street crossing is not captured. Additionally, the robots used for data collection in previous social navigation datasets [9] tend to use a simple controller for point-to-point navigation that does not explicitly exhibit socially aware navigation.

Recently, imitation learning has emerged as a useful paradigm for designing mobile robot navigation controllers [12]–[15]. In this paradigm, the desired navigation behavior is first demonstrated by an agent such as a human, the recording of which is then utilized by an imitation learning algorithm to imitate. This intuitive way of teaching a task to a robot is also easy for non-expert humans since it only requires providing demonstrations, instead of defining the rules of the task itself, which may be hard to explicitly

[1]The University of Texas at Austin, Department of Mechanical Engineering haresh.miriyala@utexas.edu
[2]The University of Texas at Austin, Department of Computer Science, ani.nair@utexas.edu, {xiao, joydeepb, hart, pstone}@cs.utexas.edu
[3] Robotics@Google {toshev, pirk}@google.com
[4] Sony AI
[5] Computational and Information Sciences Directorate, Army Research Laboratory garrett.a.warnell.civ@army.mil

[1]Link to the full version : https://arxiv.org/abs/2203.15041

Fig. 2: Five example scenarios from SCAND showing the RGB image and below it the accompanying Lidar with the monocular image from side camera on the Spot. From left to right, the scenarios have the tags “Street Crossing”, “Narrow Doorway, “Navigating Through Large Crowds”, “Vehicle Interaction”, and “Crossing Stationary Queue.”

define for social navigation. Motivated by recent successes of imitation learning in robot navigation, we posit that one way to enable autonomous agents to navigate socially is through learning from human demonstrations of socially compliant navigation behavior. However, there is a lack of large-scale datasets containing socially compliant navigation demonstrations in the wild that can be utilized for imitation learning.

To fill this gap, in this work, we introduce a dataset of demonstrations for socially compliant robot navigation in the wild. Our dataset contains 8.7 hours of human-teleoperated, socially compliant, navigation demonstrations, specifically, Velodyne lidar scans, joystick commands, odometry, camera visuals, and 6D inertial (IMU) information collected on two morphologically different mobile robots—a Clearpath Jackal and a Boston Dynamics Spot—within the University of Texas at Austin university campus. Comprising 25 miles in total of 138 trajectories, Socially CompliAnt Navigation Dataset (SCAND) is publically released[2] and also contains labeled tags of naturally occurring social interactions with every trajectory. Additionally, we show that with SCAND, it is possible to learn socially compliant local navigation policies through imitation learning.

II. DATA COLLECTION PROCEDURE

In this section, we first describe the data collection procedure used in SCAND and outline the sensor-suite present on both robots. We then describe the labeled annotations of social interactions provided with every trajectory.

A. Collecting Data

To collect multi-modal, socially compliant demonstration data for robot navigation, 4 human demonstrators—including the first two authors of this work—navigate the robot by teleoperation using a joystick. For each of the 138 trajectories in SCAND, the human demonstrator walks behind the robot at all times, maintaining on average two meters distance. Unlike other datasets for social navigation [9], we do not restrict data collection to a controlled, indoor environment or orchestrate a social scenario for data collection. Instead, similar to the JRDB dataset [8], we perform data collection in the wild in both indoor and outdoor environments. The two robots are driven around the university campus on frequently used sidewalks, roads, and lawns, and inside buildings, all with people in the scene during peak hours of high foot traffic. This includes data collected outdoors near the university's football stadium on two game days with high traffic public crowds gathered near the arena. The Spot is driven at linear and angular velocities in the range of $[0, 1.6]$ m/s and $[-1.5, 1.5]$ rad/s, respectively, and the Jackal in the range of $[0, 2.0]$ m/s and $[-1.5, 1.5]$ rad/s, respectively. Note that these velocities are within the range of many people's normal walking speed.

Fig. 3 shows the sensors present on the Clearpath Jackal and the Boston Dynamics Spot robots. Both robots have in common a VLP-16 Velodyne laser puck publishing at a frequency of 10 Hz, a 6D inertial (IMU) sensor at 16 Hz, and a front-facing Azure Kinect RGB camera at 20 Hz. In addition to these common sensors, the Jackal has a front-facing stereo camera (20 Hz) and wheel odometry (30 Hz), while the Spot has five monocular cameras on its body (publishing at 5 Hz), that are placed as shown in Fig. 3. We utilize the Boston Dynamics APK to record the visual odometry of its body frame and the joint angles of the legs on the robot. SCAND also contains transforms between the frames of each of the sensors relative to the robot's body for both robots. We utilize AMRL's software stack [20] for data collection from different sensors which we record in the rosbag format [21].

