Appendix

To complement the results presented in the paper, we provide additional experiments that evaluate the scalability and robustness of the proposed LLM-DiSC framework under more diverse settings. Each experiment is executed 40 times using the DeepSeek-r1 model.

A. Scalability Evaluation

We further assess the scalability of the proposed method by conducting experiments with 20 robots, as illustrated in Fig. 1, while keeping the environment and task configuration consistent with that in the paper.

TABLE I
DIFFERENT GROUP SIZES

Model	Size	Pass@k			Iterations
		k=1	k=5	k=10	
Unicycle	10	55%	95%	100%	2.13
	20	30%	85%	85%	2.18

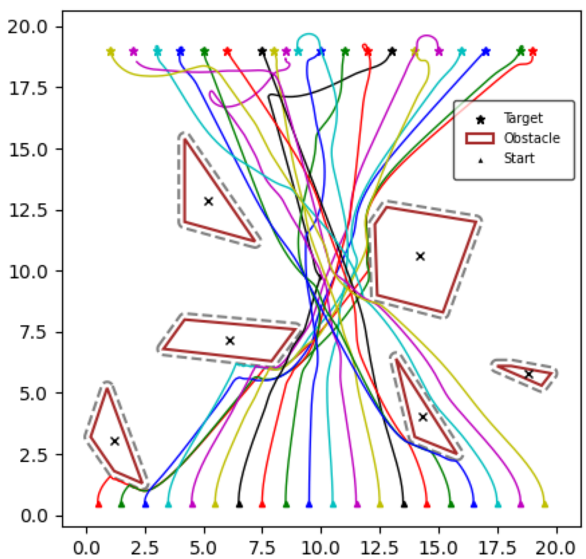

Fig. 1. Trajectories of 20 unicycle robots.

As shown in Table I, although the success rate at $k = 1$ decreases with a larger number of robots, the overall success rate still reaches 85% within a few iterations, highlighting the scalability of our method.

B. Robustness to Environment Variations

To evaluate the robustness of LLM-DiSC under varying environmental configurations, we conducted additional experiments with different obstacle distributions. Specifically, the sizes and placements of rectangular obstacles were adjusted to simulate both spacious and constrained environments. Experiments were conducted with obstacle gaps of 4, 3, and 2, respectively. Fig. 2 illustrates the robot trajectories in the scenario where the gap is set to 2.

As shown in Table II, while the success rate declines as the obstacle gap narrows, the proposed method remains effective in more challenging environments. Specifically, when the gap is reduced to 2, the method still achieves an 80% success rate after a few iterations, demonstrating its robustness under constrained conditions.

TABLE II
DIFFERENT OBSTACLE GAPS

Model	Obstacle Gap	Pass@k			Iterations
		k=1	k=5	k=10	
Unicycle	4	60%	100%	100%	1.9
	3	40%	90%	100%	2.8
	2	20%	50%	80%	4

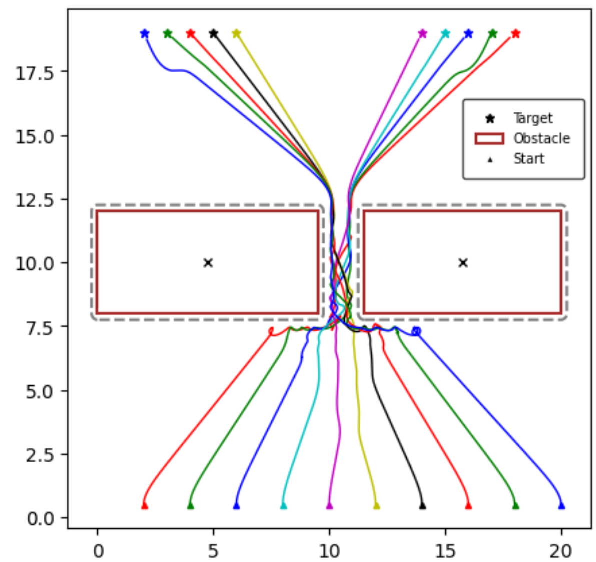

Fig. 2. Robots trajectories with obstacle gaps of 2