Online Feedback Efficient Active Target Discovery in Partially Observable Environments

Anindya Sarkar*, **Binglin Ji***, **Yevgeniy Vorobeychik**
{anindya, binglin.j, yvorobeychik}@wustl.edu,
Department of Computer Science and Engineering
Washington University in St. Louis, USA

Abstract

In various scientific and engineering domains, where data acquisition is costly—such as in medical imaging, environmental monitoring, or remote sensing—strategic sampling from unobserved regions, guided by prior observations, is essential to maximize target discovery within a limited sampling budget. In this work, we introduce Diffusion-guided Active Target Discovery (*DiffATD*), a novel method that leverages diffusion dynamics for active target discovery. *DiffATD* maintains a belief distribution over each unobserved state in the environment, using this distribution to dynamically balance exploration-exploitation. Exploration reduces uncertainty by sampling regions with the highest expected entropy, while exploitation targets areas with the highest likelihood of discovering the target, indicated by the belief distribution and an incrementally trained reward model designed to learn the characteristics of the target. *DiffATD* enables efficient target discovery in a partially observable environment within a fixed sampling budget, all without relying on any prior supervised training. Furthermore, *DiffATD* offers interpretability, unlike existing black-box policies that require extensive supervised training. Through extensive experiments and ablation studies across diverse domains, including medical imaging, species discovery and remote sensing, we show that *DiffATD* performs significantly better than baselines and competitively with supervised methods that operate under full environmental observability.

1 Introduction

A key challenge in many scientific discovery applications is the high cost associated with sampling and acquiring feedback from the ground-truth reward function, which often requires extensive resources, time, and cost. For instance, in MRI, doctors aim to minimize radiation exposure by strategically selecting brain regions to identify potential tumors or other conditions. However, obtaining detailed scans for accurate diagnosis is resource-intensive and time-consuming, involving advanced equipment, skilled technicians, and substantial patient time. Furthermore, gathering reliable feedback to guide treatment decisions often requires expert judgment, adding another layer of complexity and cost. Similarly, in a search and rescue mission, personnel must strategically decide where to sample next based on prior observations to locate a target of interest (e.g., a missing person), taking into account limitations such as restricted field of view and high acquisition costs. In this context, obtaining feedback from the ground truth reward function could involve sending a team for on-site verification or having a human analyst review the collected imagery, both of which require significant resources and expert judgment. This challenge extends to various scientific and engineering fields, such as material and drug discovery, where the objective may vary but the core problem—optimizing sampling to identify a target under constraints—remains the same. The challenge is further exacerbated by the

*Equal contribution

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

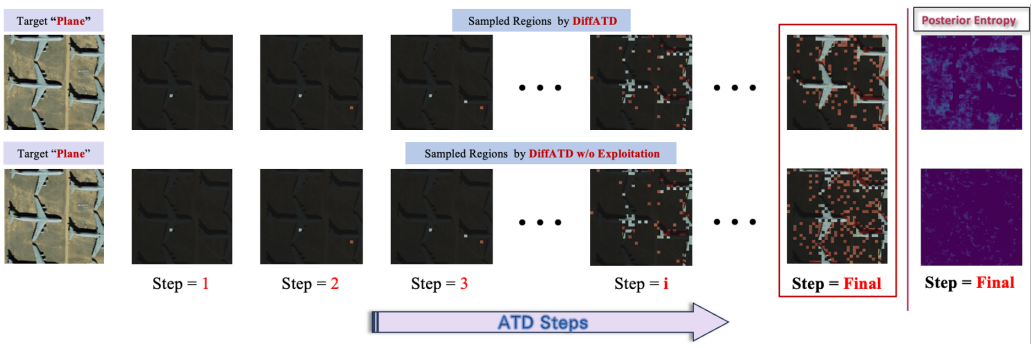

Figure 1: *DiffATD* uncovers more target pixels (e.g., plane) while exhibiting higher uncertainty in the search space. In contrast, *DiffATD without exploitation* explores the environment more effectively, resulting in lower entropy but identifies fewer target pixels.

fact that obtaining labeled data for certain rare target categories, such as rare tumors, is often not feasible. Naturally, the question emerges: ***Is it possible to design an algorithm capable of identifying the target of interest while minimizing reliance on expensive ground truth feedback, and without requiring any task-specific supervised training?***

The challenge is twofold: (i) the agent must judiciously sample the most informative region from a partially observed scene to maximize its understanding of the search space, and (ii) it must concurrently ensure that this selection contributes to the goal of identifying regions likely to include the target. One might wonder why the agent cannot simply learn to choose regions that directly unveil target regions, bypassing the need to acquire knowledge about the underlying environment. The crux of the issue lies in the inherent complexity of reasoning within an unknown, partially observable domain. Consequently, the agent must deftly balance exploration—identifying regions that yield the greatest informational gain about the search space—with exploitation—focusing on areas more likely to contain the target based on the evolving understanding of the environment. Prior approaches typically rely on off-the-shelf reinforcement learning algorithms to develop a search policy that can efficiently explore a search space and identify target regions within a partially observable environment by training on large-scale pre-labeled datasets from similar tasks. However, obtaining such supervised datasets is often impractical in practice, restricting the models' applicability in real-world scenarios. Suppose the task is identifying a rare tumor in an MRI image. In that case, it is unrealistic to expect large-scale supervised data for such a rare condition, with positive labels indicating the target.

To address this challenge, we introduce *DiffATD*, an approach capable of uncovering a target in a partially observable environment without the need for training on pre-labeled datasets, setting it apart from previous methods. Conditioned on the observations gathered so far, *DiffATD* leverages Tweedie's formula to construct a belief distribution over each unobserved region of the search space. Exploration is guided by selecting the region with the highest expected entropy of the corresponding belief distribution, while exploitation directs attention to the region with the highest likelihood score measured by the lowest expected entropy, modulated by the reward value, predicted by an online-trained reward model designed to predict the "target-ness" of a region. Figure 1 illustrates how exploitation enables *DiffATD* to uncover more target pixels (e.g., plane) while exhibiting higher uncertainty about the search space. Moreover, *DiffATD* offers interpretability, providing enhanced transparency compared to existing methods that rely on opaque black-box policies, all while being firmly grounded in the theoretical framework of Bayesian experiment design. We summarize our contributions as follows:

- We propose *DiffATD*, a novel algorithm that uncovers the target of interest within a feedback/query budget in a partially observable environment, minimizing reliance on costly ground truth feedback and avoiding the need for task-specific training with annotated data.
- *DiffATD* features a white-box policy for sample selection based on Bayesian experiment design, offering greater interpretability and transparency compared to existing black-box methods.
- We demonstrate the significance of each component in our proposed *DiffATD* method through extensive quantitative and qualitative ablation studies across a range of datasets, including medical imaging, species distribution modeling, and remote sensing.

2 Related Work

Active Target/Scene Reconstruction. There is extensive prior work on active reconstruction of scenes and/or objects [1, 2, 3]. Recently, deep learning methods have been developed to design subsampling strategies that aim to minimize reconstruction error. Fixed strategies, such as those proposed by [4, 5], involve designing a single sampling mask for a specific domain, which is then applied uniformly to all samples during inference. While effective in certain scenarios, these methods face limitations when the optimal sampling pattern varies across individual samples. To overcome this, sample-adaptive approaches [6, 7, 8, 9] have been introduced, dynamically tailoring sampling strategies to each sample during inference. [6] trained a neural network to iteratively construct an acquisition mask of M k-space lines, adding lines adaptively based on the current reconstruction and prior context. Similarly, [7] used an RL agent for this process. However, these methods rely on black-box policies, making failure cases hard to interpret. Generative approaches, like [10, 11], address this with transparent sampling policies, such as maximum-variance sampling for measurement selection. However, all these prior methods typically focus solely on optimizing for reconstruction, while our ultimate goal is identifying target-rich regions. Success for our task hinges on balancing *exploration* (obtaining useful information about the scene) and *exploitation* (uncovering targets).

Learning Based approaches for Active Target Discovery. Several RL-based approaches, such as [12, 13, 14, 15], have been proposed for active target discovery, but these methods rely on full observability of the search space and large-scale pre-labeled datasets for learning an efficient sub-sampling strategy. Training-free methods inspired by Bayesian decision theory offer a promising approach to active target discovery [16, 17, 18]. However, their reliance on full observability of the search space remains a significant limitation, rendering them ineffective in scenarios with partial observability. More Recently, a series of methods, such as [19, 20, 21], have been proposed to tackle active discovery in partially observable environments. Despite their potential, these approaches face a significant hurdle—the reliance on extensive pre-labeled datasets to fully realize their capabilities. In this work, we present a novel algorithm, ***DiffATD***, designed to overcome the limitations of prior approaches by enabling efficient target discovery in partially observable environments. ***DiffATD*** eliminates the need for pre-labeled training data, significantly enhancing its adaptability and practicality across diverse applications, from medical imaging to remote sensing.

