

LDP: A Local Diffusion Planner for Efficient Robot Navigation and Collision Avoidance

Wenhao Yu¹, Jie Peng², Huanyu Yang², Junrui Zhang¹, Yifan Duan³, Jianmin Ji^{3,4,*} and Yanyong Zhang^{3,4}

Abstract—The conditional diffusion model has been demonstrated as an efficient tool for learning robot policies, owing to its advancement to accurately model the conditional distribution of policies. The intricate nature of real-world scenarios, characterized by dynamic obstacles and maze-like structures, underscores the complexity of robot local navigation decision-making as a conditional distribution problem. Nevertheless, leveraging the diffusion model for robot local navigation is not trivial and encounters several under-explored challenges: (1) *Data Urgency* The complex conditional distribution in local navigation needs training data to include diverse policy in diverse real-world scenarios; (2) *Myopic Observation* Due to the diversity of the perception scenarios, diffusion decisions based on the local perspective of robots may prove suboptimal for completing the entire task, as they often lack foresight. In certain scenarios requiring detours, the robot may become trapped. To address these issues, our approach begins with an exploration of a diverse data generation mechanism that encompasses multiple agents exhibiting distinct preferences through target selection informed by integrated global-local insights. Then, based on this diverse training data, a diffusion agent is obtained, capable of excellent collision avoidance in diverse scenarios. Subsequently, we augment our **Local Diffusion Planner**, also known as **LDP** by incorporating global observations in a lightweight manner. This enhancement broadens the observational scope of **LDP**, effectively mitigating the risk of becoming ensnared in local optima and promoting more robust navigational decisions. Our experimental results demonstrated that the **LDP** outperforms other baseline algorithms in navigation performance, exhibiting enhanced robustness across diverse scenarios with different policy preferences and superior generalization capabilities for unseen scenarios. Moreover, we highlighted the competitive advantage of the **LDP** within real-world settings.

I. INTRODUCTION

With the rapid advancement of artificial intelligence and robotics, an increasing number of technologies are being integrated into motion planning for robot collision avoidance [1]. Many learning-based methods model the planning task as a conditional probability generation problem, where the robot’s action sequence is a latent variable with its prior distribution [2], [3]. The planning process is accomplished

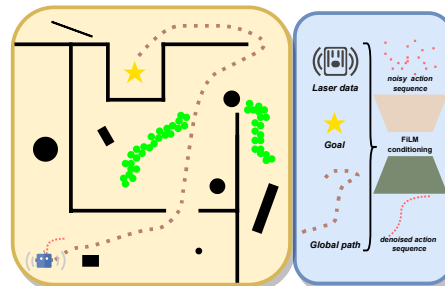


Fig. 1. The diagram illustrates the execution of our method. Obstacles are denoted by black circles and rectangles, while the trajectories of pedestrians are represented by green circles. The navigation target is marked by a yellow pentagram, and a brown dashed line delineates the global path from the robot’s starting point to its target. This system utilizes laser data, target point information (goal), and the global path to generate local action sequences in a classifier-free guidance approach.

by computing the posterior distribution based on conditions such as robot observations, final rewards, and constraints.

As is well known, the real-world environment for robot navigation is complex, encompassing various scenarios. Designing a specific policy for each scenario would require immense effort and lack the necessary flexibility and scalability. Therefore, an outstanding navigation policy must effectively handle diverse scenarios. Furthermore, the distribution of near-optimal expert policies often varies across different scenarios, highlighting the need for navigation policies capable of addressing diverse scenarios to exhibit a multimodal distribution. Given these requirements, collecting expert data from diverse scenarios to use as model training data and constructing models that better represent multimodal distributions will become crucial. Additionally, the limited local perspective of robots often fails to provide sufficient information for devising policies to tackle diverse scenarios. Robots frequently fall into suboptimal states during navigation tasks. This situation is commonly known as a local minimum problem [4].

Therefore, to address the challenges mentioned above, we have made two efforts in this paper: (1) we have collected expert policy data with multiple preferences under diverse scenarios and utilize the diffusion model, which has strong distribution modeling capabilities, to construct the policy model. (2) We incorporate global paths as an additional condition to guide the diffusion model, enhancing policies for better navigation through maze-like scenarios and similar environments. Specifically, we have collected expert policy data in three different types of scenarios, i.e., dense static, dynamic pedestrian, and maze-like. For each scenario, we have gathered expert policy data for two preferences, i.e., the original Soft Actor-Critic (SAC) policy [5] and the SAC

¹ Institute of Advanced Technology, University of Science and Technology of China (USTC), Hefei 230026, China wenhaoyu@mail.ustc.edu.cn

² School of Data Science, USTC, 230026, China

³ School of Computer Science and Technology, USTC, 230026, China

⁴ Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, 230088, China

* Corresponding author. jianmin@ustc.edu.cn

The work is partially supported by National Key R&D Program of China No.2023YFB4704500, Guangdong Province R&D Program 2020B0909050001, Anhui Province Development and Reform Commission 2021 New Energy and Intelligent Connected Vehicle Innovation Project, and the Youth Innovation Promotion Association of the Chinese Academy of Sciences.

policy guided by global paths. Using the Denoising Diffusion Probabilistic Models (DDPM) [6] algorithm, based on robot observations (local costmaps, goals, and global paths), we directly denoise and sample from the posterior trajectory distribution step by step, generating the final robot action sequence.