Although we provide visual scene information in the form of surround-view monocular images on the Spot, RGB image from the front-facing Kinect camera, and 3D Velodyne laser scans on both robots, since the focus of this work is specifically on navigation, we do not provide any labeled annotations for human detection or tracking. We refer the reader to the JRDB dataset [8] which contains detailed, high-quality annotations for solving perception-related tasks. Instead, SCAND contains joystick commands of linear and angular velocities executed by the demonstrator while teleoperating the robot socially, along with rich, multi-modal sensory information of the environment including labeled annotations of 12 different social interactions in every trajectory. Fig. 2 shows five example scenarios and their associated tags.

[2] www.cs.utexas.edu/~xiao/SCAND/SCAND.html

Dataset	# Traj.	Dist. (Km)	Dur. (min)	Sensors	Nav. method	# Robots	Location
CoBot [16]	1082	131	15600	2D Range Scanner, RGB-D Camera, Wheel Odometry	Autonomous	2	Indoors + Outdoors
L-CAS [17]	3	N/A	49	3D LiDAR	Teleoperated	1	Indoors
NCLT [18]	27	147.4	2094	3D LiDAR, RGB Camera, IMU, Wheel Odometry, GPS	Teleoperated	1	Indoors + Outdoors
FLOBOT [19]	6	N/A	27.5	3D LiDAR, RGB-D camera, Stereo Camera, 2D LiDAR, OEM incremental measuring wheel encoder, IMU	Autonomous	1	Indoors
JRDB [8]	54	N/A	64	3D LiDAR, 2D LiDAR, Omnidirectional Stereo Suite, RGB camera, RGB-D stereo camera, 6D IMU	Teleoperated	1	Indoors + Outdoors
THÖR [9]	600	N/A	60	3D LiDAR, Motion capture system, Eye-tracking Glasses	Autonomous	1	Indoors
SCAND	138	40	522	3D LiDAR, RGB-D Camera, Monocular Camera, Stereo Camera, Wheel Odometry, Visual Odometry	Teleoperated	2	Indoors + Outdoors

TABLE I: Comparison of real-world datasets for robot navigation.

Tag	Description	# Tags
Against Traffic	Navigating against oncoming traffic	22
With Traffic	Navigating with oncoming traffic	74
Street Crossing	Crossing across a street	34
Overtaking	Overtaking a person or groups of people	14
Sidewalk	Navigating on a sidewalk	57
Passing Conversational Groups	Navigating past a group of 2 or more people that are talking amongst themselves	38
Blind Corner	Navigating past a corner where the robot cannot see the other side	6
Narrow Doorway	Navigating through a doorway where the robot waits for a human to open the door	15
Crossing Stationary Queue	Walking across a line of people	6
Stairs	Walking up and/or down stairs	22
Vehicle Interaction	Navigating around a vehicle	21
Navigating Through Large Crowds	Navigating among large unstructured crowds	27

TABLE II: Descriptions of labeled tags contained in SCAND

B. Labeled Annotations of Social Interactions

We annotate each trajectory in SCAND with labels describing social interactions that occurred along the path. The labels are in the form of a list of textual captions of social interactions taking place in a trajectory, chosen from a set of twelve predefined labels of social interactions observed in SCAND. For the full list of labels, refer to Table II. We intend the labels to be useful for future studies of specific scenarios that occur during social navigation in the real-world.

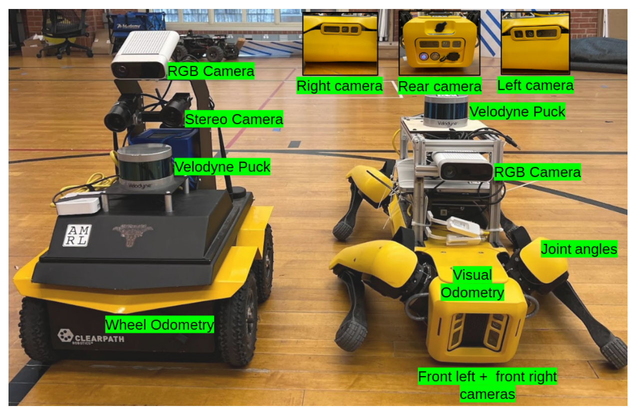

Fig. 3: Sensors present on the wheeled Jackal and the legged Spot robots. Along with this multi-modal sensor information, SCAND also contains joystick commands issued during the navigation demonstration.