3 Preliminaries

Denoising diffusion models are trained to reverse a Stochastic Differential Equation (SDE) that gradually perturbs samples x into a standard normal distribution [22]. The SDE governing the noising process is given by:

$$dx = -\frac{\beta(\tau)}{2} x \, d\tau + \sqrt{\beta(\tau)} \, dw$$

where $x_0 \in \mathbb{R}^d$ is an initial clean sample, $\tau \in [0, T]$, $\beta(\tau)$ is the noise schedule, and w is a standard Wiener process, with $x(T) \sim \mathcal{N}(0, I)$. According to [23], this SDE can be reversed once the score function $\nabla_x \log p_\tau(x)$ is known, where $\bar{w}$ is a standard Wiener process running backwards:

$$dx = \left[-\frac{\beta(\tau)}{2} x - \beta(\tau) \nabla_x \log p_\tau(x) \right] d\tau + \sqrt{\beta(\tau)} \, d\bar{w}$$

Following [24, 25], the discrete formulation of the SDE is defined as $x_\tau = x(\tau T/N)$, $\beta_\tau = \beta(\tau T/N)$, $\alpha_\tau = 1 - \beta_\tau$, and $\bar{\alpha}_\tau = \prod_{s=1}^{\tau} \alpha_s$, where N represents the number of discretized segments. The diffusion model reverses the SDE by learning the score function using a neural network parameterized by θ, where $s_\theta(x_\tau, \tau) \approx \nabla_{x_\tau} \log p_\tau(x_\tau)$. The reverse diffusion process can be conditioned on observations gathered so far y_t to generate samples from the posterior $p(x \mid y_t)$. This is achieved by substituting $\nabla_{x_\tau} \log p_\tau(x_\tau)$ with the conditional score function $\nabla_{x_\tau} \log p_\tau(x_\tau \mid y_t)$ in the above Equation. The difficulty in handling $\nabla_{x_\tau} \log p_\tau(y_t \mid x_\tau)$, stemming from the application of Bayes' rule to refactor the posterior, has inspired the creation of several approximate guidance techniques [26, 27] to compute these gradients with respect to a partially noised sample x_τ. Most of these methods utilize *Tweedie's formula*, which serves as a one-step denoising operation from $\tau \to 0$, represented as $\mathcal{T}_\tau(\cdot)$. Using a trained diffusion model, the fully denoised sample x_0 can be estimated from its noisy version x_τ as follows:

$$\hat{x}_0 = \mathcal{T}_\tau(x_\tau) = \frac{1}{\sqrt{\bar{\alpha}_\tau}} \left(x_\tau + (1 - \bar{\alpha}_\tau) s_\theta(x_\tau, \tau) \right)$$

4 Problem Formulation

In this section, we present the details of our proposed Active Target Discovery (ATD) task setup. ATD involves actively uncovering one or more targets within a search area, represented as an region x divided into N grid cells, such that $x = (x^{(1)}, x^{(2)}, ..., x^{(N)})$. ATD operates under a query budget $\mathcal{B}$, representing the maximum number of measurements allowed. Each grid cell represents a sub-region and serves as a potential measurement location. A measurement reveals the content of a specific sub-region $x^{(i)}$ for the i-th grid cell, yielding an outcome $y^{(i)} \in [0, 1]$, where $y^{(i)}$ represents the ratio of pixels in the grid cell $x^{(i)}$ that belongs to the target of interest. In each task configuration, the target's content is initially unknown and is revealed incrementally through observations from measurements. The goal is to identify as many grid cells belonging to the target as possible by strategically exploring the grid within the given budget $\mathcal{B}$. Denoting a query performed in step t as q_t, *the overall task optimization objective is*:

$$U(x^{\{(q_t)\}}; \{q_t\}) \equiv \max_{\{q_t\}} \sum_t y^{(q_t)} \text{ subject to } t \leq \mathcal{B} \tag{1}$$

With objective 1 in focus, we aim to develop a search policy that efficiently explores the search area ($x_{\text{test}} \sim X_{\text{test}}$) to identify as many target regions as possible within a measurement budget $\mathcal{B}$, and to achieve this by utilizing a pre-trained diffusion model, trained in an unsupervised manner on samples x_{train} from the training set, X_{train}.

5 Solution Approach

Crafting a sampling strategy that efficiently balances exploration and exploitation is the key to success in solving ATD. Before delving into exploitation, we first explain how ***DiffATD*** tackles exploration of the search space using the measurements gathered so far. To this end, ***DiffATD*** follows a maximum-entropy strategy, by selecting measurement q_t^{exp},

$$q_t^{\text{exp}} = \arg\max_{q_t} [I(\hat{x}_t; x \mid Q_t, \tilde{x}_{t-1})] \tag{2}$$

where I denotes the mutual information between the full search space x and the predicted search space $\hat{x}_t$, conditioned on prior observations ($\tilde{x}_{t-1}$) and the measurement locations $Q_t = \{q_0, q_1, \ldots, q_t\}$ selected up to time t, where q_t is the measurement location at time t. We now present a proposition that facilitates the simplification of Equation 2.

Proposition 1. *Maximizing $I(\hat{x}_t; x \mid Q_t, \tilde{x}_{t-1})$ w.r.t a measurement location q_t is equivalent to maximizing the marginal entropy of the estimated search space $\hat{x}_t$, i.e.,*

$$\arg\max_{q_t} [I(\hat{x}_t; x \mid Q_t, \tilde{x}_{t-1})] = \arg\max_{q_t} [H(\hat{x}_t | Q_t, \tilde{x}_{t-1})]$$

Where H denotes entropy. We derive Prop. 1 in the Appendix. The marginal entropy $H(\hat{x}_t|Q_t, \tilde{x}_{t-1})$ can be expressed as $-\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})]$, representing the expected log-likelihood of the estimated search space $\hat{x}_t$ based on the observations collected so far and the set of measurement locations Q_t. We refer to this distribution as the *belief distribution*, essential for ranking potential measurement locations from an exploration perspective. Next, we discuss our approach to estimating the *belief distribution*. To this end, ***DiffATD*** functions by executing a reverse diffusion process for a batch $\{x_\tau^{(i)}\}$ of size N_B, where $i \in 0, \ldots, N_B$, guided by an incremental set of observations $\{\tilde{x}_t\}_{t=0}^{\mathcal{B}}$, as detailed in Algorithm 1. These observations are obtained through measurements collected at reverse diffusion steps τ, that satisfy $\tau \in M$, where M is a measurement schedule. An element of this batch, $x_\tau^{(i)}$, is referred to as *particle*. These *particles* implicitly form a *belief distribution* over the entire search space x throughout reverse diffusion and are used to estimate uncertainty about x. Concretely, we model the belief distribution simply as mixtures of N_B isotropic Gaussians, with variance $\sigma_x^2\mathbf{I}$, and mean $\hat{x}_t^{(i)}$ computed via Tweedie's operator $\mathcal{T}_\tau(\{x_\tau^{(i)}\})$ as follows:

$$\hat{x}_t^{(i)} = \mathbb{E}[\hat{x}_\tau^{(i)} \mid x_\tau^{(i)}] = \frac{1}{\sqrt{\bar{\alpha}_\tau}} \left(x_\tau^{(i)} + (1 - \bar{\alpha}_\tau) s_\theta(x_\tau^{(i)}, \tau) \right) \tag{3}$$

Algorithm 1 Diffusion Dynamics Guided by Measurements

Require: $T, N_B, s_\theta, \zeta, \{\alpha_\tau\}_{\tau=0}^{T}, \{\tilde{\sigma}_\tau\}_{\tau=0}^{T}, M, Q_0 = \emptyset$

$t = 0$; $\{x_T^{(i)} \sim \mathcal{N}(0, I)\}_{i=0}^{N_B}; z \sim \mathcal{N}(0, I), \tilde{x}_0 = 0$

for $\tau = T$ to 1 **do**

  for $i = 0$ to N_B **do**

    $\hat{x}_\tau^{(i)} \leftarrow \mathcal{T}_\tau(x_\tau^{(i)}) = \frac{1}{\sqrt{\bar{\alpha}_\tau}}\left(x_\tau^{(i)} + (1-\bar{\alpha}_\tau)\hat{s}^{(i)}\right)$; Where $\hat{s}^{(i)} \leftarrow s_\theta(x_\tau^{(i)}, \tau)$.