In our experiments, the LDP outperforms other baseline algorithms across various scenarios. It exhibits exceptional learning capabilities when dealing with data from mixed scenarios, displaying remarkable robustness. Moreover, it demonstrates enhanced performance in unseen scenarios, showcasing impressive zero-shot generalization abilities. The addition of the global path as conditional guidance enhances our policy’s capacity to comprehend sample distribution, leading to more forward-thinking policy outcomes. Furthermore, our approach leverages the advantages offered by expert policy data from two preferences, resulting in more effective decision-making in diverse navigation scenarios. Ultimately, we deploy our policy algorithm in real-world scenarios and on robotic platforms, thereby showcasing its competitive advantages.

The main contributions of our work are summarized as follows:

- The paper introduces LDP, a novel local planning algorithm for robotic collision avoidance, leveraging diffusion processes.
- We provide a dataset of expert policy based on 2D laser sensing, which spans expert data across three different types of scenarios and two different preferences.
- With global paths serving as additional guiding conditions, the diffusion model can better learn the distribution of expert data and make wiser decisions.
- We conducted extensive experiments demonstrating that LDP outperforms other baseline algorithms in terms of superior navigation performance, stronger robustness, and more profound generalization capabilities. Furthermore, we validated the effectiveness of the algorithm by deploying it on physical robots, thus highlighting its practical value.

II. RELATED WORK

A. Traditional Navigation Approaches

Traditional navigation systems [1], [7] typically adopt a hierarchical paradigm that combines global and local path planning with motion control. These methods can be broadly classified into three categories: search-based planners (e.g., hybrid A* [8], JPS [9]), sampling-based planners (e.g., PRM [10] and RRT [11]), and optimization-based planners (e.g., TEB [12]). Despite their prevalent usage, a significant amount of effort is needed to tune parameters for these methods to adapt to a wide range of scenarios.

B. Learning-Driven Navigation Approaches

Imitation learning (IL) is a process that involves learning from examples provided by an expert, typically in the form of decision-making data from human operators. IL has a

broad and profound influence in areas such as robotic navigation [5], [13], [14], manipulation [3], [15], and autonomous driving [16]–[18]. Depending on the structure of the policy model construction, these methods can be bifurcated into two categories:

1) *Explicit Policy*: These methods learn the mapping from observations to actions directly, guiding the policy learning process via a regression loss. Bojarski et al. [18] employed behavior cloning (BC) to train the end-to-end deep convolutional neural network (CNN) for autonomous driving. However, these policies often grapple with the effective modeling of multimodal data distributions. They usually map one policy to one scenario, and learning from data across multiple scenarios can lead to catastrophic forgetting. Furthermore, they often face challenges in generalizing to new, unseen scenarios.

2) *Implicit Policy*: These methods employ energy-based models (EBMs) [19] to represent action distributions. Each action is assigned an energy value, and the action prediction problem is transformed into an optimization problem to find the action with the lowest energy. This implicit design approach can more effectively represent the multimodal distribution of expert actions.

C. Diffusion Model for Robotic Decision Planning

In recent years, a growing number of works have leveraged diffusion models to perform intelligent agent decision planning tasks, including imitation learning [3], [20], [21] and reinforcement learning [22]–[24]. Diffuser [22] concatenates state-action sequences of a certain length into a two-dimensional array, employs the original DDPM method for unconditional sampling, and designs a classifier based on rewards, goals, and other information to guide the inference denoising process, thereby ensuring that the generated decision sequences comply with the respective constraints. The diffusion policy [3] represents an alternative modeling method that uses the robot’s visual observations as conditions and directly guides the generation of action sequences in a classifier-free manner. MPD [20] integrates diffusion models and optimization-based methods. With the planning of start and end points and various optimization costs as conditions, it guides the generation of global motion planning in static scenarios. However, LDP accomplishes local motion planning in diverse scenarios (static and dynamic environments). NoMaD [21] harnesses the powerful distribution modeling capability of diffusion models and employs a goal mask to achieve a single policy capable of completing both navigation and exploration tasks. In contrast, LDP evaluates the navigation performance of the policy model under a mixed expert trajectory distribution of multiple scenes and preferences and introduces additional global paths as conditions for guidance.

III. METHOD

In this section, we initially outline our approach to gathering training data focusing on the *Data Urgency* challenge. Subsequently, we delve into the specifics of developing the diffusion model for local planning using the data collected to

