MATT–Diff: Multimodal Active Target Tracking by Diffusion Policy

Saida Liu[1] 2215072T@STU.KOBE-U.AC.JP
Nikolay Atanasov[2] NATANASOV@UCSD.EDU
Shumon Koga[1] KOGA@HARBOR.KOBE-U.AC.JP
[1]*Department of Computer Science and Systems Engineering, Kobe University, Kobe, Hyogo, 657-8501, Japan*
[2]*Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, 92093, USA*

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

This paper proposes MATT-Diff: Multimodal Active Target Tracking by Diffusion Policy, a control policy for active multi-target tracking using a mobile agent. The policy enables multiple behavior modes for the agent, including exploration, tracking, and target reacquisition, without prior knowledge of the target numbers, states, or dynamics. Effective target tracking demands balancing exploration for undetected or lost targets with exploitation, i.e., uncertainty reduction, of detected but uncertain ones. We generate a demonstration dataset from three expert planners including frontier-based exploration, an uncertainty-based hybrid planner switching between frontier-based exploration and RRT* tracking, and a time-based hybrid planner switching between exploration and target reacquisition based on target detection time. Our control policy utilizes a vision transformer for egocentric map tokenization and an attention mechanism to integrate variable target estimates represented by Gaussian densities. Trained as a diffusion model, the policy learns to generate multimodal action sequences through a denoising process. Evaluations demonstrate MATT-Diff's superior tracking performance against other learning-based baselines in novel environments, as well as its multimodal behavior sourced from the multiple expert planners. Our implementation is available at https://github.com/CINAPSLab/MATT-Diff.

Keywords: Diffusion policy, active target tracking, reinforcement learning

1. INTRODUCTION

The ability to track targets actively using mobile robots has widespread applications in security and surveillance, environmental monitoring, and search and rescue (Queralta et al., 2020). The complexity of this task stems from several interconnected challenges: (i) effectively exploring the environment to detect and acquire targets; (ii) accurately predicting target motion to sustain tracking under noisy observations and unmodeled dynamics; (iii) planning agent trajectories that actively minimize target uncertainty under limited field of view (FoV) constraints; and (iv) adaptively modifying target beliefs and planning strategies when targets are not detected where they are expected to be. Optimization- and sampling-based planners yield single, deterministic solutions, struggling with the multi-modal decisions inherent in uncertain real-world scenarios, particularly the exploration-exploitation dilemma. Learning policies that handle multi-modal action distributions is crucial in the contex of target tracking.

This work was supported by JST-Research & Development Program for Next Generation Edge AI Semiconductors Japan Grant Number JPMJES2514

This paper proposes MATT-Diff: Multi-Modal Active Target Tracking by Diffusion Policy. Our approach addresses agent control in complex target tracking scenarios without prior knowledge of target numbers, states, or dynamics in known environments with obstacles and limited FoV. The target states are estimated using Kalman filters with updates only when targets are within the FoV. We generate a diverse dataset of demonstrations from three planners: a frontier-based exploration searching for undetected targets and two hybrid planners combining exploration with RRT* tracking of uncertain targets based on uncertainty or detection time. Our MATT-Diff control policy employs a vision transformer network, processing egocentric occupancy-grid maps via tokenization and an attention mechanism robustly integrating a variable number of target estimates. The policy is trained using diffusion to match the multi-modal action sequences from expert demonstrations through a denoising process, embedding the processed tokens. Through numerical experiments, we demonstrate MATT-Diff's superior performance to other learning-based baselines across randomized numbers and motions of targets in an environment map not included in training data.

2. RELATED WORK

We review related work on target tracking and diffusion models.

2.1. Active Target Tracking

Active target tracking has a rich history in robotics, originating from early studies on pursuit-evasion games (LaValle et al., 1997) and later evolving into the area of sensor planning and management (Spletzer and Taylor, 2003). To handle sensor noise and uncertainty in a probabilistic formulation, information-theoretic approaches were introduced, e.g. in Le Ny and Pappas (2009); Hero and Cochran (2011) proposing nonmyopic trajectory planning for targets with known linear Gaussian dynamics. This approach has since been extended to multi-robot systems, addressing challenges such as resilience, anytime planning, asymptotic optimality, and communication constraints (Zhou et al., 2018; Schlotfeldt et al., 2018; Kantaros et al., 2019; Wang et al., 2024b). An informative path planning framework has been proposed in Meera et al. (2019) for target search and in Sudha et al. (2025) for dynamic occupancy mapping. Low-level control for handling occlusion in maintaining target's visibility has been proposed in Zhou et al. (2025). For a comprehensive overview of target tracking, particularly in the context of unmanned aerial vehicles, see Sun et al. (2024).

More recently, the field has shifted towards learning-based methods that leverage deep neural networks. Deep reinforcement learning (RL) has been employed to train policies for active target tracking, enabling robots to learn complex behaviors directly from experience or simulation by model-free (Jeong et al., 2021), model-based (Yang et al., 2023), and graph-neural network (GNN) approaches (Tzes et al., 2023). While these methods demonstrate effective learning capabilities, they often produce unimodal policies that can struggle in scenarios requiring diverse, context-dependent strategies. Our work builds on learning-based methods by specifically addressing the need for multi-modal action generation, a gap not fully explored by existing RL approaches.

2.2. Diffusion Models

Diffusion models have recently emerged as a powerful representaiton of multi-modal action distributions in robotics (Wang et al., 2023), following their success in generative AI (Song et al., 2021). Behavior cloning (BC; Torabi et al., 2018) relies on supervised imitation of expert demonstrations,

but its unimodal policy representation often collapses multi-modal action distributions into averaged behaviors, leading to degraded performance even without covariate shift (Zare et al., 2024).

