Curriculum Learning for GPS-Free Indoor Social Navigation

Gunjan Chhablani[†,*], Madhura Keshava Ummettuguli[†,*], Siva Kailas[†*]
gchhablani3@gatech.edu, mummettuguli3@gatech.edu, skailas3@gatech.edu
[†] Georgia Institute of Technology

Abstract

Tasks involving human-robot interaction demand seamless collaboration between the two within indoor settings. Habitat 3.0 [11] introduced a novel Social Navigation task where agents find, track, and follow humans while avoiding collisions. Their baselines show that performance relies heavily on human GPS availability. However, indoor GPS sensors are rarely reliable in real-life and it may be impractical to provide GPS for every human in the scene. In this work, we tackle the issue of realistic social navigation by relaxing the human GPS requirement at every time step. We achieve this via a curriculum learning strategy for training an RL policy capable of finding and tracking humans with sparse or no reliance on human GPS observations. Our experiments demonstrate the effectiveness of our curriculum strategy, achieving comparable performance to the baselines with lesser samples, using a single GPS observation at the beginning of the episode. The project website and videos can be found at `gchhablani.github.io/socnav-curr`.

1. Introduction

Embodied navigation in indoor environments has been a long-standing challenge [1] in robotics and artificial intelligence. Recent works [9, 14, 18] have leveraged deep reinforcement learning to address these tasks. Several prior works [12, 13] in this area assume that the GPS location of the goal is provided to the agent at each timestep. Some works have attempted to relax this assumption through various means such as visual odometry [4, 10] and information bottleneck [7].

Recently, Habitat 3.0 [11] proposed a Social Navigation (SocialNav) task where the agent is spawned in an indoor scene with a human and the agent is tasked with finding and following the human at a safe distance while avoiding collisions. They propose an end-to-end RL baseline which uses the human GPS location at each timestep. However, it is impractical in real environments to rely heavily on indoor GPS systems and to expect GPS availability for each human in the environment. Therefore, we attempt to address a more realistic SocialNav task by learning a policy that only has access to infrequent human GPS location during the episode. We realize this by leveraging curriculum training on the GPS location.

Curriculum learning [3] involves starting with easy versions of a task and gradually increasing difficulty until the original task is mastered. Various curriculum strategies are applied in RL scenarios, such as reverse curriculum generation [5] and accuracy-based curriculum learning [6]. In navigation tasks, [2, 16, 17] employ curriculum learning based on distance to goal or waypoint/trajectory decomposition, while [8] focuses on environment difficulty. Inspired by [15], we design a curriculum strategy that operates on the human GPS availability in SocialNav. Via a curriculum, we show that the agent with no access to GPS can learn a policy that performs at least as good as an agent with no GPS sensor, and reaches its peak finding success much earlier.

2. Methodology

We use the standard RL setting (same as [11]), modeling the SocialNav task as a partially-observable Markov Decision Process. The observations at each timestep are derived from four simulation sensors: an arm depth camera, an arm RGB camera, a humanoid detector, and a humanoid GPS sensor. We perform experiments with relaxed GPS availability condition. We present two different ways to represent the human GPS location for timesteps where the current human GPS location is unavailable:

- **ZeroGPS** (`ZGPS`): Provides $(0, 0)$ as GPS location.
- **LastGPS** (`LGPS`): Provides the last known GPS location.

2.1. Human GPS Availability Curriculum

Without human GPS sensor, the finding success rate (FS) drops significantly [11]. Thus, we hypothesize that: (1) A policy with consistent GPS information relies heavily on that information; and (2) An adaptive policy trained via curriculum on GPS availability would enable the agent to retain desirable properties for the task.