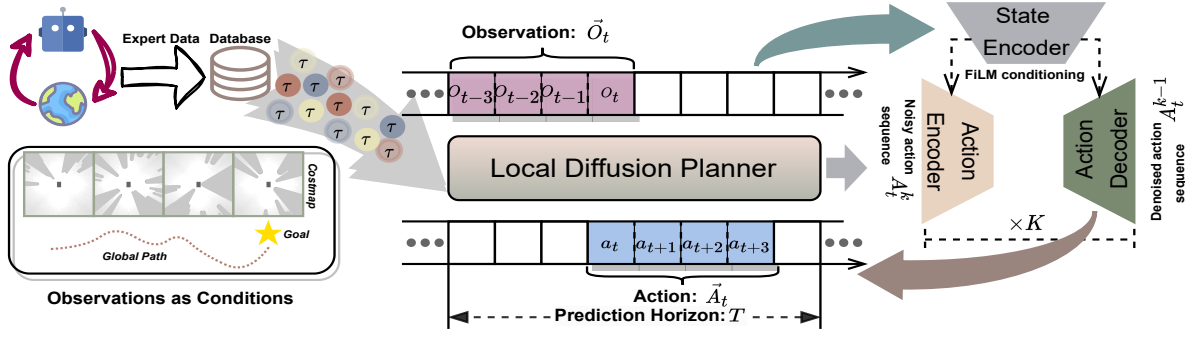


Fig. 2. An in-depth depiction of the entire process and the architecture of the local diffusion planner. The circles, marked with different colors and denoted as τ , represent expert data gathered from diverse scenarios, each reflecting various preferences through distinct levels of transparency. At time step t , the planner takes in observations \vec{O}_t from the past T_o steps and predicts the action sequence \vec{A}_t for the next T_a steps. During the training process, the loss is computed based on DDPM noise prediction over the entire sequence T . In the inference process, an action sequence can be generated for every K iterations of denoising.

address the *Myopic Observation* problem. The comprehensive workflow of our research is illustrated in Fig. 2.

A. Expert Policy Data

In recent times, the swift advancement of reinforcement learning has unveiled novel solutions for robotic motion planning challenges. We employ the SAC algorithm to learn sophisticated navigation policies. This robot navigation problem is conceptualized as a Markov Decision Process. The expert policy employs the same structure as that in [5], which has been proven to be effective. The policy’s state space is bifurcated into (1) egocentric costmaps with dimensions of 84×84 , produced by a 3D laser sensor, encompassing a complete 360-degree view; (2) the relative target pose. The action space is bidimensional and continuous, denoting the linear velocity and the angle of the front wheel for Ackermann steering robots. Given that the linear velocity can assume negative values, this expert navigation policy accommodates backward movement. Building upon the meticulously crafted state and action spaces, and enabling the robot to swiftly reach its target without any collisions, the reward function for the reinforcement learning policy has been thoughtfully devised as follows:

$$r_t = r_t^{goal} + r_t^{safe} + r_t^{shaping} + r_t^{back}, \quad (1)$$

where r_t is the sum of four parts, r_t^{goal} , r_t^{safe} , $r_t^{shaping}$ and r_t^{back} .

Specifically, r_t^{goal} represents the reward awarded to the robot upon successfully reaching the designated local target:

$$r_t^{goal} = \begin{cases} r_{arr}, & \text{if target is reached,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

r_t^{safe} denotes the penalty applied to the robot in the event of a collision:

$$r_t^{safe} = \begin{cases} r_{col}, & \text{if collision,} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

$r_t^{shaping}$ denotes the reward-shaping mechanism that converts sparse rewards into dense rewards, thereby hastening the training process of reinforcement learning algorithms.

The underlying design philosophy encompasses two main elements: (1) Actions diverting the robot from its local goal incur specific penalties; (2) To deter the agent from engaging in clever but impractical strategies—like spiraling near the target to gain higher rewards, counter to the necessity for swift navigation task completion—an additional fixed-value penalty is levied on each action.

$$r_t^{shaping} = \varepsilon(\|p_{t-1} - p_g\|_2 - \|p_t - p_g\|_2) - r_{step}, \quad (4)$$

where p_t is the position of the robot at time t , p_g is the position of the target point, r_{step} is a fixed-value penalty, and ε is a hyper-parameter.

The term r_t^{back} denotes a fixed-value penalty applied when the robot executes a reverse maneuver. This mechanism is designed to prompt the robot to reverse only when needed, instead of persistently backing up from start to end, thus aligning more closely with practical applications.

$$r_t^{back} = \begin{cases} r_{reverse}, & \text{if } v < 0, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where $r_{reverse}$ is also a fixed-value penalty and v is robot’s linear velocity at time t . In our experiments, we set $r_{arr} = 500$, $r_{col} = -500$, $\varepsilon = 200$, $r_{step} = 5$ and $r_{reverse} = -10$.

Within the scope of our research, for every experimental scenario delineated in the paper, we meticulously trained an expert policy, subjecting each to an extensive training regimen spanning three million steps. Leveraging these expert policies, we collated datasets reflecting two strategic preferences: (1) The original expert policy data. While the reward-shaping mechanism outlined in Eq. (4) offers significant benefits, such as expedited training, diminished complexity, and the facilitation of rapid task completion, it tends to engender policies that overly greedy, lacking in long-term planning, and unsuitable for tasks in maze-like scenarios. (2) Expert policy data guided by global path planning. We utilized the A* algorithm to search for global paths and provided local targets to the expert policies through a sliding window approach. This type of policy overcomes the aforementioned issues but lacks efficiency in task completion, leading to detours in certain scenarios. Our objective is to enhance the policy model’s performance by