Through iterative denoising, diffusion models learn to generate coherent action sequences conditioned on past observations, enabling policies to reproduce diverse expert strategies. Diffusion policies have demonstrated remarkable success in high-dimensional control tasks, including visuomotor control (Chi et al., 2023), language-conditioned multi-task policies (Yan et al., 2024), and dexterous robot manipulation (Song et al., 2025). Recent work has also incorporated geometric symmetries such as rotation into diffusion policies to improve generalization and sample efficiency in tasks with spatial invariances (Wang et al., 2024a). Integrating diffusion models with reinforcement learning has been explored both in online training (Ding et al., 2024) and in fine-tuning of pre-trained models (Ren et al., 2025; Li et al., 2024; Wagenmaker et al., 2025).

The application of diffusion policies to mobile robot navigation is a burgeoning area of research. A pioneering work in this domain is NoMaD (Sridhar et al., 2024), which proposes a unified policy for both goal-directed navigation and goal-agnostic exploration using a goal-masking mechanism within a transformer architecture. In DARE (Cao et al., 2025), a diffusion policy for exploration was trained to reason about partially observed environments—i.e., to act based on incomplete belief states— using expert demonstrations generated with access to the full ground-truth map. Other works have applied diffusion models not for direct policy learning but for auxiliary tasks in visual navigation, such as generating goal proposals (Shah et al., 2023) or trajectories (Zeng et al., 2025). Our work, MATT-Diff, contributes to this line of research by formulating and training a diffusion policy for active multi-target tracking, where the policy must learn to balance exploration for new/lost targets with exploitation of currently tracked targets in a multi-modal fashion.

3. PROBLEM STATEMENT

Consider a mobile agent with state $\mathbf{x}_t \in \mathbb{R}^{n_x}$ and control input $\mathbf{u}_t \in \mathbb{R}^{n_u}$, evolving in discrete time according to $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t)$. The agent operates in an environment containing N_y targets with individual states $\mathbf{y}_t^{(j)} \in \mathbb{R}^{n_y}$ for $j \in \{1, \dots, N_y\}$. Both the exact number of targets N_y and their dynamics are unknown to the agent. The agent is equipped with an onboard sensor with a limited FoV denoted by $\mathcal{F}(\mathbf{x}_t) \subset \mathbb{R}^3$ which is dependent on the agent state due to occlusion. The sensor provides measurements $\mathbf{z}_t^{(j)}$ of the j-th target if and only if its position $\mathbf{p}(\mathbf{y}_t^{(j)}) \in \mathbb{R}^3$ lies within the agent's FoV at time t. The sensor model is given by

$$\mathbf{z}_t^{(j)} = H\mathbf{y}_t^{(j)} + \boldsymbol{\eta}_t, \quad \text{if} \quad \mathbf{p}(\mathbf{y}_t^{(j)}) \in \mathcal{F}(\mathbf{x}_t), \tag{1}$$

where $\boldsymbol{\eta}_t \sim \mathcal{N}(0, R)$ is Gaussian measurement noise with covariance $R \in \mathbb{R}^{n_z \times n_z}$.

To estimate the target states, the agent employs a Kalman filter with the limited FoV constraint. The state of each hypothesized target j is estimated as a Gaussian $\mathbf{y}_t^{(j)} | \mathbf{z}_{0:t} \sim \mathcal{N}(\boldsymbol{\mu}_t^{(j)}, \Sigma_t^{(j)})$. The filter update step is executed only when the corresponding target is successfully detected and measured within the agent's FoV. Regardless of the target detection, the filter continues with its prediction step, where the matrices (A, W) describing the target dynamics $\mathbf{y}_{t+1}^{(j)} = A\mathbf{y}_t^{(j)} + \mathbf{w}_t^{(j)}$ with $\mathbf{w}_t^{(j)} \sim \mathcal{N}(\mathbf{0}, W)$ are hypothesized. Due to the process noise $\mathbf{w}_t^{(j)}$, if a target is not detected, its covariance grows to $\Sigma_{t+1}^{(j)} = A\Sigma_t^{(j)}A^\top + W$ and its determinant reflects the increased uncertainty.

A critical challenge arises in the scenario of missing targets. This occurs when the agent predicts a target to be within its FoV, i.e., $\mathbf{p}(\boldsymbol{\mu}_t^{(j)}) \in \mathcal{F}(\mathbf{x}_t)$, but the true target state $\mathbf{y}_t^{(j)}$ is not actually

within the FoV, i.e., $\mathbf{p}(\mathbf{y}_t^{(j)}) \notin \mathcal{F}(\mathbf{x}_t)$. In this case, the agent temporarily ignores the target, while its Kalman filter continues prediction updates and resumes measurement updates upon re-detection without reinitialization. Overall, the set of detected targets $\mathcal{I}_t^{(D)}$ up to time t is given by $\mathcal{I}_t^{(D)} = \mathcal{I}_{t-1}^{(D)} \cup \mathcal{I}_t^{+} \setminus \mathcal{I}_t^{-}$, where $\mathcal{I}_t^{+} := \{j \in \{1, \dots, N_y\} | \mathbf{p}(\mathbf{y}_t^{(j)}) \in \mathcal{F}(\mathbf{x}_t)\}$ denotes the set of newly discovered targets at time t, and $\mathcal{I}_t^{-} := \left\{ j \in \mathcal{I}_t^{(D)} \setminus \mathcal{I}_t^{+} \;\middle|\; \mathbf{p}(\mu_t^{(j)}) \in \mathcal{F}(\mathbf{x}_t) \right\}$ denotes the set of lost targets that were previously detected but are now missing from the FoV at time t.

The agent faces an *exploration-exploitation dilemma*: should it continue to "exploit" by moving towards detected targets to minimize the uncertainty in their states or should it "explore" the environment to search for lost targets or discover new previously unobserved targets? This paper addresses the problem of learning to autonomously navigate and make these multimodal decisions to achieve effective target tracking.

4. METHODOLOGY

Our methodology is centered around learning a multimodal diffusion policy from demonstrations collected by expert target-tracking methods. We first describe the design of the expert algorithms, which generate rich demonstration data. We, then, detail the architecture of our control policy and the diffusion-based training process.