[*] Equal contribution

During training, initially GPS information is available at every step. Then, based on upper and lower training FS score thresholds, we either decrease or increase the frequency of GPS availability, thereby making the task harder or easier. The goal of the curriculum is to eventually reach a difficulty level where GPS is never available after timestep 0 and the policy is still performant. The curriculum level K in our experiments ranges from $[1, 1500]$, corresponding to the interval in the episode at which agent has access to the current GPS sensor information. We train for 300M steps, and check every 10M steps if the curriculum level needs to be updated. Additionally, we use the first 10M as warmup steps so the agent achieves a good FS with full GPS conditions. The upper threshold is 0.9 and lower threshold is set to 0.8, 0.85, 0.88 for 0-100M, 100-200M, and 200-300M steps during training, respectively. In this work, we explore various approaches for updating the curriculum:

- **Additive** (`Add`): The curriculum increments (+50) or decrements (-25) a fixed value based on FS.
- **Multiplicative** (`Mul`): The curriculum doubles or halves its current GPS frequency depending on FS.
- **Dynamic Add** (`Dyn-Add`): Adds $[30, 60]$ or subtracts $[10, 80]$, a dynamic value that scales with training FS.

3. Experiments and Results

3.1. Metrics and Evaluation

We evaluate our policies on 500 unseen episodes and borrow the metrics from [11]. `ZGPS` evaluation assumes a fixed GPS observation of (0,0) at each step, while `LGPS` receives only the initial human GPS at every time step.

3.2. Baselines

We consider two baselines: `Full` (Human GPS observed at every timestep [11]), and `NoGPS` (No GPS sensor). The evaluation results in Figure 1 show that the `NoGPS` baseline achieves 0.92 eval FS but inconsistent performance. `Full` achieves near-perfect performance (0.98 eval FS) consistently.

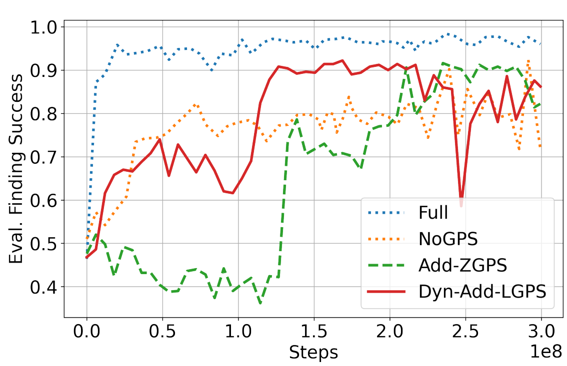

Figure 1. Eval FS for best-performing curriculum policies

Experiment	FS (↑)	FR (↑)	SPS (↑)	CR (↓)	R (↑)
`NoGPS`	0.92	**0.72**	**0.52**	0.52	**6770.28**
`Full`	**0.98**	0.68	**0.52**	0.59	6079.23
`Mul-ZGPS`	0.52	0.06	0.02	**0.40**	31.07
`Add-ZGPS`	0.92	0.66	0.44	0.60	4941.44
`Dyn-Add-ZGPS`	0.89	0.60	0.41	0.63	4784.58
`Mul-LGPS`	0.92	0.63	0.38	0.64	4044.00
`Dyn-Add-LGPS`	0.92	0.65	0.44	0.62	5364.39
`Add-LGPS`	0.91	0.71	0.51	0.54	6605.48

Table 1. Evaluation results on checkpoints with the highest FS.

3.3. ZeroGPS vs LastGPS

For brevity, we depict only the top two curriculum strategies in Figure 1. `Dyn-Add-LGPS` outperforms `Add-ZGPS`, requiring fewer iterations for similar performance, indicating its higher sample efficiency. Table 1 confirms the superior performance of `LGPS` over `ZGPS`. This result is likely due to `ZGPS` needing to implicitly remember the latest human GPS observation, whereas `LGPS` continually receives the cached human GPS.

3.4. Additive vs Multiplicative vs Dynamic Additive

From Table 1, we observe that FS for `Add-ZGPS` and `Dyn-Add-LGPS` reaches 0.92. `Mul-ZGPS` strategy performs poorly (0.52 FS) but `Mul-LGPS` performs well (0.92 FS). If the difficulty updates too rapidly, as in `Mul`, it can exacerbate the learning curve. Since `LGPS` is easier than `ZGPS` setting (see Section 3.3), the combination `Mul-ZGPS` results in too difficult of a curriculum.

3.5. GPS vs Baselines

Figure 1 shows that our best curriculum strategies (`Add-ZGPS` and `Dyn-Add-LGPS`) perform as good as the `NoGPS` baseline, but reach a high performance early on during the training. We also observe more stable curves compared to the `NoGPS` baseline towards the end of training. This shows that curriculum learning helps in learning a robust policy with very few samples for finding the human without relying too much on GPS availability.