    $x_{\tau-1}^{(i)'} \leftarrow \frac{\sqrt{\alpha_\tau}(1-\bar{\alpha}_{\tau-1})}{1-\bar{\alpha}_\tau} x_\tau^{(i)} + \frac{\sqrt{\bar{\alpha}_{\tau-1}}\beta_\tau}{1-\bar{\alpha}_\tau}\hat{x}_\tau^{(i)} + \tilde{\sigma}_\tau z$

    $x_{\tau-1}^{(i)} \leftarrow x_{\tau-1}^{(i)'} - \zeta \underbrace{\nabla_{x_\tau^{(i)}} ||[x]_{Q_t} - [\hat{x}_\tau^{(i)}]_{Q_t}||^2}_{\textit{Measurement Guidance}}$; $[.]_{Q_t}$ select elements of $[.]$ indexed by the set Q_t.

  end for

  if $\tau \in M$ **then**

    $t \leftarrow t + 1$

    Acquire new measurement at location q_t and update: $Q_t \leftarrow Q_{t-1} \cup q_t$, $\tilde{x}_t \leftarrow \tilde{x}_{t-1} \cup [x]_{q_t}$

  end if

end for

We assume that the reverse diffusion step τ corresponds to the measurement step t. Utilizing the expression 3, we can express the *belief distribution* as follows:

$$p(\hat{x}_t|Q_t, \tilde{x}_{t-1}) = \sum_{i=0}^{N_B} \alpha_i \mathcal{N}(\hat{x}_t^i, \sigma_x^2 I) \tag{4}$$

According to [28], the marginal entropy of this *belief distribution* can be estimated as follows:

$$H(\hat{x}_t|Q_t, \tilde{x}_{t-1}) \propto \sum_{i=0}^{N_B} \alpha_i \log \sum_{j=0}^{N_B} \alpha_j \exp\left\{\frac{||\hat{x}_t^{(i)} - \hat{x}_t^{(j)}||_2^2}{2\sigma_x^2}\right\} \tag{5}$$

As per Eqn. 5, uncovering the optimal measurement location (q_t^{exp}) requires computing a separate set of $\hat{x}_t$ for each possible choice of q_t. Next, we present a theorem that empowers us to select the next measurement as the region with the highest total variance across $\{\hat{x}_t^{(i)}\}$ estimated from each particle in the batch conditioned on $(Q_t, \tilde{x}_{t-1})$, effectively bypassing the need to compute a separate set of predicted search space $\hat{x}_t$ for each potential measurement location q_t.

Theorem 1. *Assuming k represents the set of possible measurement locations at step t, and that all particles have equal weights ($\alpha_i = \alpha_j, \forall i, j$), then q_t^{exp} is given by:*

$$\arg\max_{q_t} \left[\sum_{i=0}^{N_B} \log \sum_{j=0}^{N_B} \exp\left(\frac{\sum_{q_t \in k}([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2}\right)\right],$$

where $[.]_{q_t}$ selects element of $[.]$ indexed by q_t.

We prove this result in Appendix. Its consequence is that the optimal choice for the next measurement location (q_t^{exp}) lies in the region of the measurement space where the predicted search space ($\hat{x}_t^{(i)}$) exhibits the greatest disagreement across the *particles* ($i \in 0, \ldots, N_B$), quantified by a metric, denoted as $\text{expl}^{\text{score}}(q_t)$, which enables ranking potential measurement locations to optimize exploration efficiency. In cases where the measurement space consists of pixels, $\hat{x}_t^{(i)}$ represents predicted estimates of the full image based on observed pixels $\tilde{x}_{t-1}$ so far.

$$\text{expl}^{\text{score}}(q_t) = \left[\sum_{i=0}^{N_B}\sum_{j=0}^{N_B} \frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2}\right] \tag{6}$$

We now delve into how *DiffATD* leverages the observations collected so far to address exploitation effectively. To this end, *DiffATD*: (i) employs a *reward model* (r_ϕ) parameterized by ϕ, which is incrementally trained on supervised data obtained from measurements, enabling it to predict whether a specific measurement location belongs to the target of interest; (ii) estimates $\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})]_{q_t}$,

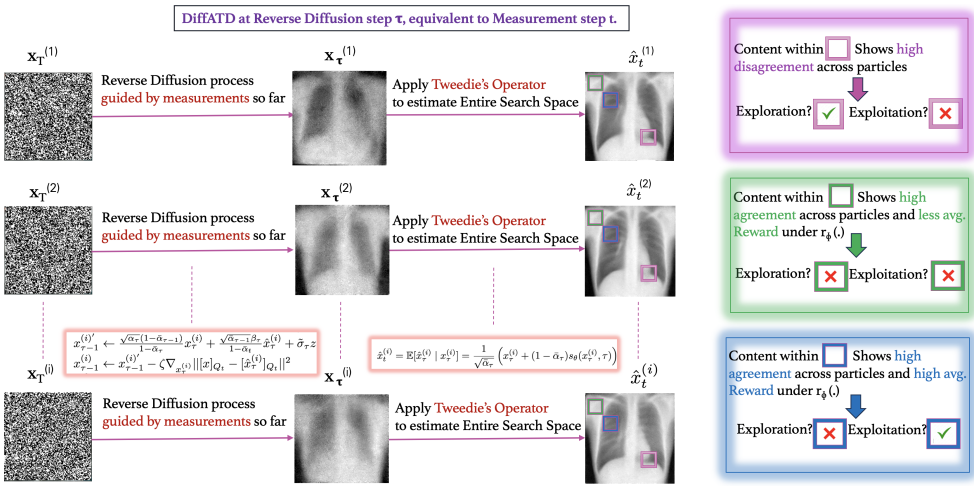

Figure 2: An Overview of *DiffATD*.

representing the expected log-likelihood score at a measurement location q_t, denoted as $\text{likeli}^{\text{score}}(q_t)$, as evaluating it is essential for prioritizing potential measurement locations from an exploitation standpoint. The following theorem establishes an expression for computing $\text{likeli}^{\text{score}}(q_t)$.

Theorem 2. *The expected log-likelihood score at a measurement location q_t can be expressed:*

$$\text{likeli}^{\text{score}}(q_t) = \sum_{i=0}^{N_B}\sum_{j=0}^{N_B} \exp\left\{-\frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2}\right\}$$

We present the derivation of Theorem 2 in the Appendix. $\text{likeli}^{\text{score}}(q_t)$ serves as a key factor in calculating the exploitation score at measurement location q_t, where the exploitation score, $\text{exploit}^{\text{score}}(q_t)$, is defined as *the reward-weighted expected log-likelihood*, as shown below:

$$\text{exploit}^{\text{score}}(q_t) = \underbrace{\text{likeli}^{\text{score}}(q_t)}_{\textit{Expected log-likelihood}} \times \underbrace{\sum_{i=0}^{N_B} r_\phi([\hat{x}_t^{(i)}]_{q_t})}_{\textit{reward}} \tag{7}$$

According to Eqn. 7, a measurement location q_t is preferred for exploitation if it satisfies two conditions: (1) the predicted content at q_t, across the *particles*, is highly likely to correspond to the target of interest, as determined by the reward model (r_ϕ) that predicts whether a specific measurement location belongs to the target; (2) the predicted content at q_t, denoted as $[\hat{x}_t^{(i)}]_{q_t}$, shows the highest degree of consensus among the *particles*, indicating that the content at q_t is highly predictable. Additionally, we demonstrate that maximizing the exploitation score is synonymous with maximizing the expected reward at the current time step, highlighting that the exploitation score operates as a purely greedy strategy. We establish the relationship between $\text{exploit}^{\text{score}}(q_t)$ and a pure greedy strategy and present the proof in the Appendix.

Theorem 3. *Maximizing $\text{exploit}^{\text{score}}(q_t)$ w.r.t a measurement location q_t is equivalent to maximizing expected reward as stated below:* $\arg\max_{q_t}\left[\text{exploit}^{\text{score}}(q_t)\right] \equiv \arg\max_{q_t} \mathbb{E}_{\hat{x}_t \sim p(\hat{x}_t|Q_t,\tilde{x}_{t-1})}[r_\phi[\hat{x}_t]_{q_t}]$.