4.1. Expert Planners for Multimodal Demonstrations

We describe three planners that can be used to generate a diverse dataset of target-tracking behaviors, which demonstrate the trade-off between exploration and exploitation effectively.

Frontier-Based Exploration Planner For a pure exploration planner, we utilize a well-known frontier-based occupancy map exploration approach (Yamauchi, 1997). We represent the environment as a discrete occupancy grid $\mathcal{E} \subset \mathbb{Z}^2$ with probabilistic occupancy values $\mathbf{M}_{\text{prob}}(i,j) \in [0,1]$, where a cell (i,j) is classified as free if $\mathbf{M}_{\text{prob}}(i,j) < 0.5$, unknown if $\mathbf{M}_{\text{prob}}(i,j) = 0.5$, and occupied if $\mathbf{M}_{\text{prob}}(i,j) > 0.5$. The obstacle and free regions are defined as $\mathcal{O} = \{\mathbf{p} \in \mathcal{E} \mid \mathbf{M}_{\text{prob}}(\mathbf{p}) > 0.5\}$ and $\mathcal{E}_{\text{free}} = \mathcal{E} \setminus \mathcal{O}$, respectively. Given our assumption that the ground-truth map is known, the free space $\mathcal{E}_{\text{free}}$ is divided into "explored" $\mathcal{E}_{\text{explore},t}$ and "unexplored" $\mathcal{E}_{\text{free}} \setminus \mathcal{E}_{\text{explore},t}$ regions. The frontier region is defined as $\mathcal{E}_{\text{frontier},t} = \partial\mathcal{E}_{\text{explore},t} \setminus \mathcal{O}$, representing the interface between known free and unknown space. The frontier exploration planner selects the next frontier by score minimization $\mathbf{p}_t^* \in \arg\min_{\mathbf{p} \in \mathcal{E}_{\text{frontier},t}} S_t(\mathbf{p})$, where $S_t(\mathbf{p})$ is a weighted combination of distance, visitation frequency, and expected coverage gain. A collision-free path to $\mathbf{p}_t^*$ is planned with RRT* (Karaman and Frazzoli, 2011) within a safety margin from obstacles and is tracked using curvature-based lookahead control. At each step, frontier points are scored based on a weighted function of distance, visitation frequency, and expected coverage gain, encouraging persistent exploration before the entire map is covered. Once the environment becomes fully explored, low-penalty revisits are favored instead of stalling, allowing the agent to reobserve previously lost or moving targets.

Uncertainty-Based Hybrid Planner Once at least one target is detected, a planner must decide whether to continue exploring for new targets or to track existing ones. We design a hybrid exploration-exploitation planner that makes this decision based on the targets' uncertainty. Namely, when all currently detected target states are known with high confidence (i.e., low uncertainty), the

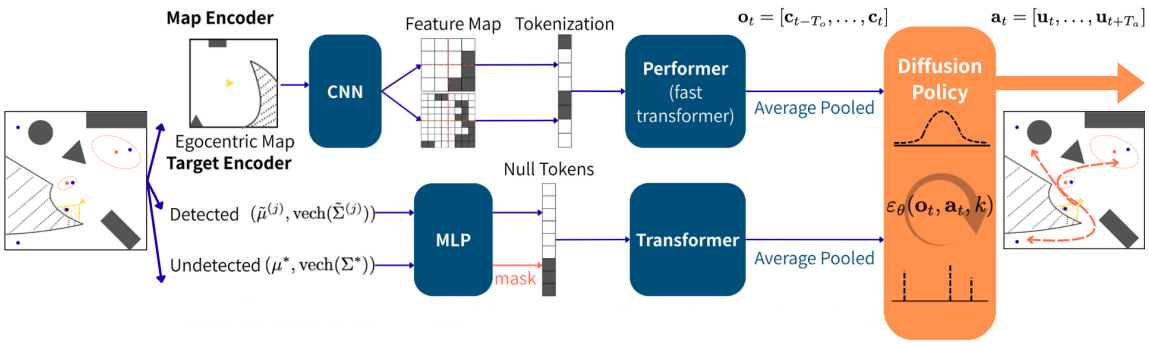

Figure 1: Our MATT-Diff architecture consists of a map encoder and a target encoder. The map encoder converts a local egocentric map into patch tokens via CNN and feeds them into a Performer transformer. The target encoder processes detected target beliefs with masking for undetected targets through self-attention to produce context-aware embeddings. A diffusion policy performs a denoising process to generate multimodal action sequences.

planner prioritizes exploration to discover new targets. Conversely, if any target's uncertainty exceeds a predefined threshold, the planner switches to tracking mode, prioritizing the reduction of uncertainty for the most unconfidently localized target. The target uncertainty is measured by its differential entropy, which is proportional to the log-determinant of its covariance matrix, $\log\det(\Sigma_t^{(j)})$ (Le Ny and Pappas, 2009). In the tracking mode, the planner uses RRT* to generate a path toward the mean of the most uncertain target $\boldsymbol{\mu}_t^{(j^*)}$, where $j^* = \arg\max_{j \in \mathcal{I}_t^{(D)}} \log\det(\Sigma_t^{(j)})$.

Time-Based Hybrid Planner As a third expert, we use a time-based hybrid exploration-exploitation planner. The planner first starts with frontier-based exploration and, once a target is detected, to estimate its state as accurately as possible, the planner keeps tracking the detected target for a fixed-time interval. Tracking is done via RRT* with goal iteratively set to mean of the tracked target. After the fixed time has passed, the planner switches back to frontier-based exploration to search for undiscovered targets.