4. Conclusion and Future Work

In this work, we use curriculum training to relax the requirement of human GPS availability in the SocialNav task. Our approach achieves comparable success rates to the `NoGPS` conditions with better stability and using fewer training samples, demonstrating the effectiveness of our curriculum learning. In future, we aim to improve on other metrics such as collision rate and SPS, which currently lag behind `NoGPS`. Additionally, we will explore strategies to encourage active exploration, as we observed instances where the agent moves in circles until a human is visible.

5. Acknowledgement

We would like to thank Dr. Sehoon Ha (https://faculty.cc.gatech.edu/~sha9/) for providing us insights and feedback on the formulation of this project.

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 preprint arXiv:1807.06757*, 2018. 1

[2] Kaushik Balakrishnan, Punarjay Chakravarty, and Shubham Shrivastava. An a* curriculum approach to reinforcement learning for rgbd indoor robot navigation. *arXiv preprint arXiv:2101.01774*, 2021. 1

[3] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009. 1

[4] Samyak Datta, Oleksandr Maksymets, Judy Hoffman, Stefan Lee, Dhruv Batra, and Devi Parikh. Integrating egocentric localization for more realistic point-goal navigation agents. In *Conference on Robot Learning*, pages 313–328. PMLR, 2021. 1

[5] Carlos Florensa, David Held, Markus Wulfmeier, Michael Zhang, and Pieter Abbeel. Reverse curriculum generation for reinforcement learning. In *Conference on robot learning*, pages 482–495. PMLR, 2017. 1

[6] Pierre Fournier, Olivier Sigaud, Mohamed Chetouani, and Pierre-Yves Oudeyer. Accuracy-based curriculum learning in deep reinforcement learning. *arXiv preprint arXiv:1806.09614*, 2018. 1

[7] Guillermo Grande, Dhruv Batra, and Erik Wijmans. Realistic pointgoal navigation via auxiliary losses and information bottleneck. *arXiv preprint arXiv:2109.08677*, 2021. 1

[8] Luca Marzari, Davide Corsi, Enrico Marchesini, and Alessandro Farinelli. Curriculum learning for safe mapless navigation. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pages 766–769, 2022. 1

[9] Arsalan Mousavian, Alexander Toshev, Marek Fiser, Jana Kosecka, Ayzaan Wahid, and James Davidson. Visual representations for semantic target driven navigation. 2019. 1

[10] Ruslan Partsey, Erik Wijmans, Naoki Yokoyama, Oles Dobosevych, Dhruv Batra, and Oleksandr Maksymets. Is mapping necessary for realistic pointgoal navigation? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17232–17241, 2022. 1

[11] Xavier Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta Desai, Alexander Clegg, Michal Hlavac, So Yeon Min, et al. Habitat 3.0: A co-habitat for humans, avatars, and robots. In *The Twelfth International Conference on Learning Representations*, 2023. 1, 2

[12] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A Platform for Embodied AI Research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1

[13] Erik Wijmans, Abhishek Kadian, Ari Morcos, Stefan Lee, Irfan Essa, Devi Parikh, Manolis Savva, and Dhruv Batra. Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames. In *International Conference on Learning Representations*, 2019. 1

[14] Yi Wu, Yuxin Wu, Georgia Gkioxari, and Yuandong Tian. Building generalizable agents with a realistic and rich 3d environment. 2018. 1

[15] Wojciech Zaremba and Ilya Sutskever. Learning to execute. *arXiv preprint arXiv:1410.4615*, 2014. 1

[16] Jiwen Zhang, Jianqing Fan, Jiajie Peng, et al. Curriculum learning for vision-and-language navigation. *Advances in Neural Information Processing Systems*, 34:13328–13339, 2021. 1

[17] Wang Zhu, Hexiang Hu, Jiacheng Chen, Zhiwei Deng, Vihan Jain, Eugene Ie, and Fei Sha. Babywalk: Going farther in vision-and-language navigation by taking baby steps. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2539–2556, 2020. 1

[18] 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. *arXiv preprint arXiv:1609.05143*, 2016. 1