Towards Socially Compliant Navigation: Hybrid Parameter Optimization for Falcon in Dynamic Environments

Yang Li
City University of Hong Kong
yang.li.mne@my.cityu.edu.hk

Congfei Li
City University of Hong Kong
Harbin Institute of Technology
congfei.li@my.cityu.edu.hk

Yuxiang Sun
City University of Hong Kong
yx.sun@cityu.edu.hk

Abstract

Socially compliant navigation in dynamic indoor environments populated by humans is challenging. There are many models, such as Falcon, try to balance task efficiency and social norms amid complex human-machine interactions. To improve the Falcon model performance, this report uses a hybrid optimization strategy. The strategy combines proportional-constrained parameter coupling tuning with grid search. It aims to address the core challenges of dimensionality explosion and multi-objective balance in parameter optimization for reward functions. The strategy first presets proportions for the pair of core parameters related to task completion, which reduces dimensionality and enhances efficiency. Then, to ensure the comprehensive exploration, it employs a systematic grid search, which traverses coupled parameter combinations and social compliance-related parameters. Experimental results demonstrate that this strategy achieves effective gains in terms of multiple evaluation metrics.

1. Introduction

In crowded environments, robots must simultaneously satisfy task completion (*e.g.*, reaching targets) and social compliance (*e.g.*, respecting human personal space). However, the reward function parameters in existing models, such as Falcon [1], are often empirically tuned, potentially leading to suboptimal balancing between navigation efficiency and social constraints.

As one of the core regulatory mechanisms in reinforcement learning, the parameter configuration of the reward function directly influences how the agent allocates weight between task completion and avoiding conflicts. Unreasonable parameters may result in low task success rates or frequent social intrusions.

By precisely adjusting parameters, the practicality of Falcon in real-world scenarios can be enhanced at low cost without altering the model's core architecture.

Our team systematically optimized the reward function for the Falcon model by selecting parameters critical to task completion and social compliance, employing grid search to identify the optimal parameter combination.

After parameter optimization, the model achieves significantly higher task success rates on the validation set of the Social-HM3D dataset [2] compared to the original Falcon model (reaching up to 15% success rate), while maintaining a reasonable human personal space, which validates the effectiveness of parameter fine-tuning.

2. Method

The essence of the reward function is a weighted combination of multiple reward components, where the parameter values directly determine the weight distribution between the two objectives. The reward function for the Falcon model is defined as follows:

$$\text{Total Reward} = r_{\mathrm{S}} + r_{\mathrm{M}} + r_{\mathrm{D}} + r_{\mathrm{R}} \tag{1}$$

where:

- r_{S}: social navigation reward.
- r_{M}: multiple agent navigation reward.
- r_{D}: distance to goal reward.
- r_{R}: rearrange reward.

To mitigate the conflict between efficiency and social compliance, our team selects one parameter from each of the first three reward functions to represent their respective contributions to controlling these reward categories:

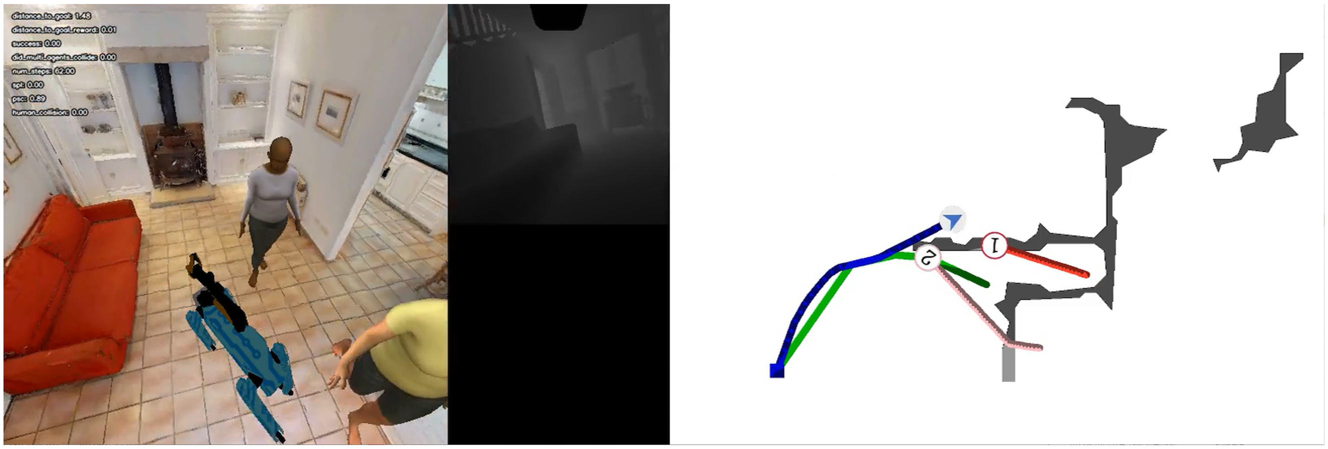

Figure 1. A snapshot of a qualitative demonstration video. Left: Third-person RGB view. Center: Robot's first-person depth image. Right: Map with agents' historical trajectory (blue for robot, green for ground truth, other for human).

- **facing human distance threshold parameter**: the distance threshold for human rewards. When the robot's distance from a human falls below this threshold, the robot receives an additional reward based on its orientation.
- **trajectory cover penalty**: the penalty applied to prevent the robot from blocking a human's future path.
- **task completion reward**: the reward granted when the robot successfully reaches the goal within 1.0 meters.

In Fig. 1, the human labeled number 2, with a pink historical trajectory, has a walking direction toward the upper-left corner of the map. Before the robot reaches this human, the robot has already made an early turn, shifting from its original direction toward the right side of the map to the upper-right corner. This is because the Falcon model predicts the human's movement trajectory and performs an evasive maneuver due to the trajectory cover penalty. That inspires the trajectory cover penalty is crucial for collision, social compliance, and path efficiency.

This study employs parameter optimization techniques, Parameter Coupled Tuning [3] and Grid Search [4], to optimize the Falcon model's reward function parameters efficiently.

We utilize proportional constraint-based parameter coupled tuning for the first and third parameters of the reward function. We establish a coupled interaction mechanism between them by imposing a predefined mathematical relationship (here, it means a fixed ratio) to constrain their variation patterns. Specifically, these two parameters are adjusted synchronously at a fixed ratio. We conduct training and performance analysis for parameter combinations corresponding to different ratios. This strategy significantly reduces the complexity of the parameter space, substantially shortens the experimental cycle, and prevents hardware resource consumption on redundant or ineffective parameter combinations.

Although parameter coupled tuning achieves alignment between the parameter space and research objectives, to further ensure the comprehensiveness of parameter search and the reliability of results, we systematically traverse and optimize the coupled parameter combinations and the second parameter using grid search. We aim to achieve a balanced approach of efficiency, specificity, and comprehensiveness in the parameter optimization process through the synergistic application of these two techniques.

3. Experiments

3.1. Dataset

We use the official data provided by the *RoboSense Challenge 2025* [5] held at IROS 2025. This competition builds upon the legacy of the *RoboDepth Challenge 2023* [6, 7] at ICRA 2023 and the *RoboDrive Challenge 2024* [8, 9] at ICRA 2024, continuing the collective effort to advance robust and scalable robot perception. Each track in this competition is grounded on an established benchmark designed for evaluating real-world robustness and generalization [1, 10–14]. The Social-HM3D dataset contains 844 high-fidelity 3D indoor scenes reconstructed from real building-scale scans, constituting the largest benchmark suite for social navigation in this research area.

3.2. Experimental Setup

Our team primarily focuses on the three parameters mentioned in previous section. The model structure and the other parameters remain unchanged. The core evaluation metrics include:

3.2.1 Success Rate

Success Rate (SR) quantifies the fraction of episodes where the robot successfully reaches the goal. SR ranges from 0 to 1, where 1 indicates robot completed navigation in all episodes. The mathematical representation of SR is:

$$\mathrm{SR} = \frac{1}{N}\sum_{i=1}^{N} S_i, \tag{2}$$

where:
- S_i: 1 if the robot successfully reaches the goal within 1.0 meter in episode i and the goal being oracle-visible from the final pose, 0 otherwise.
- N: Total number of episodes.

3.2.2 Success weighted by Path Length

Success weighted by Path Length (SPL) quantifies path efficiency. SPL ranges from 0 to 1, where 1 indicates robot completed navigation via the shortest path. The formula is shown as following:

$$\mathrm{SPL} = \mathrm{SR} \cdot \frac{l_i}{\max(p_i, l_i)}, \tag{3}$$

where:
- l_i: Length of the shortest path to the goal in episode i.
- p_i: Path length taken by the agent in episode i.