4.2. Policy Network Architecture

The MATT-Diff policy network is designed to process heterogeneous sensor and target-density inputs to generate a sequence of future agent actions. The agent pose is used only for coordinate transformations within the encoders to obtain egocentric representations. The architecture consists of two encoders, a map encoder and a target encoder, whose outputs are fused and used to condition a U-Net (Ronneberger et al., 2015) in the diffusion model to produce multi-modal action sequences through iterative denoising. An overview of the architecture is shown in Fig. 1.

Map Encoder An egocentric occupancy-grid map, representing a subset of the global map of the environment in the agent's local coordinate frame, is generated from the ground-truth occupancy-grid map and the current agent pose. Specifically, the global map is transformed by the agent pose $\mathbf{x}_t = [p_t, \theta_t]$ through rotation and translation to obtain an egocentric view centered at the agent's position p_t and aligned with its orientation θ_t. This local map is then processed by a Convolutional

Neural Network (CNN) to extract multi-resolution features.The features are divided into a sequence of non-overlapping patches, i.e., map tokens, which are linearly embedded and passed to a transformer encoder to obtain a latent representation of the local environment geometry (Xiao et al., 2023). We use a Performer (Choromanski et al., 2021), a fast transformer model with linear time and space complexity, to efficiently handle the high dimensionality of the egocentric map.

Target Encoder To handle a variable number of targets, we use an attention module. For a detected target's Gaussian density $(\boldsymbol{\mu}, \Sigma)$, the encoder uses an input feature $(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma}) = (\mathbf{q}(\mathbf{x}, \boldsymbol{\mu}), \frac{\Sigma}{\log\det(\overline{\Sigma})})$, where $\mathbf{q}(\mathbf{x}, \mathbf{p})$ transforms the position $\mathbf{p}$ into the agent's coordinate frame given the agent pose $\mathbf{x}$, and $\overline{\Sigma}$ is a threshold covariance matrix defined later. Conversely, for undetected targets, the encoder input is set as $(\boldsymbol{\mu}^*, \Sigma^*) = (\mathbf{q}(\mathbf{x}, \mathbf{0}), \frac{\overline{\Sigma}}{\log\det(\overline{\Sigma})})$. This representation centers the target belief at the environment's origin (viewed from the agent's frame) and assigns a sufficiently large covariance $\overline{\Sigma}$ so that the target belief is uninformative by covering nearly the entire environment. The state vectors for both detected and undetected targets are passed through a Multi-Layer Perceptron (MLP) to create target embeddings, which are then processed by a self-attention layer (transformer encoder). During this step, undetected targets are masked based on the condition $\log\det(\Sigma^*) \geq 1$. Finally, mean pooling is applied to these representations to generate a single fixed-size context vector that summarizes the entire multi-target set.

4.3. Diffusion Policy Training

Our policy is trained as a conditional Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020) to generate a T_a-step future action sequence $\mathbf{a}_t = [\mathbf{u}_t, \ldots, \mathbf{u}_{t+T_a-1}]$ given a T_o-step past observation context sequence $\mathbf{o}_t = [\mathbf{c}_{t-T_o+1}, \ldots, \mathbf{c}_t]$, where each $\mathbf{c}_t$ denotes the fused map-target embedding obtained from the encoders described in Section 4.2. The network ε_θ is trained to predict the added Gaussian noise at each diffusion timestep k. Here, ε^k denotes the sampled noise at step k, and $\mathbf{a}_t^k = \mathbf{a}_t + \varepsilon^k$ represents the noisy version of the clean expert action sequence $\mathbf{a}_t$. The training objective minimizes the mean squared error between the true and predicted noise:

$$\mathcal{L}(\theta) = \|\varepsilon^k - \varepsilon_\theta(\mathbf{o}_t, \mathbf{a}_t^k, k)\|^2. \tag{2}$$

At inference time, an action sequence is generated by starting with pure noise $\mathbf{a}_K \sim \mathcal{N}(0, \mathbf{I})$ and iteratively applying the learned denoising function for $k = K, \ldots, 1$ by

$$\mathbf{a}_t^{k-1} = \alpha(\mathbf{a}_t^k - \gamma\varepsilon_\theta(\mathbf{o}_t, \mathbf{a}_t^k, k) + \mathcal{N}(0, \sigma^2 I)), \tag{3}$$

where $\alpha, \gamma, \sigma > 0$ are noise scheduling functions of k. This iteration produces a clean action plan $\mathbf{a}_t^0$ which is implemented as an output of the diffusion policy.

5. EVALUATION

We train MATT-Diff and compare its performance against a behavior cloning (BC) baseline (Torabi et al., 2018), a deep reinforcement learning (RL) baseline adapted from Jeong et al. (2021), and an ablation model of MATT-Diff eliminating the map encoder to justify the contribution of the egocentric location awareness provided in our architecture. The BC baseline learns a deterministic policy π_ϕ that predicts actions $\mathbf{u}_t$ from observations $\mathbf{o}_t$ (encoded map and target beliefs). The parameters ϕ are optimized to minimize the mean squared error $\mathcal{L}_{\text{BC}} = \left\|\mathbf{u}_t^{\text{exp}} - \pi_\phi(\mathbf{o}_t)\right\|^2$, where

$\mathbf{u}_t^{\text{exp}}$ denotes expert actions. The RL basline uses the Deep Q-Network (DQN) for target tracking proposed by Jeong et al. (2021). The agent is trained by setting the reward to maximize as $R_t = -\lambda \sum_{j \in \mathcal{I}_t^{(D)}} \log(\det(\Sigma_t^{(j)})$, namely, minimizing the differential entropy, where $\lambda > 0$ is a scaling factor to stabilize the DQN learning process (set as $\lambda = 0.1$ in our experiments). MATT-Diff (w/o map encoder) is implemented based solely on the target encoder to generate actions, serving to validate the need for map tokenization. The training, simulation, and evaluation are done on a computer with Ubuntu 24.04, Intel Core Ultra 7, Nvidia GeForce RTX 5080.

5.1. Experiment Setup

The experiment setup for both training and evaluation is described below.