It is important to note that we randomly initialize the reward model parameters (ϕ) and update them progressively using *binary cross-entropy loss* as new supervised data is acquired after each measurement. The training objective of reward model r_ϕ at measurement step t is defined as below:

$$\min_\phi [\sum_{i=1}^{t} -(y^{(q_i)} \log(r_\phi([x]_{q_t})) + (1 - y^{(q_i)}) \log(1 - r_\phi([x]_{q_t})) \tag{8}$$

Initially, the reward model is imperfect, making the exploitation score unreliable. However, this is not a concern, as *DiffATD* prioritizes exploration in the early stages of the discovery process. Next, we integrate the components to derive the proposed *DiffATD* sampling strategy for efficient active target discovery. This involves ranking each potential measurement location (q_t) at time t by considering both exploration and exploitation scores, as defined below:

$$\text{Score}(q_t) = \kappa(\mathcal{B}) \cdot \text{expl}^{\text{score}}(q_t) + (1 - \kappa(\mathcal{B})) \cdot \text{likeli}^{\text{score}}(q_t) \tag{9}$$

Here, $\kappa(\mathcal{B})$ is a function of the measurement budget that enables *DiffATD* to achieve a budget-aware balance between exploration and exploitation. After computing the scores (as defined in 9) for each measurement location q_t, *DiffATD* selects the location with the highest score for sampling. The functional form of $\kappa(\mathcal{B})$ is a design choice that depends on the specific problem. For instance, a simple approach is to define $\kappa(\mathcal{B})$ as a linear function of the budget, such as $\kappa(\mathcal{B}) = \frac{\mathcal{B}-t}{\mathcal{B}+t}$. We empirically demonstrate that this simple choice of $\kappa(\mathcal{B})$ is highly effective across a wide range of application domains. We detail the *DiffATD* algorithm in 2 and an overview in Figure 2.

Algorithm 2 DiffATD Algorithm

Require: $T, N_B, s_\theta, \zeta, \{\alpha_\tau\}_{\tau=0}^{T}, \{\tilde{\sigma}_\tau\}_{\tau=0}^{T}, M, Q_0 = \emptyset, \mathcal{B}$

$t = 0$; $\{x_T^{(i)} \sim \mathcal{N}(0, I)\}_{i=0}^{N_B}; z \sim \mathcal{N}(0, I), \tilde{x}_0 = 0, \mathcal{D}_t = \emptyset, \{k\} =$ Set of All measurements, $R = 0$
for $\tau = T$ to 1 **do**
  for $i = 0$ to N_B **do**
    $\hat{x}_\tau^{(i)} \leftarrow \mathcal{T}_\tau(x_\tau^{(i)}) = \frac{1}{\sqrt{\bar{\alpha}_\tau}} \left(x_\tau^{(i)} + (1 - \bar{\alpha}_\tau)\hat{s}^{(i)} \right)$; Where $\hat{s}^{(i)} \leftarrow s_\theta(x_\tau^{(i)}, \tau)$
    $x_{\tau-1}^{(i)'} \leftarrow \frac{\sqrt{\alpha_\tau}(1-\bar{\alpha}_{\tau-1})}{1-\bar{\alpha}_\tau} x_\tau^{(i)} + \frac{\sqrt{\bar{\alpha}_{\tau-1}}\beta_\tau}{1-\bar{\alpha}_t} \hat{x}_\tau^{(i)} + \tilde{\sigma}_\tau z$
    $x_{\tau-1}^{(i)} \leftarrow x_{\tau-1}^{(i)'} - \zeta \nabla_{x_\tau^{(i)}} ||[x]_{Q_t} - [\hat{x}_\tau^{(i)}]_{Q_t}||^2$
  end for
  if $\tau \in M$ and $\mathcal{B} > 0$ **then**
    Compute $\text{expl}^{\text{score}}(q_t)$ and $\text{exploit}^{\text{score}}(q_t)$ using Eqn. 6 and 7 respectively for each $q_t \in k$.
    Compute $\text{score}(q_t)$ using Eqn. 9 for each $q_t \in k$ and sample a location q_t with the highest score.
    $t \leftarrow t + 1, \mathcal{B} \leftarrow \mathcal{B} - 1, \{k\} \leftarrow \{k\} \setminus q_t$
    Update: $Q_t \leftarrow Q_{t-1} \cup q_t, \tilde{x}_t \leftarrow \tilde{x}_{t-1} \cup [x]_{q_t}$.; Update: $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{[x]_{q_t}, y^{(q_t)}\}, R$ += $y^{(q_t)}$
    Train r_ϕ with updated $\mathcal{D}_t$ and optimize ϕ with 8.
  end if
end for
Return R.

6 Experiments

Evaluation metrics. Since ATD seeks to maximize the identification of measurement locations containing the target of interest, we assess performance using the ***success rate (SR) of selecting measurement locations that belong to the target*** during exploration in partially observable environments. Therefore, SR is defined as:

$$\text{SR} = \frac{1}{L} \sum_{i=1}^{L} \frac{1}{\min\{\mathcal{B}, U_i\}} \sum_{t=1}^{\mathcal{B}} y_i^{(q_t)}; L = \text{number of tasks.} \tag{10}$$

Here, U_i denotes the maximum number of measurement locations containing the target in the i-th search task. We evaluate *DiffATD* and the baselines across different measurement budgets $\mathcal{B} \in \{150, 200, 250, 300\}$ for various target categories and application domains, spanning from medical imaging to remote sensing. Our code and models are publicly available at this link.

Baselines. We compare our proposed *DiffATD* policy to the following baselines:

- ***Random Search (RS)***, in which unexplored measurement locations are selected uniformly at random.
- ***Max-Ent*** approach ranks a set of unexplored measurement locations based on $\text{expl}^{\text{score}}(q_t)$ and then selects q_t corresponding to the maximum value of $\text{expl}^{\text{score}}(q_t)$.
- ***Greedy-Adaptive (GA)*** approach ranks a set of unexplored measurement locations based on $\text{exploit}^{\text{score}}(q_t)$ and then selects q_t corresponding to the maximum value of $\text{exploit}^{\text{score}}(q_t)$. Like *DiffATD*, GA approach updates (r_ϕ) using Equation 8 after each observation.

Active Discovery of Objects from Overhead Imagery We begin the evaluation by comparing the performance of *DiffATD* with the baselines in terms of SR. For this comparison, we consider various targets (e.g., truck, plane) from the DOTA dataset [29] and present the results in Table 1. The empirical outcome indicates that *DiffATD* significantly outperforms all baselines, with improvements ranging from **7.01%** to **17.23%** over the most competitive method across all measurement budgets. We also present visualizations of *DiffATD*'s exploration strategy in Fig. 3, demonstrating its effectiveness in ATD. As depicted in Fig. 3, *DiffATD* efficiently explores the search space, uncovering distinct target clusters within the measurement budget, highlighting its effectiveness in balancing exploration and exploitation. Similar visualizations with different target categories are in the Appendix.

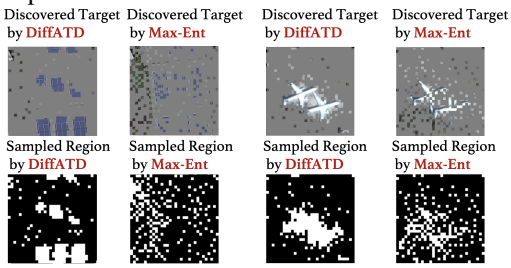

Figure 3: Active Discovery of Plane, Playground.

Table 1: *SR* Comparison with Baseline Approaches

Active Discovery with targets, e.g., plane, truck			
Method	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$
RS	0.1990	0.2487	0.2919
Max-Ent	0.4625	0.5524	0.6091
GA	0.4586	0.5961	0.6550
DiffATD	**0.5422**	**0.6379**	**0.7309**

Active Discovery of Species Next, we evaluate the efficacy of *DiffATD* in uncovering an unknown species from iNaturalist [30]. The target species distribution is formed by dividing a large geospatial region into equal-sized grids and counting target species occurrences on each grid, as detailed in Appendix. We compare *DiffATD* with the baselines in terms of *SR*, and across varying $\mathcal{B}$. The results are presented in Table 2. We observe significant improvements in the performance of the proposed *DiffATD* approach compared to all baselines in each measurement budget setting, ranging from **8.48%** to **13.86%** improvement relative to the most competitive method. In Fig. 4, we present a qualitative comparison of the exploration strategies of *DiffATD* and GA, illustrating that *DiffATD* uncovers more target clusters compared to the GA approach within a fixed sampling budget.

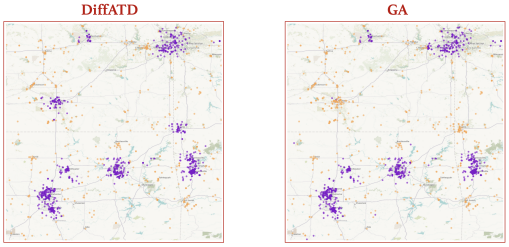

Figure 4: ATD of Species (purple → targets).