Algorithm 1 Expert Data Collecting

Collecting data on two types of preferred expert policies in each scenario

```

1: Initialize the collection of expert data  $\mathcal{T}^{train}$ 
2: for  $\pi_s \in \Pi$  do
3:   Initialize episodes of collected data  $N$ , global path
   planning algorithm  $A^*$  and size of sliding window  $w$ 
4:   Clear the trajectory buffer  $\mathcal{T}$ 
5:   for episode = 1 to  $N/2$  do
6:     for step  $t = 1$  to  $T_{ep}$  do
7:        $gp_t = A^*(p_t, g_t)$ 
8:        $a_t = \pi_s(s_t), s_t = \{c_t, g_t\}$ 
9:        $\mathcal{T} \leftarrow \mathcal{T} \cup \{c_t, g_t, gp_t, a_t\}$ 
10:    end for
11:  end for
12:  for episode =  $N/2 + 1$  to  $N$  do
13:    for step  $t = 1$  to  $T_{ep}$  do
14:       $gp_t = A^*(p_t, g_t)$ 
15:       $lg_t = gp_t[w]$ 
16:       $a_t = \pi_s(s_t), s_t = \{c_t, lg_t\}$ 
17:       $\mathcal{T} \leftarrow \mathcal{T} \cup \{c_t, g_t, gp_t, a_t\}$ 
18:    end for
19:  end for
20:   $\mathcal{T}^{train} \leftarrow \mathcal{T}^{train} \cup \mathcal{T}$ 
21: end for

```

learning from the mixed data of these two types of preferred policies.

B. Local Diffusion Planner

The objective of our research is to develop a local motion planning algorithm for robots, by leveraging multimodal expert strategy data encompassing a variety of environments and preferences. Therefore, we formulate the task as a conditional generation problem via diffusion model:

$$\max_{\theta} \mathbb{E}_{\tau \sim \mathcal{T}, \tau = \{\mathcal{O}, A\}} \log p_{\theta}(A^0 | \mathcal{O}),$$

where τ is the expert trajectory data used for training, A^0 is the final generated sequence of actions, \mathcal{O} represents the robot's observations serving as conditions for the diffusion model, and p_{θ} refers to the reverse denoising process within the diffusion model.

In this paper, we structure the training data according to a receding-horizon action prediction framework, as outlined in [3]. Here, $\tau_t = \{\mathcal{O}_t, A_t\}$ signifies the chosen robot training trajectory at time t , where \mathcal{O}_t and A_t respectively denote the corresponding observation sequence $\{o_{t-(T_o-1)}, \dots, o_{t-1}, o_t\}$ and action sequence $\{a_t, a_{t+1}, \dots, a_{t+(T_a-1)}\}$ for that trajectory. T_o and T_a represent the lengths of the observation and action sequences, respectively, with the length of the trajectory τ , denoted as T_{τ} , being equal to $T_o + T_a - 1$. It should be emphasized that, throughout this paper, all superscripts associated with time instances refer to diffusion time steps, and all subscripts associated with time instances pertain to motion time steps.

As we have discussed earlier, \mathcal{O} consists of three parts: costmaps C , goals G , and global paths GP , i.e., $\mathcal{O} = \{C, G, GP\}$. It is important to note that here, the global paths

Algorithm 2 LDP Training and Evaluation

LDP Training

```

1: initialize training iterations  $N$ , batch size  $M$ 
2: for  $\mathcal{T} \subset \mathcal{T}^{train}$  do
3:   for  $n = 1$  to  $N$  do
4:      $\mathcal{B} \sim \{C_i, G_i, GP_i, A_i^0\}_{i=1}^M$ 
5:     randomly select a diffusion timestep  $k \sim \mathcal{U}(1, K)$ 
     and acquire noisy action sequences  $A^k$ 
6:     compute loss function  $\mathcal{L}(\theta)$  and update LDP
7:   end for
8: end for

```

LDP Evaluation

```

1: initialize sample scale  $\beta$  and classifier-free guidance
   scale  $\omega$ 
2: for episode = 1 to  $E$  do
3:   for step  $t = 1$  to  $T_{ep}$  do
4:     sample  $A^K \sim \mathcal{N}(0, \beta I)$ 
5:     for  $k = K$  to 1 do
6:        $\bar{\epsilon} = \epsilon_{\theta}(A^k, k) + \omega(\epsilon_{\theta}(A^k, \mathcal{O}, k) - \epsilon_{\theta}(A^k, k))$ 
7:        $(\mu_{k-1}, \sum_{k-1}) \leftarrow \text{Denoise}(A^k, \bar{\epsilon})$ 
8:        $A^{k-1} \sim \mathcal{N}(\mu_{k-1}, \beta \sum_{k-1})$ 
9:     end for
10:    generate the final action sequence  $A^0$ 
11:    execute each action within the  $A^0$  sequence,
    requiring  $T_a$  time steps
12:    obtain the next robot observation for the subse-
    quent decision-making
13:  end for
14: end for

```

act merely as conditions for the diffusion process, rather than providing an additional local target for the policy during inference, as is done in our method of collecting expert policy data.