Agent dynamics The agent's dynamics are set to a two-dimensional single-integrator model, i.e., the agent's position $\mathbf{p}_t \in \mathbb{R}^2$ follows $\mathbf{p}_{t+1} = \mathbf{p}_t + \mathbf{u}_t$ with control input $\mathbf{u}_t \in \mathbb{R}^2$, in both training and evaluation scenarios. To represent FoV, the heading angle is set to $\theta_{t+1} = \arctan(u_t^{(2)}/u_t^{(1)})$ where $u_t^{(i)} \in \mathbb{R}$ for $i \in \{1, 2\}$ is i-th element in $\mathbf{u}_t$.

Target configulations and estimation The target dynamics are set to follow a Brownian velocity model in 2-D, i.e., with target dimension $n_y = 2$, transition matrix $A = I$, and process noise covariance $W = \text{diag}(w_x^2, w_y^2)$, where (w_x, w_y) are sampled uniformly from $[0.8, 1.2]$ at the beginning of each episode. The sensor model is set as $H = I$ and the measurement noise covariance is set as $R = \text{diag}(r_x^2, r_y^2)$ with $r_x = r_y = 0.05$. We consider the case where the ground-truth noise covariances in targets' process noise and the measurement noise are unknown to the agent. To represent a conservative state estimation in the agent's internal Kalman filter, the estimated process noise covariance $\hat{W}$ is set to $\hat{W} = \text{diag}(90, 40)$ and the estimated measurement noise covariance $\hat{R}$ is set to a significantly higher magnitude than the ground-truth. For each configuration, we run 20 randomized episodes across the unseen map and report the mean and the standard deviation of the performance metrics. Both the training and evaluation scenarios follow this setup across randomized number of moving targets among $N_y \in \{3, \dots, 6\}$ and randomized initial configurations for both agent's and targets' states.

Environment maps The environment maps are sourced from the HouseExpo dataset (Li et al., 2020). The expert data generation by the three planners introduced in Section 4.1 and the training of MATT-Diff, MATT-Diff (w/o map encoder), BC, and DQN are done over four different maps, while the evaluation is done on a map not included in the training maps to evaluate the performance in an out-of-distribution (OOD) environment.

5.2. Episode-Averaged Results

The performance of MATT-Diff, MATT-Diff(w/o map encoder), DQN, and BC is measured using three metrics: root mean squared error (RMSE), negative log-likelihood (NLL) (Pinto et al., 2021), and differential entropy of the target estimates. For undetected targets, as handled by the target encoder introduced in Section 4.2, the Gaussian densities are set to $\mathcal{N}(\mathbf{0}, \overline{\Sigma})$ with sufficiently large $\overline{\Sigma}$, to approximate a uniform distribution over the environment. For detected targets with densities $\mathcal{N}(\boldsymbol{\mu}^{(j)}, \Sigma^{(j)})$ for $j \in \mathcal{I}^{(D)}$ and ground-truth target states $\{\mathbf{y}^{(j)}\}$ for $j \in \mathcal{I} = \{1, \dots, N_y\}$, the evaluation metrics are given by RMSE $= \sum_{j \in \mathcal{I}^{(D)}} ||\mathbf{y}^{(j)} - \boldsymbol{\mu}^{(j)}||/|I^{(D)}|$,

Table 1: The average and standard deviation of the performance metrics over 20 randomized episodes in an out-of-distribution (OOD) map. The lowset average values (i.e., the best results) are highlighted in bold. MATT-Diff outperforms all baselines across all metrics.

Method	RMSE	NLL	Entropy
MATT-Diff	**268.840 ± 155.540**	**13.124 ± 1.468**	**13.332 ± 1.017**
MATT-Diff (w/o map encoder)	273.525 ± 148.701	13.295 ± 1.321	13.501 ± 0.911
BC	299.776 ± 133.638	13.464 ± 1.351	13.582 ± 1.014
DQN	297.049 ± 134.074	13.335 ± 1.878	13.465 ± 1.589

Entropy $= \sum_{j \in \mathcal{I}^{(D)}} \log\left(\det\left(\Sigma^{(j)}\right)\right) + (N_y - |\mathcal{I}^{(D)}|) \log\left(\det\left(\overline{\Sigma}\right)\right)$, and

$$\text{NLL} = -\sum_{j \in \mathcal{I}^{(D)}} \log\left(p_{\mathcal{N}}\left(\mathbf{y}^{(j)} | \boldsymbol{\mu}^{(j)}, \Sigma^{(j)}\right)\right) - \sum_{j \in \mathcal{I} \setminus \mathcal{I}^{(D)}} \log\left(p_{\mathcal{N}}(\mathbf{y}^{(j)} | \mathbf{0}, \overline{\Sigma})\right),$$

where $p_{\mathcal{N}}(\mathbf{x}|\boldsymbol{\mu}, \Sigma)$ is the probability density function of a Gaussian with mean $\boldsymbol{\mu}$ and covariance Σ. For all three metrics, lower values are desirable for successful target tracking.

We evaluate the trained MATT-Diff, MATT-Diff (w/o map encoder), DQN, and BC by analyzing the average and standard deviation of the metrics over 20 randomized episodes in an OOD map, summarized in Table 1. As shown in Table 1, MATT-Diff achieves the lowest averaged value in all metrics among all the implemented learning-based approaches, exhibiting the strongest zero-shot generalization to an OOD environment. This observation justifies the MATT-Diff's capability to effectively fuse spatial awareness and target densities to balance exploration and tracking in a multimodal fashion.