Table 2: *SR* Comparison with Baseline Approaches

Active Discovery of Coccinella Septempunctata			
Method	$\mathcal{B} = 100$	$\mathcal{B} = 150$	$\mathcal{B} = 200$
RS	0.1241	0.1371	0.1932
Max-Ent	0.4110	0.5044	0.5589
GA	0.4345	0.5284	0.5826
DiffATD	**0.4947**	**0.5732**	**0.6401**

Active Discovery of Skin Disease Next, we test the impact of *DiffATD* in the medical imaging setting. To this end, we compare the performance of *DiffATD* with the baselines in terms of *SR* using target classes from the skin imaging dataset [31]. We compare the performance across varying measurement budgets $\mathcal{B}$. The results are presented in Table 3. We observe significant improvements in the performance of the proposed *DiffATD* approach compared to all baselines in each measurement budget setting, ranging from **7.92%** to **10.18%** improvement relative to the most competitive method. In Fig. 5, we present a qualitative comparison of the exploration strategies of *DiffATD* and Max-Ent. As shown in Fig. 5, *DiffATD* efficiently explores the search space, uncovering two distinct target clusters within the measurement budget, highlighting its effectiveness in balancing exploration and exploitation. Several such visualizations and the results with other datasets are in Appendix.

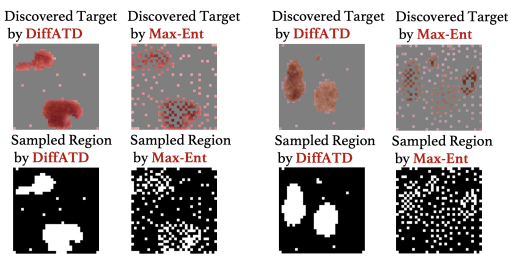

Figure 5: Active Discovery of Skin Disease.

Table 3: *SR* Comparison with Baselines.

Active Discovery of Malignant Skin Diseases			
Method	$\mathcal{B} = 150$	$\mathcal{B} = 200$	$\mathcal{B} = 250$
RS	0.2777	0.2661	0.2695
Max-Ent	0.5981	0.5816	0.5717
GA	0.8396	0.8261	0.7943
DiffATD	**0.9061**	**0.8974**	**0.8752**

Active Discovery of Bone Suppression Next, we evaluate *DiffATD* on the Chest X-Ray dataset [32], with results shown in Table 4. The findings follow a similar trend to previous datasets, with *DiffATD* achieving notable performance improvements of **41.08%** to **59.22%** across all measurement budgets. These empirical outcomes further reinforce the efficacy of *DiffATD* in active target discovery. Visualizations of *DiffATD*'s exploration strategy are provided in Fig. 6, with additional examples are presented in the Appendix.

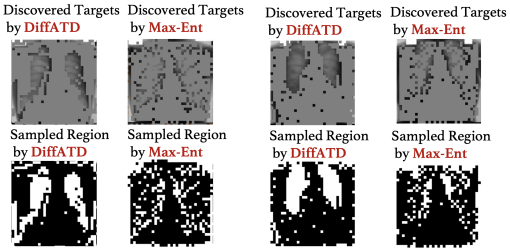

Figure 6: Active Discovery of Bone Suppression.

Table 4: *SR* Comparison with Chest X-Ray Dataset

Active Discovery of Bone Suppression			
Method	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$
RS	0.1925	0.2501	0.2936
Max-Ent	0.1643	0.2194	0.2616
GA	0.1607	0.2010	0.2211
DiffATD	**0.3065**	**0.3733**	**0.4142**

Qualitative Comparison with GA Approach We compare the exploration strategy of *DiffATD* with that of *Greedy-Adaptive* (GA) using CelebA samples, as shown in Fig. 7. *DiffATD* achieves a better balance between exploration and exploitation, leading to higher success rates and more efficient target discovery within a fixed budget. In contrast, GA focuses solely on exploitation, relying on an incrementally learned reward model r_ϕ. These visual comparisons highlight the limitations of purely greedy approaches, such as GA, especially when targets have complex or irregular structures with spatially isolated regions. **GA often struggles early in the search, when training samples are limited, the reward model tends to overfit noise, misguiding the discovery process.** In contrast, *DiffATD* minimizes early reliance on reward-based exploitation, gradually increasing it as more measurements are collected and the reward model strengthens. This adaptive process makes *DiffATD* more resilient and reliable than GA. More qualitative results are in the Appendix.

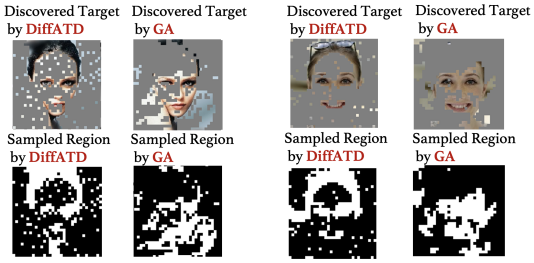

Figure 7: Active Discovery of Eye, hair, nose, lip.

Effect of $\kappa(\mathcal{B})$ We conduct experiments to assess the impact of $\kappa(\mathcal{B})$ on *DiffATD*'s active discovery performance. Specifically, we investigate how amplifying the exploration weight, by setting $\kappa(\mathcal{B}) = \max\{0, \kappa(\alpha \cdot \mathcal{B})\}$ with $\alpha > 1$, and enhancing the exploitation weight by setting $\alpha < 1$, influence the overall effectiveness of the approach. We present results for $\alpha \in \{0.2, 1, 5\}$ using the DOTA and Skin datasets, as shown in Table 5. The best performance is achieved with $\alpha = 1$, and the results suggest that **extreme values of α (either too low or too high) hinder performance**, as both exploration and exploitation are equally crucial to achieve superior performance.

Table 5: Effect of $\kappa(\mathcal{B})$

Performance across varying α with $\mathcal{B} = 200$			
Dataset	$\alpha = 0.2$	$\alpha = 1.0$	$\alpha = 5.0$
DOTA	0.5052	**0.5422**	0.4823
Skin	0.8465	**0.8974**	0.8782

Comparison with Supervised and Fully Observable Approach To evaluate the efficacy of the proposed *DiffATD* framework, we compare its performance to a fully supervised SOTA semantic segmentation approach, SAM [33], that operates under full observability of the search space, referred to as *FullSEG*. Note that during inference, *FullSEG* selects the top $\mathcal{B}$ most probable target regions for measurement in a one-shot manner. We present the result for diverse target categories with $\mathcal{B} = 300$ in Table 6. We observe that ***DiffATD* outperforms SAM for rare targets**, such as skin diseases, while performing comparably to *FullSEG* for non-rare targets like vehicles, planes, etc., despite never observing the full search area and being trained entirely in an unsupervised manner, unlike *FullSEG*. This further showcases the strength of *DiffATD* in active target discovery in a partially observable environment.

Active Discovery of Targets		
Method	Skin	DOTA
FullSEG	0.6221	**0.8731**
DiffATD	**0.8642**	0.7309

Table 6: SR Comparison.

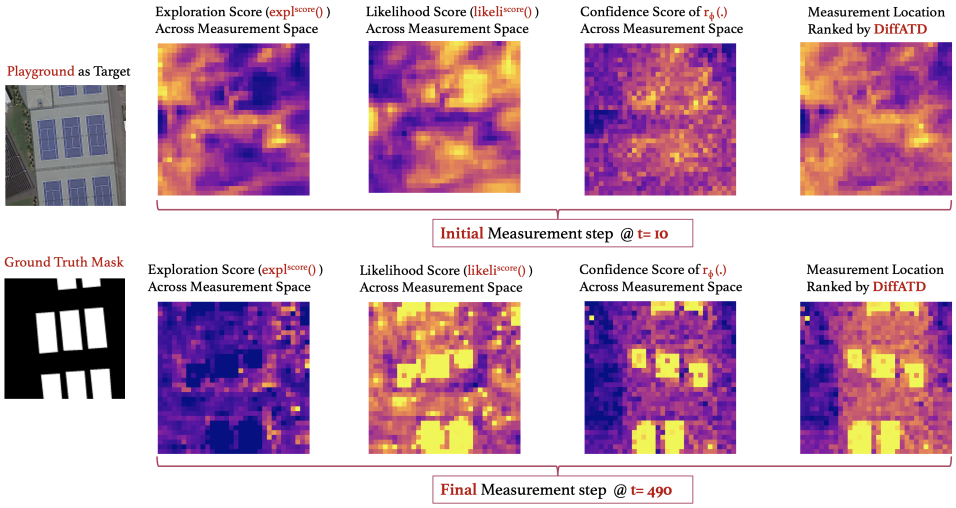

Figure 8: Explanation of *DiffATD*. Brighter indicates a higher value.