3.2.3 Path Safety Compliance

Path Safety Compliance (PSC) quantifies how well a robot maintains safe distances from humans during navigation. PSC ranges from 0 to 1, where 1 indicates full compliance. It is calculated as the ratio of compliant steps to the total number of steps taken:

$$\mathrm{PSC} = \frac{1}{T}\sum_{i=1}^{S} C_i, \tag{4}$$

where:
- C_i: 1 if robot maintains a distance of at least 1.0 meter from any humans, 0 otherwise.
- T: The total number of steps taken by the robot.

3.3. Implementation Details

The experiments are conducted on a computing cluster equipped with 8 NVIDIA A5000 GPUs (24GB memory each), interconnected with 4 NVLink bridges to facilitate efficient inter-GPU communication. The software environment was built on the Ubuntu 18.04 system.

For model training, we employed the Proximal Policy Optimization (PPO) algorithm to optimize the neural network policy. The total training steps were set to 15,000,000, with a learning rate of -2.5×10^{-4}.

δ	trajectory cover penalty		
	-5.0×10^{-4}	-2.5×10^{-4}	-1.0×10^{-4}
0.2	0.622	**0.77**	0.698
0.1	0.738	0.734	0.753
0.08	0.748	0.727	0.73

Table 1. Effect of δ and Trajectory Cover Penalty on **Success Rate**. The bolded values represent the highest success rate.

δ	trajectory cover penalty		
	-5.0×10^{-4}	-2.5×10^{-4}	-1.0×10^{-4}
0.20	0.532	**0.706**	0.643
0.10	0.688	0.683	0.699
0.08	0.694	0.670	0.681

Table 2. Effect of δ and Trajectory Cover Penalty in terms of **Success weighted by Path Length**. The bolded values represent the highest success rate.

3.4. Ablation Study

Our team defines the ratio of task completion reward and facing human distance threshold parameter as δ:

$$\delta = \frac{c}{f}, \tag{5}$$

where:
- c : task completion reward.
- f : facing human distance threshold parameter.

By parameter coupled tuning, we choose three parameter combination:
- $\delta = 0.20$: facing human distance threshold parameter=10.0 and task completion reward=2.0.
- $\delta = 0.10$: facing human distance threshold parameter=15.0 and task completion reward=1.5.
- $\delta = 0.08$: facing human distance threshold parameter=17.0 and task completion reward=1.3.

3.4.1 Basic Analysis

Tab. 1 investigates the variation patterns of SR concerning δ and trajectory cover penalty. The data reveal a significant optimal solution for parameter coupling. When $\delta = 0.2$ and trajectory cover penalty = -2.5×10^{-4}, SR reaches its peak value of 0.770 (highlighted in bold), indicating the highest probability of task completion under this combination. If δ decreases (e.g., to 0.08) or the trajectory penalty deviates from the medium intensity (-2.5×10^{-4}) level, SR declines to varying degrees.

Tab. 2 focuses on SPL, where higher values indicate more efficient paths during task completion. The

δ	trajectory cover penalty		
	-5.0×10^{-4}	-2.5×10^{-4}	-1.0×10^{-4}
0.20	0.901	**0.911**	0.867
0.10	0.902	0.902	0.908
0.08	0.902	0.909	0.909

Table 3. Effect of δ and Trajectory Cover Penalty on **Path Safety Compliance**. The bolded values represent the highest success rate.

result shows strong overlap with Tab. 1's optimal parameters: SPL peaks at 0.706 (bolded) when $\delta = 0.2$ and trajectory cover penalty = -2.5×10^{-4}. Additionally, when $\delta = 0.1$, SPL exhibits a gradual upward trend as the penalty weakens, indicating reduced sensitivity to trajectory penalties and enhanced stability in path efficiency under this δ setting.

Tab. 3 measures PSC. Like the previous two tables, the optimal PSC value occurs at $\delta = 0.2$ and trajectory cover penalty = -2.5×10^{-4} ($PSC = 0.911$, bolded). Most parameter combinations maintain PSC above 0.9, indicating the Falcon system generally excels in social space compliance. Social compliance is less susceptible to interference from task reward and penalty intensity parameters.

3.4.2 Task Efficiency Analysis

When $\delta = 0.20$ and trajectory cover penalty=-2.5×10^{-4}, $SR = 0.770$ and $SPL = 0.706$, there was a difference of 0.064 between them. This combination represents the joint optimal solution for both Tab. 1 and Tab. 2. However, compared to other parameter combinations, the improvement in SR is significantly greater than that in SPL, indicating that path efficiency imposes relative constraints on task success in this scenario. When $\delta = 0.1$ and trajectory cover penalty=-1.0×10^{-4}, $SR = 0.753$ and $SPL = 0.699$, with a difference of 0.054. These values are closer, indicating that this parameter coupling achieves a more balanced trade-off between mission success and path efficiency.