5.3. Per-Episode Results

We analyze the temporal behavior and performance metrics for MATT-Diff and the expert planners. Fig. 2 and Fig. 3 show the temporal evolution of NLL and the corresponding trajectory visualizations for MATT-Diff, the frontier-based expert, and the time-based expert within a single episode. Although this specific episode does not represent the peak performance, it was selected for its clear illustration of the distinct behavioral phases inherent in our MATT-Diff policy. Specifically, the NLL curve of MATT-Diff (green) in Fig. 2 exhibits six characteristic phases: (1) an initial *exploration* phase, where after an early detection, the NLL rises as the agent prioritizes exploring the frontier space rather than immediate tracking, showing a trend similar to the frontier-based planner (timesteps 0–75); (2) a *tracking* phase, where the NLL drops sharply upon target traking and remains low (timesteps 75–100); (3) a renewed *exploration* phase, marked by a steady, monotonic increase in NLL as the agent temporarily search other regions (timesteps 100–375); (4) a combined *tracking and re-acquisition* phase, where the NLL fluctuates rapidly as the agent iteratively loses and rediscovers the target (timesteps 375–500); (5) a transition back to *exploration*, indicated by a gradual increase in NLL as the agent diverges to search frontier spaces (timesteps 500–750); (6) a final *reacqisition* phase, where the NLL fluctuates repeatedly as the agent successfully re-acquires the target through active searching (timesteps 750–1000). In contrast, the frontier-based expert (blue) continually expands its exploration frontier, demonstrating no explicit tracking or re-acquisition behavior, as evidenced by the brief drops in its curve. Meanwhile, the time-based expert (red) focuses

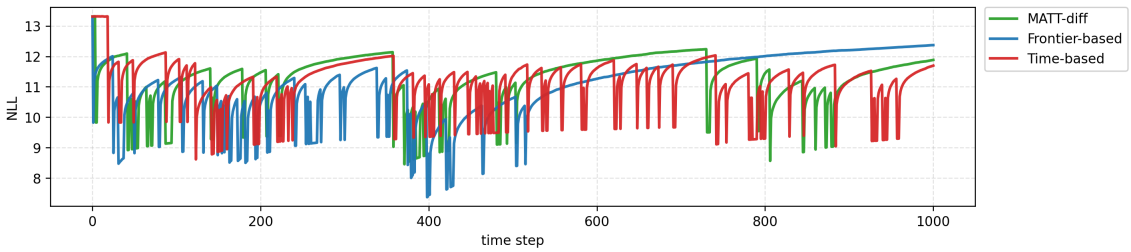

Figure 2: Temporal evolution of NLL for the frontier-based (blue) expert, the time-based (red) expert, and MATT-diff (green) in one episode.

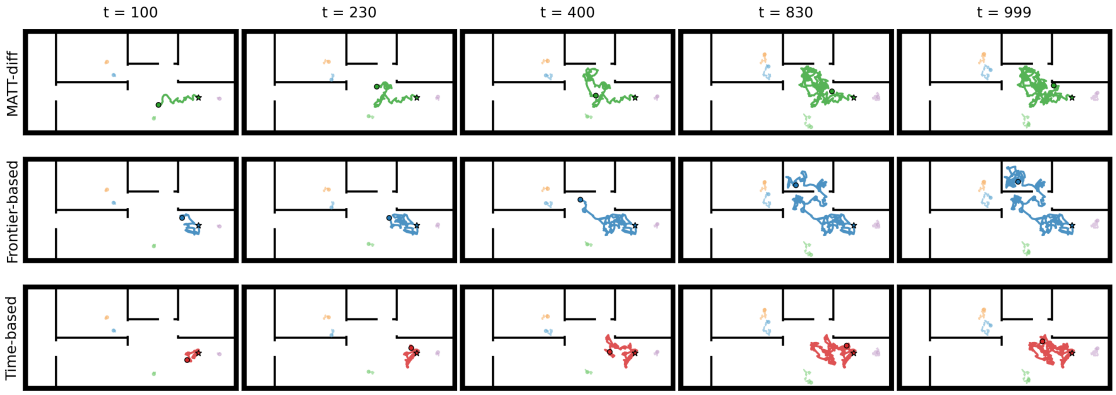

Figure 3: Trajectory snapshots of MATT-Diff and the expert planners over a single episode. MATT-Diff achieves a good balance of exploration, target tracking, and re-acquisition.

heavily on a fixed tracking duration after detection, resulting in the highest minimum NLL among the three planners, while still maintaining a certain low level.

Fig. 3 qualitatively compares the behaviors of the proposed and expert planners in the same episode as Fig. 2. The frontier-based expert tends to focus on broad coverage, sweeping unexplored regions but not following already detected targets. In contrast, MATT-Diff maintains target visibility through local tracking, occasionally performing short re-acquisition loops and then expanding its path toward unseen areas. This results in a balanced behavior between exploration and tracking that corresponds to the temporal profile discussed above.

We also observe certain failure modes depending on local geometric constraints. As illustrated in Fig. 4, MATT-Diff occasionally fails to maintain safety, resulting in collisions that trigger early termination. This effectively limits the agent's exposure to tracking and re-acquisition opportunities, whereas the expert planners consistently demonstrate robust collision avoidance.

5.4. Limitations and Future Directions

While MATT-Diff exhibits effective multi-modal behaviors for target tracking, several limitations remain.

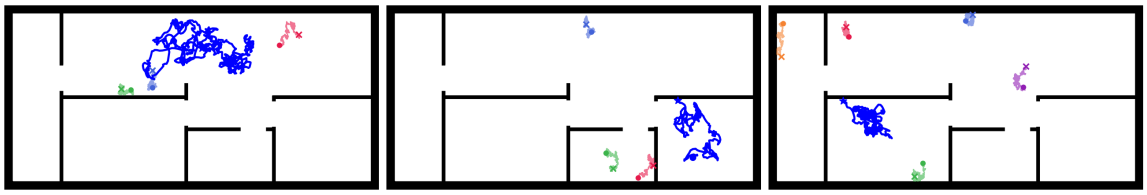

Figure 4: Trajectories followed by MATT-Diff across multiple episodes. The policy occasionally exhibits insufficient collision avoidance, causing the agent to get stuck in obstacle regions.