Visualizing *DiffATD*'s Exploration Behavior at Different Stage We illustrate the behavior of *DiffATD* with an example. As shown in Fig. 8, during the initial search phase, *DiffATD* prioritizes the exploration score when ranking measurement locations. However, as the search nears its final stages, rankings are increasingly driven by the exploitation score, as defined in Eqn. 7. **As the search progresses, the confidence of r_ϕ and $\text{likeli}^{\text{score}}(.)$ become more accurate, making the $\text{exploit}^{\text{score}}(.)$ more reliable, which justifies *DiffATD*'s increasing reliance on it in the later phase.** More such visualizations are presented in the Appendix.

Comparison With Vision-Language Model We conduct comparisons between DiffATD and two State-of-the-art VLMs: GPT-4o and Gemini, on the DOTA dataset. The corresponding empirical results are summarized in the following Table 7. We observe that DiffATD consistently outperforms VLM baselines across different measurement budgets. **These findings underscore DiffATD's advantage over alternatives like VLMs in the ATD setting, driven by its principled balance of exploration and exploitation based on the maximum entropy framework.** Additionally, in Figure 9, we present comparative exploration strategies for different targets from DOTA to provide further intuition.

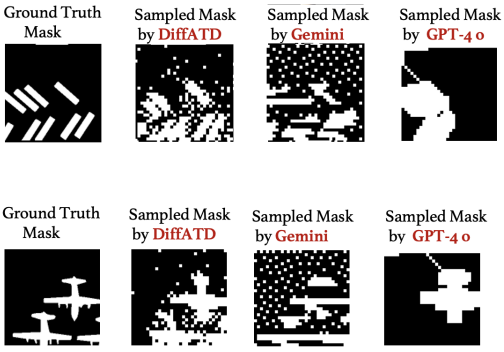

Figure 9: Active Discovery of Plane, Truck.

Table 7: Quantitative Comparison with VLM using DOTA Overhead Imagery Dataset

Active Discovery with targets, e.g., plane, truck			
Method	$\mathcal{B} = 150$	$\mathcal{B} = 180$	$\mathcal{B} = 300$
GPT4-o	0.3383	0.3949	0.5678
Gemini	0.4027	0.4608	0.6453
DiffATD	**0.4537**	**0.5092**	**0.7309**

7 Conclusions

We present Diffusion-guided Active Target Discovery (*DiffATD*), a novel approach that harnesses diffusion dynamics for efficient target discovery in partially observable environments within a fixed sampling budget. *DiffATD* eliminates the need for supervised training, addressing a key limitation of prior approaches. Additionally, *DiffATD* offers interpretability, in contrast to existing black-box policies. Our extensive experiments across diverse domains demonstrate its efficacy. We believe our work will inspire further exploration of this impactful problem and encourage the adoption of *DiffATD* in real-world applications across diverse fields.

Acknowledgement This research was partially supported by the National Science Foundation (CCF-2403758, IIS-2214141), Army Research Office (W911NF2510059), Office of Naval Research (N000142412663), Amazon.

References

[1] Dinesh Jayaraman and Kristen Grauman. Look-ahead before you leap: end-to-end active recognition by forecasting the effect of motion. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*, pages 489–505. Springer, 2016.

[2] Dinesh Jayaraman and Kristen Grauman. Learning to look around: Intelligently exploring unseen environments for unknown tasks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1238–1247, 2018.

[3] Bo Xiong and Kristen Grauman. Snap angle prediction for 360 panoramas. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–18, 2018.

[4] Iris Huijben, Bastiaan S Veeling, and Ruud JG van Sloun. Deep probabilistic subsampling for task-adaptive compressed sensing. In *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[5] Cagla D Bahadir, Alan Q Wang, Adrian V Dalca, and Mert R Sabuncu. Deep-learning-based optimization of the under-sampling pattern in mri. *IEEE Transactions on Computational Imaging*, 6:1139–1152, 2020.

[6] Hans Van Gorp, Iris Huijben, Bastiaan S Veeling, Nicola Pezzotti, and Ruud JG Van Sloun. Active deep probabilistic subsampling. In *International Conference on Machine Learning*, pages 10509–10518. PMLR, 2021.

[7] Tim Bakker, Herke van Hoof, and Max Welling. Experimental design for mri by greedy policy search. *Advances in Neural Information Processing Systems*, 33:18954–18966, 2020.

[8] Tianwei Yin, Zihui Wu, He Sun, Adrian V Dalca, Yisong Yue, and Katherine L Bouman. End-to-end sequential sampling and reconstruction for mri. *arXiv preprint arXiv:2105.06460*, 2021.

[9] Tristan SW Stevens, Nishith Chennakeshava, Frederik J de Bruijn, Martin Pekař, and Ruud JG van Sloun. Accelerated intravascular ultrasound imaging using deep reinforcement learning. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1216–1220. IEEE, 2022.

[10] Thomas Sanchez, Igor Krawczuk, Zhaodong Sun, and Volkan Cevher. Closed loop deep bayesian inversion: Uncertainty driven acquisition for fast mri. 2020.

[11] Oisin Nolan, Tristan SW Stevens, Wessel L van Nierop, and Ruud JG van Sloun. Active diffusion subsampling. *arXiv preprint arXiv:2406.14388*, 2024.

[12] Burak Uzkent and Stefano Ermon. Learning when and where to zoom with deep reinforcement learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12345–12354, 2020.

[13] Anindya Sarkar, Michael Lanier, Scott Alfeld, Jiarui Feng, Roman Garnett, Nathan Jacobs, and Yevgeniy Vorobeychik. A visual active search framework for geospatial exploration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 8316–8325, 2024.

[14] Anindya Sarkar, Nathan Jacobs, and Yevgeniy Vorobeychik. A partially-supervised reinforcement learning framework for visual active search. *Advances in Neural Information Processing Systems*, 36:12245–12270, 2023.

[15] Quan Nguyen, Anindya Sarkar, and Roman Garnett. Amortized nonmyopic active search via deep imitation learning. *arXiv preprint arXiv:2405.15031*, 2024.

[16] Roman Garnett, Yamuna Krishnamurthy, Xuehan Xiong, Jeff Schneider, and Richard Mann. Bayesian optimal active search and surveying. *arXiv preprint arXiv:1206.6406*, 2012.

[17] Shali Jiang, Gustavo Malkomes, Geoff Converse, Alyssa Shofner, Benjamin Moseley, and Roman Garnett. Efficient nonmyopic active search. In *International Conference on Machine Learning*, pages 1714–1723. PMLR, 2017.

[18] Shali Jiang, Roman Garnett, and Benjamin Moseley. Cost effective active search. *Advances in Neural Information Processing Systems*, 32, 2019.

[19] Samrudhdhi B Rangrej, Chetan L Srinidhi, and James J Clark. Consistency driven sequential transformers attention model for partially observable scenes. *arXiv preprint arXiv:2204.00656*, 2022.

[20] Aleksis Pirinen, Anton Samuelsson, John Backsund, and Kalle Aström. Aerial view goal localization with reinforcement learning. *arXiv preprint arXiv:2209.03694*, 2022.

[21] Anindya Sarkar, Srikumar Sastry, Aleksis Pirinen, Chongjie Zhang, Nathan Jacobs, and Yevgeniy Vorobeychik. Gomaa-geo: Goal modality agnostic active geo-localization. *arXiv preprint arXiv:2406.01917*, 2024.

[22] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[23] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

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

[25] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

[26] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.

[27] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

[28] John R Hershey and Peder A Olsen. Approximating the kullback leibler divergence between gaussian mixture models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, pages IV–317. IEEE, 2007.

[29] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018.

[30] *iNaturalist*. https://www.inaturalist.org/.

[31] Veronica Rotemberg, Nicholas Kurtansky, Brigid Betz-Stablein, Liam Caffery, Emmanouil Chousakos, Noel Codella, Marc Combalia, Stephen Dusza, Pascale Guitera, David Gutman, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific data*, 8(1):34, 2021.

[32] Bram Van Ginneken, Mikkel B Stegmann, and Marco Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical image analysis*, 10(1):19–40, 2006.

[33] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[34] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5549–5558, 2020.