3.4.3 Social Compliance Analysis

Under the jointly optimal parameter combination ($\delta = 0.20$, trajectory cover penalty=-2.5×10^{-4}), the difference between SR and SPL, Δ_{SR-SPL}, is maximal, yet PSC reaches its peak (0.911). At first glance, this phenomenon appears consistent with the conventional understanding that task efficiency conflicts with social compliance. However, examining the overall patterns across all parameter combinations reveals that the Δ_{SR-SPL} does not exhibit a simple linear relationship with PSC. For instance, when Δ_{SR-SPL} is minimal (0.049, corresponding to $\delta = 0.08$, trajectory cover penalty=-1.0×10^{-4}), SR and SPL are at moderate levels, yet PSC remains relatively high (0.909). This indicates that when the Δ_{SR-SPL} falls within a reasonable range, synergistic gains with PSC can be achieved by precisely regulating the internal balance between task success and path efficiency.

4. Conclusion

One of the core challenges in the social navigation task is the high sensitivity of model performance to parameters and the lack of universal tuning criteria. This study proposes a hybrid optimization strategy combining proportional-constrained parameter coupling adjustment with grid search, balancing optimization efficiency and result comprehensiveness. The experimental data demonstrate that the Falcon model achieves simultaneous excellence across task completion, path efficiency, and social compliance. However, due to limited time, the study only optimized three parameters, and the parameter levels were limited to predefined discrete values, failing to cover a broader parameter range, which may overlook potentially superior parameter combinations. Future research could conduct additional experiments or incorporate the model-decision making process visualization to further support this study.

References

[1] Zeying Gong, Tianshuai Hu, Ronghe Qiu, and Junwei Liang. From cognition to precognition: A future-aware framework for social navigation. *arXiv preprint arXiv:2409.13244*, 2024.

[2] Santhosh Kumar Ramakrishnan, Aaron Gokaslan, Erik Wijmans, Oleksandr Maksymets, Alexander Clegg, John M Turner, Eric Undersander, Wojciech Galuba, Andrew Westbury, Angel X Chang, Manolis Savva, Yili Zhao, and Dhruv Batra. Habitat-matterport 3d dataset (HM3d): 1000 large-scale 3d environments for embodied AI. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

[3] Zehui Lu, Wanxin Jin, Shaoshuai Mou, and B.D.O. Anderson. Cooperative tuning of multi-agent optimal control systems. *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 571–576, 2022.

[4] Monica and Parul Agrawal. A survey on hyperparameter optimization of machine learning models. *2024 2nd International Conference on Disruptive Technologies (ICDT)*, pages 11–15, 2024.

[5] Lingdong Kong, Shaoyuan Xie, Zeying Gong, Ye Li, Meng Chu, Ao Liang, Yuhao Dong, Tianshuai Hu, Ronghe Qiu, Rong Li, Hanjiang Hu, Dongyue Lu, Wei Yin, Wenhao Ding, Linfeng Li, Hang Song, Wenwei Zhang, Yuexin Ma, Junwei Liang, Zhedong Zheng, Lai Xing Ng, Benoit R. Cottereau, Wei Tsang Ooi, Ziwei Liu, Zhanpeng Zhang, Weichao Qiu, Wei Zhang, Ji Ao, Jiangpeng Zheng, Siyu Wang, Guang Yang, Zihao Zhang, Yu Zhong, Enzhu Gao, Xinhan Zheng, Xueting Wang, Shouming Li, Yunkai Gao, Siming Lan, Mingfei Han, Xing Hu, Dusan Malic, Christian Fruhwirth-Reisinger, Alexander Prutsch, Wei Lin, Samuel Schulter, Horst Possegger, Linfeng Li, Jian Zhao, Zepeng Yang, Yuhang Song, Bojun Lin, Tianle Zhang, Yuchen Yuan, Chi Zhang, Xuelong Li, Youngseok Kim, Sihwan Hwang, Hyeonjun Jeong, Aodi Wu, Xubo Luo, Erjia Xiao, Lingfeng Zhang, Yingbo Tang, Hao Cheng, Renjing Xu, Wenbo Ding, Lei Zhou, Long Chen, Hangjun Ye, Xiaoshuai Hao, Shuangzhi Li, Junlong Shen, Xingyu Li, Hao Ruan, Jinliang Lin, Zhiming Luo, Yu Zang, Cheng Wang, Hanshi Wang, Xijie Gong, Yixiang Yang, Qianli Ma, Zhipeng Zhang, Wenxiang Shi, Jingmeng Zhou, Weijun Zeng, Kexin Xu, Yuchen Zhang, Haoxiang Fu, Ruibin Hu, Yanbiao Ma, Xiyan Feng, Wenbo Zhang, Lu Zhang, Yunzhi Zhuge, Huchuan Lu, You He, Seungjun Yu, Junsung Park, Youngsun Lim, Hyunjung Shim, Faduo Liang, Zihang Wang, Yiming Peng, Guanyu Zong, Xu Li, Binghao Wang, Hao Wei, Yongxin Ma, Yunke Shi, Shuaipeng Liu, Dong Kong, Yongchun Lin, Huitong Yang, Liang Lei, Haoang Li, Xinliang Zhang, Zhiyong Wang, Xiaofeng Wang, Yuxia Fu, Yadan Luo, Djamahl Etchegaray, Yang Li, Congfei Li, Yuxiang Sun, Wenkai Zhu, Wang Xu, Linru Li, Longjie Liao, Jun Yan, Benwu Wang,