The training of the model leverages the DDPM algorithm, incorporating classifier-free guidance [25] for its execution. The ultimate action sequence, A^0 , is derived by initially sampling from Gaussian noise A^K . Through adjacent diffusion time steps, from A^k to A^{k-1} , the action sequence is subjected to noise perturbation. This methodical alteration facilitates the denoising and refinement of the sequence itself. The perturbation noise $\bar{\epsilon}_{\theta}(A^k, \mathcal{O}, k)$ is defined as follows [26]:

$$\epsilon_{\theta}(A^k, k) + \omega(\epsilon_{\theta}(A^k, \mathcal{O}, k) - \epsilon_{\theta}(A^k, k)), \quad (6)$$

where $\omega > 1$ represents the guidance scale, a factor in identifying the expert action sequence that optimally aligns with the robot's current observations from the expert dataset, and ϵ_{θ} denotes the noise model. Eq. (6) is inspired by an implicit classifier $p(\mathcal{O}|A^k) \propto p(A^k|\mathcal{O})/p(A^k)$. The gradient of the logarithmic probability of this classifier $\nabla_{A^k} \log p(\mathcal{O}|A^k) \propto \nabla_{A^k} \log p(A^k|\mathcal{O}) - \nabla_{A^k} \log p(A^k) \propto \epsilon(A^k, \mathcal{O}, k) - \epsilon(A^k, k)$ is utilized to guide the generation of $\bar{\epsilon}$. In the training phase, we aim to optimize the reverse diffusion process p_{θ} , which is parameterized by the noise model, pursuing the following objective:

$$\mathcal{L}(\theta) := \mathbb{E}_{k \sim \mathcal{U}(1, K), \epsilon \sim \mathcal{N}(0, I)} [\|\epsilon - \epsilon_{\theta}(A^k, \mathcal{O}, k)\|^2]. \quad (7)$$

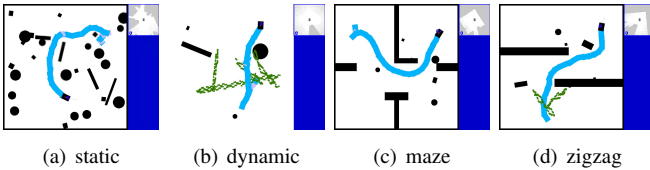


Fig. 3. Four different simulation scenarios are displayed. The black rectangles and circles are obstacles, the green dots represent pedestrian trajectories, and the blue box on the right shows the robot’s local sensor map.

In the inference phase, the ultimate expert action sequence is derived via a stepwise sampling process, as outlined in the formula $A^{k-1} \sim \mathcal{N}(\mu_\theta(A^{k-1}, \mathcal{O}, k-1), \beta \sum_{k-1})$.

The design of the diffusion model network structure follows the approach of work [3], which has been validated as efficient and outstanding.

IV. EXPERIMENTS

A. Environments, baselines, and metrics

In this study, we meticulously gather expert data reflecting two preferences across three diverse scenarios, subsequently training corresponding navigation policies. Besides, we evaluate the performance of models trained with mixed policy data and assess their zero-shot generalization capabilities in unseen scenarios. Fig. 3 illustrates environments for four different scenarios respectively, where a, c, and d represent enclosed scenarios, while b represents an open scenario.

We compare LDP with the following baseline algorithms: LSTM-GMM [27], IBC [28], and DT [29]. We strive to adjust the training details to optimize the performance of the methods as well as possible. Notice that, the input content for these baseline algorithms consists of $\{C, G, GP\}$, which aligns with the guiding conditions of LDP.

We introduce five distinct metrics to systematically assess the navigation performance of all methods within a range of scenarios.

- *Success rate (SUCC)*: the rate of episodes in which the robot reaches the target pose without collision.
- *Collision rate (Coll)*: the rate of episodes in which the robot collides.
- *Stuck rate (Stuck)*: the rate of episodes in which the robot sticks.
- *Average time (TIME)*: the average cost time at each success episode.
- *Success weighted by Path Length (SPL)* [30]: assess navigation performance by examining both the success rate and the length of the robot’s trajectory.

B. Experiments on simulation scenarios

In the following, the experimental results for various methods are presented based on the average outcomes from 1,000 randomly generated environments for each scenario. The training data for each scenario consists of 2,000 episodes.

1) *Comparative experiments*: Tab. II presents the navigation performance of all algorithms outlined in our paper. Each entry comprises three values: the performance metrics of models trained on single-scenario data (2000 episodes),

TABLE I
HYPERPARAMETERS

Hyperparameter	Value
Hyperparameters for expert policy	
Training steps	3×10^6
Encoder&Actor&Critic learning rate	5×10^{-4}
Temperature learning rate	10^{-4}
Init temperature	0.01
Batch size	1024
Obstacle collision radius for A*	0.5
Size of the sliding window	20
Hyperparameters for local diffusion planner	
Training epochs	300
Episodes of each scenario expert data	2000
Batch Size	1024
Learning rate	10^{-4}
Weight decay	10^{-6}
Prediction horizon T	8
Observation horizon T_o	2
Action horizon T_a	4
Number of diffusion iterations	100

the performance metrics of models trained on single-scenario data (6000 episodes), and the performance metrics of models trained on mixed data from three scenarios (6000 episodes). The LDP algorithm surpasses other baseline algorithms in success rate, runtime, and SPL. Comparing the first and last two items, we can conclude that increasing the quantity of training data to some extent can enhance the navigation performance across a wide range of algorithms and scenarios. Contrasting the second and third items, we can infer that enriching the diversity of data while maintaining the same training data quantity poses a challenge for baseline algorithms, resulting in a decline in navigation performance. However, thanks to the superior distribution modeling capability of LDP, its navigation performance may remain stable or even improve. An interesting observation is that LSTM-GMM, when trained on data from a single dense static scenario, cannot accurately reach the target point. It tends to wander near the target point, resulting in a high stuck rate. Even increasing the training data volume does not resolve this issue. However, introducing data from diverse scenarios can significantly enhance the performance of LSTM-GMM. Notably, in the zigzag scenario, the performance of LDP underscores its robust zero-shot generalization capability in unseen scenarios.