[35] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: section 5 discusses our algorithm and section 6 includes the results from our experiments, both of which claims in the abstract and introduction reflect.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [NA]

   Justification: We design a novel algorithm for a new task.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide complete and detailed theoretical proof in section 5 and Appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We discuss the experiment details in main paper and in the Appendix. We also provide the code for reference.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We use public available datasets for all our experiments and we provide the code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We discuss these details in the main paper and in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We provide these details in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide these details in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We follow the NeurIPS Code of Ethics. Our research does not involve human participants, and our data were obtained in a way that respects the corresponding licenses, and processed to preserve anonymity.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss the potential impacts of our work in the Appendix, which also includes considerations that should made even when our method is used as intended.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We judge that our work poses no immediate risk of misuse of our released data.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All works whose data are used by us are appropriately credited and the corresponding licenses were respected.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The README file in our code includes documentation on how to use the data and the code.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our work does not involve human participants.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our work does not involve human participants.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: Our method doesn't involve LLM.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

Appendix: Online Feedback Efficient Active Target Discovery in Partially Observable Environments

Overview of the Contents

A Proof of Theoretical Results

A.1 Proof of Proposition 1

Proof.

$$q_t^{\text{exp}} = \arg\max_{q_t} \left[I(\hat{x}_t; x \mid Q_t, \tilde{x}_{t-1}) \right]$$

$$= \arg\max_{q_t} \left[\mathbb{E}_{p(\hat{x}_t|x,Q_t)p(x|\tilde{x}_{t-1})} \left[\log p(\hat{x}_t \mid Q_t, \tilde{x}_{t-1}) - \log p(\hat{x}_t \mid x, Q_t, \tilde{x}_{t-1}) \right] \right]$$

$$= \arg\max_{q_t} \left[H(\hat{x}_t \mid Q_t, \tilde{x}_{t-1}) - \underbrace{H(\hat{x}_t \mid x, Q_t, \tilde{x}_{t-1})}_{\text{remains independent of the } q_t\text{. Therefore, this term can be disregarded when optimizing for } q_t} \right].$$

$$= \arg\max_{q_t} \left[H(\hat{x}_t \mid Q_t, \tilde{x}_{t-1}) \right]$$

□

A.2 Proof of Theorem 1

Proof. We demonstrate that maximizing the marginal entropy of the belief distribution as defined in Equation 5 does not require computing a separate set of particles for every possible choice of measurement location q_t when the action corresponds to selecting a measurement location. Since the measurement locations $Q_t = Q_{t-1} \cup q_t$ only differ in the newly selected indices q_t within the arg max, the elements of each particle $x_t^{(i)}$ remain unchanged across all possible Q_t, except at the indices specified by q_t. Exploiting this structure, we decompose the squared L_2 norm into two parts: one over the indices in q_t and the other over those in Q_{t-1}. The term associated with Q_{t-1} becomes a constant in the arg max and can be disregarded. Consequently, the formulation reduces the computing of the squared L_2 norms exclusively for the elements corresponding to q_t. We denote k as the set of possible measurement locations at step t. Utilizing Equation 5, we can write,

$$q_t^{\text{exp}} = \arg\max_{q_t} \sum_{i=0}^{N_B} \alpha_i \log \sum_{j=0}^{N_B} \alpha_j \exp \left\{ \frac{||\hat{x}_t^{(i)} - \hat{x}_t^{(j)}||_2^2}{2\sigma_x^2} \right\}$$

$$q_t^{\text{exp}} = \arg\max_{q_t} \sum_{i,j} \log \left(\exp \left(\frac{\|\hat{x}_t^{(i)} - \hat{x}_t^{(j)}\|^2}{2\sigma_x^2} \right) \right) \text{(By assuming, } \alpha_i = \alpha_j, \, \forall i, j)$$

$$q_t^{\text{exp}} = \arg\max_{q_t} \sum_{i,j} \log \left(\exp \left(\frac{\sum_{a \in Q_t} ([\hat{x}_t^{(i)}]_a - [\hat{x}_t^{(j)}]_a)^2}{2\sigma_x^2} \right) \right)$$

$$q_t^{\text{exp}} = \arg\max_{q_t} \sum_{i,j} \log \left(\exp \left(\frac{\sum_{q_t \in k} ([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2 + \sum_{r \in Q_{t-1}} ([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right) \right)$$

$$q_t^{\text{exp}} = \arg\max_{q_t} \sum_{i,j} \log \left(\prod_{q_t \in k} \exp \left(\frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} \right) \prod_{r \in Q_{t-1}} \exp \left(\frac{([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right) \right)$$

$$q_t^{\text{exp}} \propto \arg\max_{q_t} \sum_{i,j} \left(\sum_{q_t \in k} \frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} + \underbrace{\sum_{r \in Q_{t-1}} \frac{([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2}}_{\text{We can ignore as it doesn't depend on the choice of } q_t.} \right)$$

$$q_t^{\text{exp}} \propto \arg\max_{q_t} \sum_{i,j} \sum_{q_t \in k} \frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2}.$$

$$q_t^{\text{exp}} = \arg\max_{q_t} \left[\sum_{i=0}^{N_B} \log \sum_{j=0}^{N_B} \exp\left(\frac{\sum_{q_t \in k} ([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} \right) \right]$$

□

A.3 Proof of Theorem 3

A pure greedy strategy selects the measurement location that maximizes the expected reward at the current time step, as defined below:

$$= \arg\max_{q_t} \mathbb{E}_{\hat{x}_t \sim p(\hat{x}_t|Q_t, \tilde{x}_{t-1})}[r_\phi[\hat{x}_t]_{q_t}]$$

$$= \arg\max_{q_t} \underbrace{\int p(\hat{x}_t|Q_t, \tilde{x}_{t-1}) r_\phi[\hat{x}_t]_{q_t} d\hat{x}_t}_{J(q_t)}$$

In order to maximize J(q_t) w.r.t q_t, we do the following:

$$\nabla_{q_t} J(q_t) = \arg\max_{q_t} \int \nabla_{q_t} p(\hat{x}_t|Q_t, \tilde{x}_{t-1}) r_\phi[\hat{x}_t]_{q_t} d\hat{x}_t$$

$$= \arg\max_{q_t} \int p(\hat{x}_t|Q_t, \tilde{x}_{t-1}) \nabla_{q_t} \log p(\hat{x}_t|Q_t, \tilde{x}_{t-1}) r_\phi[\hat{x}_t]_{q_t} d\hat{x}_t$$

We obtain the last equality using the following identity:

$$p_\theta(x) \nabla_\theta \log p_\theta(x) = \nabla_\theta p_\theta(x)$$

We simplify the expression using the definition of expectation, as shown below:

$$= \arg\max_{q_t} \mathbb{E}_{\hat{x}_t \sim p(\hat{x}_t|Q_t, \tilde{x}_{t-1})}[\quad \underbrace{\nabla_{q_t} \log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})}_{\text{Maximizing expected log-likelihood score, i.e., } \text{likeli}^{\text{score}}(.).} \quad \underbrace{r_\phi[\hat{x}_t]_{q_t}}_{\text{Reward at current time.}} \quad]$$

By definition, it is equivalent to the following

$$= \arg\max_{q_t} [\text{exploit}^{\text{score}}(q_t)] \qquad \text{(as defined in Equation 7.)}$$

A.4 Proof of Theorem 2

Proof. We start with the definition of entropy H:

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})] = -H(\hat{x}_t|Q_t, \tilde{x}_{t-1})$$

Substituting the expression of $H(\hat{x}_t|Q_t, \tilde{x}_{t-1})$ as defined in Equation 5, and by setting $\alpha_i = \alpha_j = 1$, we obtain:

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})] \propto -\sum_{i=0}^{N_B} \log \sum_{j=0}^{N_B} \exp\left\{ \frac{||\hat{x}_t^{(i)} - \hat{x}_t^{(j)}||_2^2}{2\sigma_x^2} \right\}$$

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})] \propto -\sum_{i,j} \log \left(\exp\left(\frac{\sum_{a \in Q_t} ([\hat{x}_t^{(i)}]_a - [\hat{x}_t^{(j)}]_a)^2}{2\sigma_x^2} \right) \right)$$

Assuming k is the set of potential measurement locations at time step t, and $Q_t = Q_{t-1} \cup q_t$, where $q_t \in k$.