III. ANALYSIS

In this section, we discuss preliminary analysis using our dataset which show that navigation policies learned through imitation learning using SCAND generates socially compliant and safe navigation behaviors.

We apply the behavior cloning algorithm [22] to learn a reactive local planner that predicts joystick action commands. The local planner takes as its input past three LiDAR scans, odometry information of the robot for each of the scans with respect to the most recent observation and `move_base` global path information ten meters into the future. The local planner predicts the next ten action commands (forward velocity v and angular velocity ω) which is learned from the joystick commands issued by the human demonstrator in SCAND. To evaluate the learned local planner, we perform a human subject study in the real world with a static and dynamic environment as shown in Fig. 4. The human subjects are asked to evaluate the social compliance and

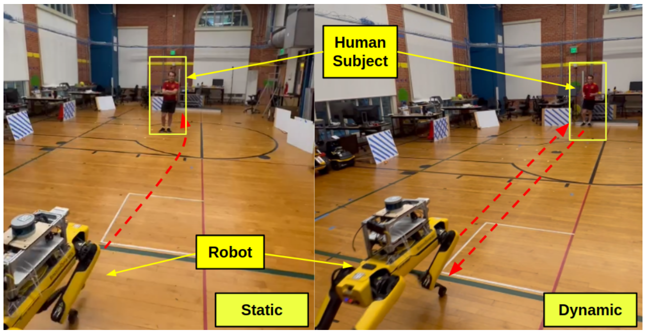

Fig. 4: Evaluating the local planner agent trained using Behavior Cloning on SCAND. Scenario on the left shows a stationary human in the robot's path and the scenario on the right shows a human walking to the location of the robot. The robot is evaluated on social compliance and safety as it navigates to its goal position.

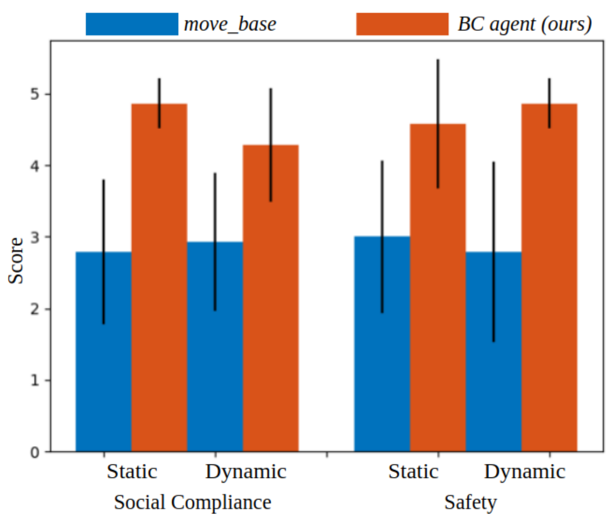

Fig. 5: Mean and standard deviation of scores assigned by the fourteen human participants in the evaluation study for the learned local planner.

safety of the robot, ranking from 1 (lowest) to 5 (highest). Fig. 5 shows the responses of fourteen human participants. On average, more humans felt the imitation learning agent trained on SCAND was more socially compliant (SCAND mean=4.39, sd=0.99; `move_base` mean=2.86, sd=0.82) and safer (SCAND mean=4.71, sd=0.70; `move_base` mean=2.89, sd=1.18) than the `move_base` agent. The results for both questions are statistically significant as tested by a One-Way Analysis of Variance (ANOVA) (*Safe* $F_{1,55} = 47.87, p < 0.001$; *Socially Compliant* $F_{1,55} = 38.67, p < 0.001$). This is expected since the `move_base` agent is not designed to exhibit social compliance.