Safety guarantee Although the policy learns collision avoidance from expert demonstrations implicitly, it lacks explicit safety guarantees as observed in Fig. 4. While recent works have explored enforcing hard constraints on diffusion model outputs during inference (Römer et al., 2025; Cheng et al., 2025), integrating these mechanisms into the finetuning phase remains a challenge. An interesting question for future research is how to explicitly embed safeguards into the training loop such as policy learning through a constrained inference layer to ensure provably safe action generation.

Tuning the policy specifically for one map Because diffusion policies inherently represent multimodal action distributions, the specific mode executed at each timestep depends on stochastic sampling from the diffusion process. Fine-tuning the policy with reinforcement learning (Ren et al., 2025; Wagenmaker et al., 2025) may help the agent learn exactly when to select each behavior mode, improving consistency and task performance.

Training with other target estimators We remark that our policy is trained based on target estimation provided by a Kalman filter. As discussed in Section 3, this currently relies on a heuristic approach of abruptly removing the estimates of lost targets upon missing observations. More advanced target estimation methods that explicitly incorporate detection probabilities and missing observation models, such as the Probabilistic Hypothesis Density (PHD) filter (Vo and Ma, 2006), can be used in the design. A key challenge for future work is to determine the network inputs and architecture capable of processing detection probabilities and data association probabilities provided by advanced probabilistic inference techniques.

6. CONCLUSION

This paper proposed a novel network design and training methodology for active multimodal target tracking via a diffusion policy, named MATT-Diff. The exploration-exploitation dilemma faced in tracking targets with unknown number, states, and dynamics was addressed by MATT-Diff using demonstrations from three planners with distinct exploration and exploitation behaviors. The network architecture of MATT-Diff applies CNN and vision-performer to the egocentric map and an attention mechanism to handle varying number of targets. The diffusion policy was trained to predict noise added to the expert action sequences given the observation sequences, and performed a denoising process starting from random noise to generate a multimodal action sequence. Our evaluation showed that MATT-Diff outperforms other learning-based approaches in an out-of-distribution environment and exhibits multimodal behavior derived from the expert planners.

References

Yuhong Cao, Jeric Lew, Jingsong Liang, Jin Cheng, and Guillaume Sartoretti. Dare: Diffusion policy for autonomous robot exploration. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11987–11993. IEEE, 2025.

Xiaoyuan Cheng, Xiaohang Tang, and Yiming Yang. Safe and stable control via lyapunov-guided diffusion models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=aY97JGello.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, page 02783649241273668, 2023.

Krzysztof Marcin Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Quincy Davis, Afroz Mohiuddin, Lukasz Kaiser, David Benjamin Belanger, Lucy J Colwell, and Adrian Weller. Rethinking attention with performers. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=Ua6zuk0WRH.

Shutong Ding, Ke Hu, Zhenhao Zhang, Kan Ren, Weinan Zhang, Jingyi Yu, Jingya Wang, and Ye Shi. Diffusion-based reinforcement learning via q-weighted variational policy optimization. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=UWUUVKtKeu.

Alfred O. Hero and Douglas Cochran. Sensor management: Past, present, and future. *IEEE Sensors Journal*, 11(12):3064–3075, 2011. doi: 10.1109/JSEN.2011.2167964.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Heejin Jeong, Hamed Hassani, Manfred Morari, Daniel D Lee, and George J Pappas. Deep reinforcement learning for active target tracking. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1825–1831. IEEE, 2021.

Yiannis Kantaros, Brent Schlotfeldt, Nikolay Atanasov, and George J Pappas. Asymptotically optimal planning for non-myopic multi-robot information gathering. In *Robotics: Science and Systems*, pages 22–26, 2019.

Sertac Karaman and Emilio Frazzoli. Sampling-based algorithms for optimal motion planning. *The international journal of robotics research*, 30(7):846–894, 2011.

Steven M LaValle, Hector H González-Banos, Craig Becker, and J-C Latombe. Motion strategies for maintaining visibility of a moving target. In *Proceedings of international conference on robotics and automation*, volume 1, pages 731–736. IEEE, 1997.

Jerome Le Ny and George J Pappas. On trajectory optimization for active sensing in gaussian process models. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 6286–6292. IEEE, 2009.

Steven Li, Rickmer Krohn, Tao Chen, Anurag Ajay, Pulkit Agrawal, and Georgia Chalvatzaki. Learning multimodal behaviors from scratch with diffusion policy gradient. *Advances in Neural Information Processing Systems*, 37:38456–38479, 2024.

Tingguang Li, Danny Ho, Chenming Li, Delong Zhu, Chaoqun Wang, and Max Q-H Meng. Houseexpo: A large-scale 2d indoor layout dataset for learning-based algorithms on mobile robots. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5839–5846. IEEE, 2020.

Ajith Anil Meera, Marija Popović, Alexander Millane, and Roland Siegwart. Obstacle-aware adaptive informative path planning for uav-based target search. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 718–724. IEEE, 2019.

Juliano Pinto, Yuxuan Xia, Lennart Svensson, and Henk Wymeersch. An uncertainty-aware performance measure for multi-object tracking. *IEEE Signal Processing Letters*, 28:1689–1693, 2021.

Jorge Pena Queralta, Jussi Taipalmaa, Bilge Can Pullinen, Victor Kathan Sarker, Tuan Nguyen Gia, Hannu Tenhunen, Moncef Gabbouj, Jenni Raitoharju, and Tomi Westerlund. Collaborative multirobot search and rescue: Planning, coordination, perception, and active vision. *IEEE Access*, 8: 191617–191643, 2020.

Allen Z. Ren, Justin Lidard, Lars Lien Ankile, Anthony Simeonov, Pulkit Agrawal, Anirudha Majumdar, Benjamin Burchfiel, Hongkai Dai, and Max Simchowitz. Diffusion policy policy optimization. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=mEpqHvbD2h.