2) *Ablation study*: In the design of the LDP approach, we have additionally introduced global paths GP as conditions for the diffusion model to guide wiser decision-making in complex scenarios. Tab. III illustrates that LDP outperforms LDP without GP , particularly in dense static and maze-like scenarios. In Fig. 4, we showcase a maze-like scenario where LDP effectively navigates around maze walls and reaches the target point, while LDP without GP gets obstructed by the walls, failing to complete the navigation task. The experimental results suggest that additional GP conditions can better assist LDP in modeling data distributions and guiding wiser decision-making.

In our paper, we collected expert data reflecting two different preferences. Therefore, in Tab. IV, we explore the

TABLE II
PERFORMANCE OF DIFFERENT METHODS

Methods	Succ(%) \uparrow	Coll(%) \downarrow	Stuck(%) \downarrow	Time(t) \downarrow	SPL \uparrow
Static Scenario (28 Obstacles)					
LSTM-GMM	0.475-0.427-0.639	0.206-0.110-0.201	0.319-0.463-0.161	44.797-42.964-39.113	0.209-0.202-0.310
IBC	0.635-0.829-0.741	0.164-0.127-0.073	0.201-0.044-0.186	41.612-31.971-37.109	0.299-0.464-0.404
DT	0.679-0.772-0.752	0.134-0.135-0.110	0.187-0.093-0.138	36.891-30.092-36.242	0.347-0.441-0.379
LDP	0.952-0.955-0.957	0.020-0.014-0.014	0.028-0.031-0.029	24.365-24.146-23.636	0.648-0.649-0.654
Dynamic Scenario (4 Obstacles & 4 Pedestrians)					
LSTM-GMM	0.693-0.715-0.657	0.265-0.251-0.270	0.042-0.034-0.066	31.543-28.012-34.716	0.396-0.443-0.338
IBC	0.422-0.719-0.732	0.451-0.253- 0.227	0.127-0.028-0.041	52.150-34.152-32.657	0.191-0.369-0.415
DT	0.508-0.744-0.708	0.487-0.243-0.258	0.005 -0.013-0.034	20.738- 20.980 -23.045	0.375- 0.533 -0.481
LDP	0.740-0.748-0.755	0.255-0.242-0.238	0.005-0.010-0.007	20.480-22.272-20.943	0.533-0.523-0.536
Maze-like Scenario (Maze Map & 6 Obstacles)					
LSTM-GMM	0.418-0.796-0.746	0.459-0.152-0.198	0.123-0.052-0.056	52.189-32.517-33.521	0.190-0.407-0.371
IBC	0.715-0.812-0.777	0.221-0.144-0.114	0.064-0.044-0.109	25.692-22.052-30.773	0.427-0.507-0.426
DT	0.686-0.861-0.785	0.237-0.106-0.178	0.077-0.033-0.037	27.657-20.092-30.231	0.395-0.564-0.421
LDP	0.915-0.930-0.930	0.065-0.057-0.058	0.020-0.013-0.012	20.248-19.261-19.163	0.609-0.637-0.631
Zigzag Scenario (Unseen)					
LSTM-GMM	0.500	0.439	0.061	38.002	0.271
IBC	0.532	0.330	0.138	33.490	0.333
DT	0.464	0.451	0.085	32.464	0.283
LDP	0.720	0.267	0.013	22.264	0.510

TABLE III
PERFORMANCE GUIDED BY GLOBAL PATH CONDITIONS

Methods	Succ(%) \uparrow	Coll(%) \downarrow	Stuck(%) \downarrow	Time(t) \downarrow	SPL \uparrow
Static Scenario (28 Obstacles)					
LDP	0.952	0.020	0.028	24.365	0.648
LDP <i>w.o. GP</i>	0.934	0.017	0.049	25.264	0.632
Dynamic Scenario (4 Obstacles & 4 Pedestrians)					
LDP	0.740	0.255	0.005	20.480	0.533
LDP <i>w.o. GP</i>	0.737	0.248	0.015	21.903	0.521
Maze-like Scenario (Maze Map & 6 Obstacles)					
LDP	0.915	0.065	0.020	20.248	0.609
LDP <i>w.o. GP</i>	0.877	0.095	0.028	19.689	0.595