Xueliang Ren, Xiaoyu Yue, Jixian Zheng, Jinfeng Wu, Shurui Qin, Wei Cong, and Yao He. The RoboSense challenge: Sense anything, navigate anywhere, adapt across platforms. https://robosense2025.github.io, 2025.

[6] Lingdong Kong, Yaru Niu, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit Cottereau, Liangjun Zhang, Hesheng Wang, Wei Tsang Ooi, Ruijie Zhu, Ziyang Song, Li Liu, Tianzhu Zhang, Jun Yu, Mohan Jing, Pengwei Li, Xiaohua Qi, Cheng Jin, Yingfeng Chen, Jie Hou, Jie Zhang, Zhen Kan, Qiang Lin, Liang Peng, Minglei Li, Di Xu, Changpeng Yang, Yuanqi Yao, Gang Wu, Jian Kuai, Xianming Liu, Junjun Jiang, Jiamian Huang, Baojun Li, Jiale Chen, Shuang Zhang, Sun Ao, Zhenyu Li, Runze Chen, Haiyong Luo, Fang Zhao, and Jingze Yu. The RoboDepth challenge: Methods and advancements towards robust depth estimation. *arXiv preprint arXiv:2307.15061*, 2023.

[7] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit R. Cottereau, and Wei Tsang Ooi. RoboDepth: Robust out-of-distribution depth estimation under corruptions. In *Advances in Neural Information Processing Systems*, volume 36, pages 21298–21342, 2023.

[8] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Yaru Niu, Wei Tsang Ooi, Benoit R. Cottereau, Lai Xing Ng, Yuexin Ma, Wenwei Zhang, Liang Pan, Kai Chen, Ziwei Liu, Weichao Qiu, Wei Zhang, Xu Cao, Hao Lu, Ying-Cong Chen, et al. The RoboDrive challenge: Drive anytime anywhere in any condition. *arXiv preprint arXiv:2405.08816*, 2024.

[9] Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Benchmarking and improving bird's eye view perception robustness in autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(5):3878–3894, 2025.

[10] Shaoyuan Xie, Lingdong Kong, Yuhao Dong, Chonghao Sima, Wenwei Zhang, Qi Alfred Chen, Ziwei Liu, and Liang Pan. Are VLMs ready for autonomous driving? an empirical study from the reliability, data, and metric perspectives. In *IEEE/CVF International Conference on Computer Vision*, pages 6585–6597, 2025.

[11] Ye Li, Lingdong Kong, Hanjiang Hu, Xiaohao Xu, and Xiaonan Huang. Is your LiDAR placement optimized for 3D scene understanding? In *Advances in Neural Information Processing Systems*, volume 37, pages 34980–35017, 2024.

[12] Meng Chu, Zhedong Zheng, Wei Ji, Tingyu Wang, and Tat-Seng Chua. Towards natural language-guided drones: GeoText-1652 benchmark with spatial relation matching. In *European Conference on Computer Vision*, pages 213–231, 2024.

[13] Ao Liang et al. Perspective-invariant 3D object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 27725–27738, 2025.

[14] Rong Li, Yuhao Dong, Tianshuai Hu, Ao Liang, et al. 3EED: Ground everything everywhere in 3D. *arXiv preprint arXiv:2511.01755*, 2025.