IV. CONCLUSIONS AND FUTURE WORK

In this work, we introduce the Socially CompliAnt Navigation Dataset (SCAND), a large-scale dataset of demonstrations for mobile robot social navigation. SCAND contains 8.7 hours, 138 trajectories, 25 miles of socially compliant driving demonstrations, collected on two morphologically different robots. In addition to the multi-modal sensory data streams from the two robots, SCAND also contains labeled annotations of social interactions for all trajectories. We illustrate the usefulness of SCAND by training a behavior cloning agent on the demonstrations from SCAND and show that it is possible to learn a socially compliant local planner for mobile robot navigation. We further validate the performance of the behavior cloned local planner through human trials on two social navigation scenarios and show that the participants perceived the imitation learning agent to be relatively more socially compliant and safe, compared to a naive `move_base` agent. While we show here that the BC agent was able to handle simple social navigation scenarios, better imitation learning algorithms may be needed to handle more sophisticated social navigation scenarios that are present in SCAND. Although SCAND includes a wide variety of social navigation scenarios, there may be novel interactions that are less frequent. To improve generalizability of a learning based approach to unseen situations, exploring representation learning for social navigation with SCAND is a promising future direction. Another interesting future research direction is to explore Real-to-Sim transfer with SCAND and improve parameterized simulated social navigation environments to generate more realistic social interactions between virtual agents, directly benefiting data hungry approaches such as reinforcement learning.

ACKNOWLEDGEMENT

This work has taken place in the Learning Agents Research Group (LARG) and Autonomous Mobile Robotics Laboratory (AMRL) at UT Austin. LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. AMRL research is supported in part by NSF (CAREER2046955, IIS-1954778, SHF-2006404), ARO (W911NF-19-2- 0333, W911NF-21-20217), DARPA (HR001120C0031), Amazon, JP Morgan, and Northrop Grumman Mission Systems. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, p. 4282–4286, May 1995. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.51.4282

[2] X. Xiao, B. Liu, G. Warnell, and P. Stone, "Motion control for mobile robot navigation using machine learning: a survey," 2020.

[3] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," 2018.

[4] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," *IEEE Access*, vol. 9, p. 10357–10377, 2021. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2021.3050338

[5] R. Mirsky, X. Xiao, J. Hart, and P. Stone, "Prevention and resolution of conflicts in social navigation – a survey," 2021.

[6] S. Schaal, "Learning from demonstration," in *Proceedings of the 9th International Conference on Neural Information Processing Systems*, ser. NIPS'96. Cambridge, MA, USA: MIT Press, 1996, p. 1040–1046.

[7] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889008001772

[8] R. Martin-Martin, M. Patel, H. Rezatofighi, A. Shenoi, J. Gwak, E. Frankel, A. Sadeghian, and S. Savarese, "Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1–1, 2021. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2021.3070543

[9] A. Rudenko, T. P. Kucner, C. S. Swaminathan, R. T. Chadalavada, K. O. Arras, and A. J. Lilienthal, "Thör: Human-robot navigation data collection and accurate motion trajectories dataset," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 676–682, 2020.

[10] N. Tsoi, M. Hussein, J. Espinoza, X. Ruiz, and M. Vázquez, "Sean: Social environment for autonomous navigation," in *Proceedings of the 8th International Conference on Human-Agent Interaction*, 2020, pp. 281–283.

[11] L. J. Manso, P. Nuñez, L. V. Calderita, D. R. Faria, and P. Bachiller, "Socnav1: A dataset to benchmark and learn social navigation conventions," *Data*, vol. 5, no. 1, p. 7, 2020.

[12] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. Theodorou, and B. Boots, "Agile autonomous driving using end-to-end deep imitation learning," 2019.

[13] H. Karnan, G. Warnell, X. Xiao, and P. Stone, "VOILA: visual-observation-only imitation learning for autonomous navigation," *CoRR*, vol. abs/2105.09371, 2021. [Online]. Available: https://arxiv.org/abs/2105.09371

[14] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," 2016.

[15] Y. Lecun, U. Muller, J. Ben, E. Cosatto, and B. Flepp, "Off-road obstacle avoidance through end-to-end learning." 01 2005.

[16] J. Biswas and M. M. Veloso, "Localization and navigation of the cobots over long-term deployments," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1679–1694, 2013. [Online]. Available: http://ijr.sagepub.com/content/32/14/1679.abstract

[17] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3d lidar-based tracking," in *In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, Canada, September 2017.

[18] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.

[19] Z. Yan, S. Schreiberhuber, G. Halmetschlager, T. Duckett, M. Vincze, and N. Bellotto, "Robot perception of static and dynamic objects with an autonomous floor scrubber," *Intelligent Service Robotics*, vol. 13, no. 3, pp. 403–417, 2020.

[20] J. Biswas, "Amrl autonomy stack," https://github.com/ut-amrl/graph_navigation, 2013.

[21] Stanford Artificial Intelligence Laboratory et al., "Robotic operating system." [Online]. Available: https://www.ros.org

[22] M. Bain and C. Sammut, "A framework for behavioural cloning," in *Machine Intelligence 15*, 1995.