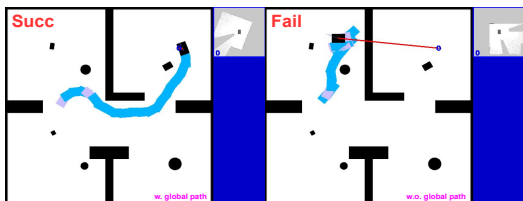


Fig. 4. Global Path Influence: Navigation Success vs. Failure in One Scene

impact of training data composition with varying preferences on model performance. All experiments are trained on the same amount of data for the same training steps. The “No.1 expert”, “No.2 expert”, and “mixed expert” respectively represent data collected from original SAC (2000 episodes), SAC guided by global paths (2000 episodes), and a mixture of both, with 1000 episodes each. Our experimental results demonstrate that mixed preference data can enhance the policy’s performance in dynamic and maze-like scenarios. While in static scenes, the success rate of the mixed data policy slightly lags behind that of the No.2 expert data policy, it outperforms in terms of average execution time and SPL. In the maze-like scenario presented in Fig. 5, LDP with mixed expert data efficiently completes the navigation task. Conversely, LDP with No.2 expert data initially encounters obstacles due to erroneous decisions but eventually succeeds after prolonged exploration, while LDP with No.1 expert data remains trapped and unable to finish the task. Providing data with mixed preferences is meaningful as it allows the policy to leverage the advantages of learning different preference policies, resulting in more efficient and accurate task completion.

TABLE IV
PERFORMANCE TRAINED BY EXPERT DATA WITH TWO PREFERENCES

Methods	Succ(%) \uparrow	Coll(%) \downarrow	Stuck(%) \downarrow	Time(t) \downarrow	SPL \uparrow
Static Scenario (28 Obstacles)					
LDP <i>w. No.1 expert</i>	0.908	0.027	0.065	23.667	0.633
LDP <i>w. No.2 expert</i>	0.956	0.012	0.032	24.877	0.642
LDP <i>w. mixed expert</i>	0.952	0.020	0.028	24.365	0.648
Dynamic Scenario (4 Obstacles & 4 Pedestrians)					
LDP <i>w. No.1 expert</i>	0.732	0.248	0.020	20.499	0.530
LDP <i>w. No.2 expert</i>	0.726	0.266	0.008	22.777	0.494
LDP <i>w. mixed expert</i>	0.740	0.255	0.005	20.480	0.533
Maze-like Scenario (Maze Map & 6 Obstacles)					
LDP <i>w. No.1 expert</i>	0.868	0.099	0.033	19.800	0.589
LDP <i>w. No.2 expert</i>	0.873	0.094	0.033	19.904	0.592
LDP <i>w. mixed expert</i>	0.915	0.065	0.020	20.248	0.609

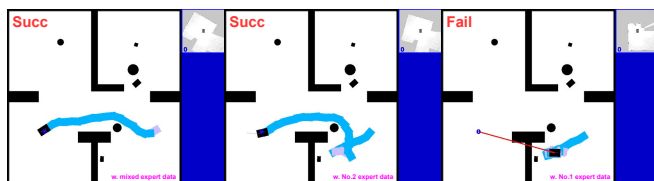


Fig. 5. Performance of policies learned under expert data with different preferences.

In the field of robot navigation, global paths can provide global information to local planners to assist them in making better decisions. Similarly, global maps *GM* can serve this role as well. We conducted experimental comparisons of two diffusion model conditions in Tab. V. All experiments are trained for the same number of steps. LDP with *GP* demonstrates higher success rates and SPL with slightly longer average running times in dense static and maze-like complex structured scenarios. In dynamic scenarios, LDP with *GP* exhibits slightly lower success rates than LDP with *GM*, but it boasts shorter average running times and higher SPL values. The experimental results suggest that *GP* offers more direct and efficient guidance compared to *GM*, particularly excelling in complex structured scenarios such as dense static and maze-like environments.

C. Deploy to real-world Ackermann steering robot

We’ve implemented the LDP algorithm on a real Ackerman robot to evaluate its performance in real-world situations. This experimental robot is built on the Agilex

TABLE V

PERFORMANCE GUIDED BY DIFFERENT GLOBAL CONDITIONS

Methods	Succ(%) \uparrow	Coll(%) \downarrow	Stuck(%) \downarrow	Time(t) \downarrow	SPL \uparrow
Static Scenario (28 Obstacles)					
LDP w. GP	0.952	0.020	0.028	24.365	0.648
LDP w. GM	0.930	0.026	0.044	24.251	0.634
Dynamic Scenario (4 Obstacles & 4 Pedestrians)					
LDP w. GP	0.740	0.255	0.005	20.480	0.533
LDP w. GM	0.744	0.238	0.018	21.777	0.527
Maze-like Scenario (Maze Map & 6 Obstacles)					
LDP w. GP	0.915	0.065	0.020	20.248	0.609
LDP w. GM	0.903	0.077	0.020	19.596	0.608

hunter2.0 chassis and features a 32-line 3D RoboSense LiDAR. It's powered by an RTX 3090 GPU and measures $0.95m \times 0.75m \times 1.45m$ in size. For a more detailed overview of the simulation and physical experiment results, please refer to our video.



Fig. 6. Schematic diagram of real robots and test scenarios.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a novel local diffusion planner, LDP, designed for robot collision avoidance, employing diffusion processes. Our approach involves gathering expert data representing two different preferences from three diverse scenarios to train our model. Our series of experiments indicate that LDP exhibits better navigation performance, and stronger robustness in learning from expert data across scenes. Additionally, it can learn wiser and more visionary policies from multi-preference expert data and demonstrate strong generalization ability in unseen scenarios. Our real-world experiments also demonstrate the practical value of LDP.