Ralf Römer, Alexander von Rohr, and Angela Schoellig. Diffusion predictive control with constraints. In *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, pages 791–803. PMLR, 2025.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Brent Schlotfeldt, Dinesh Thakur, Nikolay Atanasov, Vijay Kumar, and George J Pappas. Anytime planning for decentralized multirobot active information gathering. *IEEE Robotics and Automation Letters*, 3(2):1025–1032, 2018.

Dhruv Shah, Ajay Sridhar, Nitish Dashora, Kyle Stachowicz, Kevin Black, Noriaki Hirose, and Sergey Levine. ViNT: A foundation model for visual navigation. In *7th Annual Conference on Robot Learning*, 2023. URL https://arxiv.org/abs/2306.14846.

Mingchen Song, Xiang Deng, Zhiling Zhou, Jie Wei, Weili Guan, and Liqiang Nie. A survey on diffusion policy for robotic manipulation: Taxonomy, analysis, and future directions. *Authorea Preprints*, 2025.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

John R Spletzer and Camillo J Taylor. Dynamic sensor planning and control for optimally tracking targets. *The International Journal of Robotics Research*, 22(1):7–20, 2003.

Ajay Sridhar, Dhruv Shah, Catherine Glossop, and Sergey Levine. Nomad: Goal masked diffusion policies for navigation and exploration. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 63–70. IEEE, 2024.

Sanjeev Ramkumar Sudha, Marija Popović, and Erlend M Coates. An informative planning framework for target tracking and active mapping in dynamic environments with asvs. *arXiv preprint arXiv:2508.14636*, 2025.

Nianyi Sun, Jin Zhao, Qing Shi, Chang Liu, and Peng Liu. Moving target tracking by unmanned aerial vehicle: A survey and taxonomy. *IEEE Transactions on Industrial Informatics*, 20(5): 7056–7068, 2024.

Faraz Torabi, Garrett Warnell, and Peter Stone. Behavioral cloning from observation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4950–4957. International Joint Conferences on Artificial Intelligence Organization, 7 2018. doi: 10.24963/ijcai.2018/687. URL https://doi.org/10.24963/ijcai.2018/687.

Mariliza Tzes, Nikolaos Bousias, Evangelos Chatzipantazis, and George J. Pappas. Graph neural networks for multi-robot active information acquisition. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3497–3503, 2023. doi: 10.1109/ICRA48891.2023.10160723.

B-N Vo and W-K Ma. The gaussian mixture probability hypothesis density filter. *IEEE Transactions on signal processing*, 54(11):4091–4104, 2006.

Andrew Wagenmaker, Mitsuhiko Nakamoto, Yunchu Zhang, Seohong Park, Waleed Yagoub, Anusha Nagabandi, Abhishek Gupta, and Sergey Levine. Steering your diffusion policy with latent space reinforcement learning. *arXiv preprint arXiv:2506.15799*, 2025.

Dian Wang, Stephen Hart, David Surovik, Tarik Kelestemur, Haojie Huang, Haibo Zhao, Mark Yeatman, Jiuguang Wang, Robin Walters, and Robert Platt. Equivariant diffusion policy. In *8th Annual Conference on Robot Learning*, 2024a. URL https://openreview.net/forum?id=wD2kUVLT1g.

Shuaikang Wang, Yiannis Kantaros, and Meng Guo. Uncertainty-bounded active monitoring of unknown dynamic targets in road-networks with minimum fleet. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4584–4590. IEEE, 2024b.

Zhendong Wang, Jonathan J Hunt, and Mingyuan Zhou. Diffusion policies as an expressive policy class for offline reinforcement learning. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=AHvFDPi-FA.

Xuesu Xiao, Tingnan Zhang, Krzysztof Marcin Choromanski, Tsang-Wei Edward Lee, Anthony Francis, Jake Varley, Stephen Tu, Sumeet Singh, Peng Xu, Fei Xia, Sven Mikael Persson, Dmitry Kalashnikov, Leila Takayama, Roy Frostig, Jie Tan, Carolina Parada, and Vikas Sindhwani. Learning model predictive controllers with real-time attention for real-world navigation. In Karen Liu, Dana Kulic, and Jeff Ichnowski, editors, *Proceedings of The 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pages 1708–1721. PMLR, 14–18 Dec 2023. URL https://proceedings.mlr.press/v205/xiao23a.html.

Brian Yamauchi. A frontier-based approach for autonomous exploration. In *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97.'Towards New Computational Principles for Robotics and Automation'*, pages 146–151. IEEE, 1997.

Ge Yan, Yueh-Hua Wu, and Xiaolong Wang. Dnact: Diffusion guided multi-task 3d policy learning. *arXiv preprint arXiv:2403.04115*, 2024.

Pengzhi Yang, Shumon Koga, Arash Asgharivaskasi, and Nikolay Atanasov. Policy learning for active target tracking over continuous $SE(3)$ trajectories. In *5th Annual Learning for Dynamics & Control Conference*, 2023. URL https://openreview.net/forum?id=6dp5uz72ql.

Maryam Zare, Parham M Kebria, Abbas Khosravi, and Saeid Nahavandi. A survey of imitation learning: Algorithms, recent developments, and challenges. *IEEE Transactions on Cybernetics*, 2024.

Yiming Zeng, Hao Ren, Shuhang Wang, Junlong Huang, and Hui Cheng. Navidiffusor: Cost-guided diffusion model for visual navigation. *arXiv preprint arXiv:2504.10003*, 2025.

Lifeng Zhou, Vasileios Tzoumas, George J Pappas, and Pratap Tokekar. Resilient active target tracking with multiple robots. *IEEE Robotics and Automation Letters*, 4(1):129–136, 2018.

Minnan Zhou, Mustafa Shaikh, Vatsalya Chaubey, Patrick Haggerty, Shumon Koga, Dimitra Panagou, and Nikolay Atanasov. Control strategies for pursuit-evasion under occlusion using visibility and safety barrier functions. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12863–12869. IEEE, 2025.