In future work, two aspects could be further explored: (1) Collecting higher quality and more diverse expert data to train superior navigation policies; (2) Improving the real-time performance of LDP. Using flow-based methods [31] or consistency models [32] instead of DDPM to accelerate diffusion model sampling would facilitate the practical deployment of the LDP method.

REFERENCES

- [1] X. Xiao, B. Liu, G. Warnell, and P. Stone, "Motion planning and control for mobile robot navigation using machine learning: A survey," *Autonomous Robots*, 2022.
- [2] Z. J. Cui, Y. Wang, N. M. M. Shafiullah, and L. Pinto, "From play to policy: Conditional behavior generation from uncurated robot data," *arXiv preprint arXiv:2210.10047*, 2022.
- [3] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *arXiv preprint arXiv:2303.04137*, 2023.
- [4] M. Wang and J. N. Liu, "Fuzzy logic-based real-time robot navigation in unknown environment with dead ends," *RAS*, 2008.
- [5] W. Yu, J. Peng, Q. Qiu, H. Wang, L. Zhang, and J. Ji, "Pathrl: An end-to-end path generation method for collision avoidance via deep reinforcement learning," *arXiv preprint arXiv:2310.13295*, 2023.
- [6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in NIPS*, 2020.

- [7] J. Peng, Y. Chen, Y. Duan, Y. Zhang, J. Ji, and Y. Zhang, "Towards an online rrt-based path planning algorithm for ackermann-steering vehicles," in *IEEE ICRA*, 2021.
- [8] K. Kurzer, "Path planning in unstructured environments: A real-time hybrid a* implementation for fast and deterministic path generation for the kth research concept vehicle," Master's thesis, 2016.
- [9] D. Harabor and A. Grastien, "Online graph pruning for pathfinding on grid maps," in *AAAI*, 2011.
- [10] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *IEEE TRO*, 1996.
- [11] S. M. LaValle, J. J. Kuffner, B. Donald, *et al.*, "Rapidly-exploring random trees: Progress and prospects," *Algorithmic and computational robotics: new directions*, 2001.
- [12] C. Rösmann, W. Feiten, T. Wösch, F. Hoffmann, and T. Bertram, "Trajectory modification considering dynamic constraints of autonomous robots," in *ROBOTIK*, 2012.
- [13] S. Yao, G. Chen, Q. Qiu, J. Ma, X. Chen, and J. Ji, "Crowd-aware robot navigation for pedestrians with multiple collision avoidance strategies via map-based deep reinforcement learning," in *IEEE/RSS IROS*, 2021.
- [14] Q. Qiu, S. Yao, J. Wang, J. Ma, G. Chen, and J. Ji, "Learning to socially navigate in pedestrian-rich environments with interaction capacity," *arXiv preprint arXiv:2203.16154*, 2022.
- [15] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," in *arXiv*, 2024.
- [16] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," *Advances in NIPS*, 1988.
- [17] G. You, X. Chu, Y. Duan, J. Peng, J. Ji, Y. Zhang, and Y. Zhang, "P 3 o: Transferring visual representations for reinforcement learning via prompting," in *IEEE ICME*, 2023.
- [18] B. *et al.*, "End to end learning for self-driving cars," *arXiv:1604.07316 [cs]*, 2016.
- [19] Y. Du and I. Mordatch, "Implicit generation and generalization in energy-based models," *arXiv preprint arXiv:1903.08689*, 2019.
- [20] J. Carvalho, A. T. Le, M. Baierl, D. Koert, and J. Peters, "Motion planning diffusion: Learning and planning of robot motions with diffusion models," in *IEEE/RSS IROS*, 2023.
- [21] A. Sridhar, D. Shah, C. Glossop, and S. Levine, "Nomad: Goal masked diffusion policies for navigation and exploration," *arXiv preprint arXiv:2310.07896*, 2023.
- [22] M. Janner, Y. Du, J. B. Tenenbaum, and S. Levine, "Planning with diffusion for flexible behavior synthesis," *arXiv preprint arXiv:2205.09991*, 2022.
- [23] Z. Zhu, H. Zhao, H. He, Y. Zhong, S. Zhang, Y. Yu, and W. Zhang, "Diffusion models for reinforcement learning: A survey," *arXiv preprint arXiv:2311.01223*, 2023.
- [24] H. He, C. Bai, K. Xu, Z. Yang, W. Zhang, D. Wang, B. Zhao, and X. Li, "Diffusion model is an effective planner and data synthesizer for multi-task reinforcement learning," *Advances in NIPS*, 2024.
- [25] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.
- [26] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741*, 2021.
- [27] A. Mandelkar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, "What matters in learning from offline human demonstrations for robot manipulation," *arXiv preprint arXiv:2108.03298*, 2021.
- [28] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and J. Tompson, "Implicit behavioral cloning," in *CoRL*. PMLR, 2022.
- [29] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, "Decision transformer: Reinforcement learning via sequence modeling," *Advances in NIPS*, 2021.
- [30] N. Yokoyama, S. Ha, and D. Batra, "Success weighted by completion time: A dynamics-aware evaluation criteria for embodied navigation," in *IEEE/RSS IROS*, 2021.
- [31] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," *arXiv preprint arXiv:2209.03003*, 2022.
- [32] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," 2023.