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})] \propto -\sum_{i,j} \log \left(\exp \left(\frac{\sum_{q_t \in k} ([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2 + \sum_{r \in Q_{t-1}} ([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right) \right)$$

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})] \propto -\sum_{i,j} \log \left(\prod_{q_t \in k} \exp \left(\frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} \right) \prod_{r \in Q_{t-1}} \exp \left(\frac{([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right) \right)$$

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})] \propto -\sum_{i,j} \left(\sum_{q_t \in k} \frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} + \sum_{r \in Q_{t-1}} \frac{([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right)$$

Next, we compute the expected log-likelihood at a specified measurement location q_t. Accordingly, we ignore all the terms that do not depend on q_t. This simple observation helps us to simplify the above expression as follows:

$$\underbrace{\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})]\Big|_{q_t}}_{\textit{The expected log-likelihood at a measurement location } q_t} \propto \sum_{i,j} \left(-\frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} \right)$$

Equivalently, we can write the above expression as:

$$\mathbb{E}_{\hat{x}_t}[\log p(\hat{x}_t|Q_t, \tilde{x}_{t-1})]\Big|_{q_t} \propto \left(\underbrace{\sum_{i=0}^{N_B} \sum_{j=0}^{N_B} \exp \left\{ -\frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} \right\}}_{\text{likeli}^{\text{score}}(q_t)} \right)$$

By definition, the L.H.S of the above expression is the same as $\text{likeli}^{\text{score}}(q_t)$. □

B Empirical Analysis of Active Target Discovery across Diverse Domains

B.1 Active Discovery of Different Parts of Human Face

We compare *DiffATD* with the baselines with the human face as the target from the CelebA dataset [34] and report the findings in Table 8. These results reveal a similar trend to that observed with the previous datasets. We observe significant performance gains with *DiffATD* over all baselines, with improvements of 42.60% to 60.79% across all measurement budgets, demonstrating its mastery in balancing exploration and exploitation for effective active target discovery. We present visualizations of *DiffATD*'s exploration strategy in Fig.10, with additional visualizations are in appendix D.

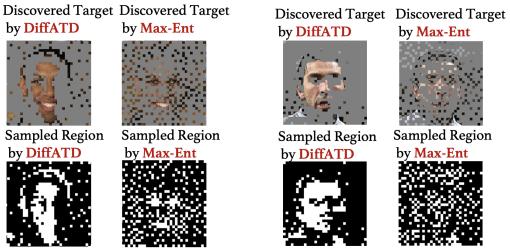

Figure 10: Active Discovery of regions with Face.

Table 8: *SR* Comparison with CelebA Dataset

Active Discovery with face as the target			
Method	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$
RS	0.1938	0.2441	0.2953
Max-Ent	0.2399	0.3510	0.4498
GA	0.2839	0.3516	0.4294
DiffATD	**0.4565**	**0.5646**	**0.6414**

B.2 Active Discovery of Hand-written Digits

Here, we evaluate performance by comparing *DiffATD* with the baselines using the *SR* metric. We compare the performance across varying measurement budgets $\mathcal{B}$. The results are presented in Table 9. We observe significant improvements in the performance of the proposed *DiffATD* approach compared to all baselines in each measurement budget setting, ranging from 16.30% to 45.23% improvement relative to the most competitive method. These empirical results are consistent with those from other datasets we explored, such as DOTA, and CelebA. We present a qualitative comparison of the exploration strategies of *DiffATD* and Max-Ent through visual illustration. As shown in Fig. 11, *DiffATD* efficiently explores the search space, identifying the underlying true handwritten digit within the measurement budget, highlighting its effectiveness in balancing exploration and exploitation. We provide several such visualizations in the following section.

Table 9: *SR* comparison with MNIST Dataset

Active Discovery of Handwritten Digits			
Method	$\mathcal{B} = 100$	$\mathcal{B} = 150$	$\mathcal{B} = 200$
RS	0.1270	0.1518	0.1943
Max-Ent	0.5043	0.6285	0.8325
GA	0.0922	0.4133	0.5820
DiffATD	**0.7324**	**0.8447**	**0.9682**

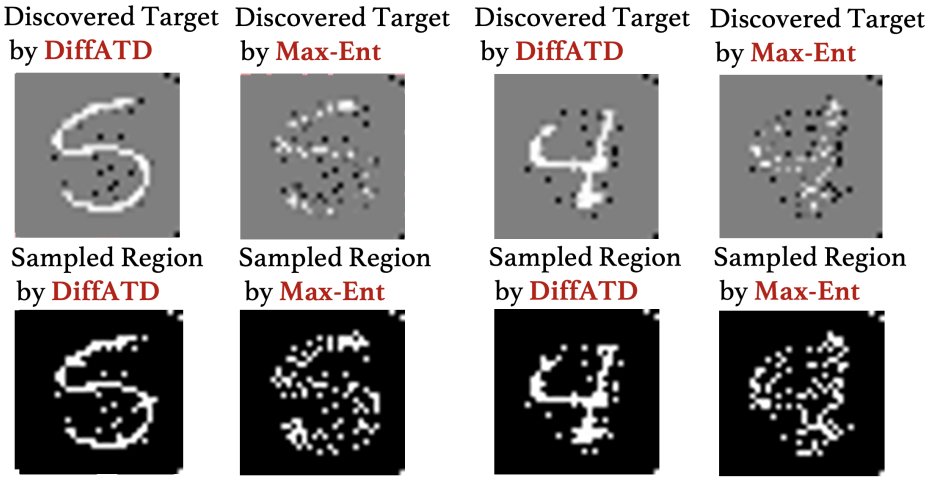

Figure 11: Visualization of Active Discovery of Handwritten Digits.

C Qualitative Comparison with Greedy Adaptive Approach

C.1 Qualitative Comparisons using Lung Disease, Skin Disease, DOTA, and CelebA Images

In this section, we present additional comparative visualizations of *DiffATD*'s exploration strategy against *Greedy-Adaptive* (GA), using samples from diverse datasets, including DOTA, CelebA, and Lung images. These visualizations are shown in Figures 12,13,14, 15. In each case, *DiffATD* strikes a strategic balance between exploration and exploitation, leading to a notably higher success rate and more efficient target discovery within a fixed budget, compared to the greedy strategy, which focuses solely on exploitation based on an incrementally learned reward model r_ϕ. These qualitative observations emphasize key challenges in the active target discovery problem that are difficult to address with entirely greedy strategies like GA. For instance, the greedy approach struggles when the target has a complex and irregular structure with spatially isolated regions. Due to its nature, once it identifies a target in one region, it begins exploiting neighboring areas, guided by the reward model that assigns higher scores to similar regions. As a result, such methods struggle

to discover spatially disjoint targets, as shown in these visualizations. Interestingly, the greedy approach also falters in noisy measurement environments due to its over-reliance on the reward model. Early in the search process, when training samples are limited, the model tends to overfit the noise, misguiding the discovery process. In contrast, our approach minimizes reliance on reward-based exploitation in the early stages. As more measurements are collected, and the reward model becomes incrementally more robust, our method begins to leverage reward-driven exploitation more effectively. This gradual adaptation makes our approach significantly more resilient and reliable, particularly in noisy environments, compared to the greedy strategy.

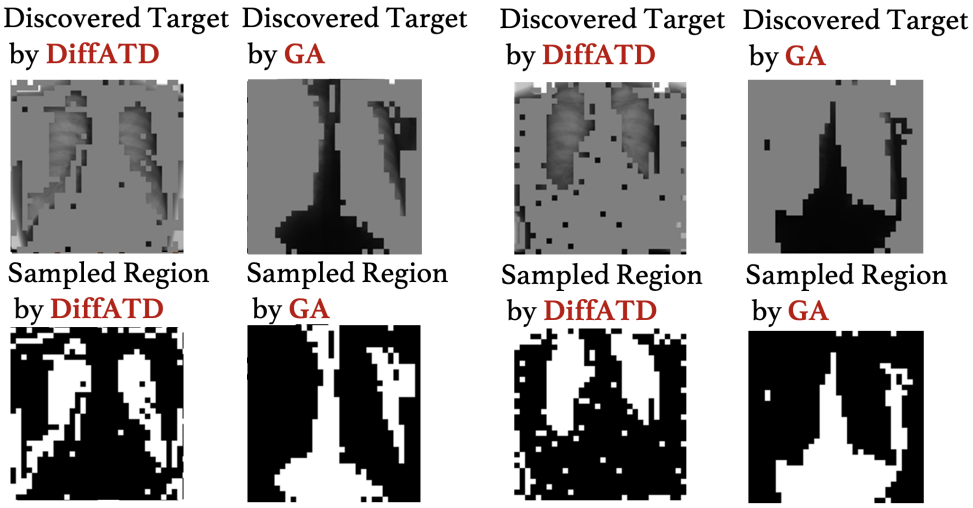

Figure 12: Visualization of Active Discovery of Lung Disease.

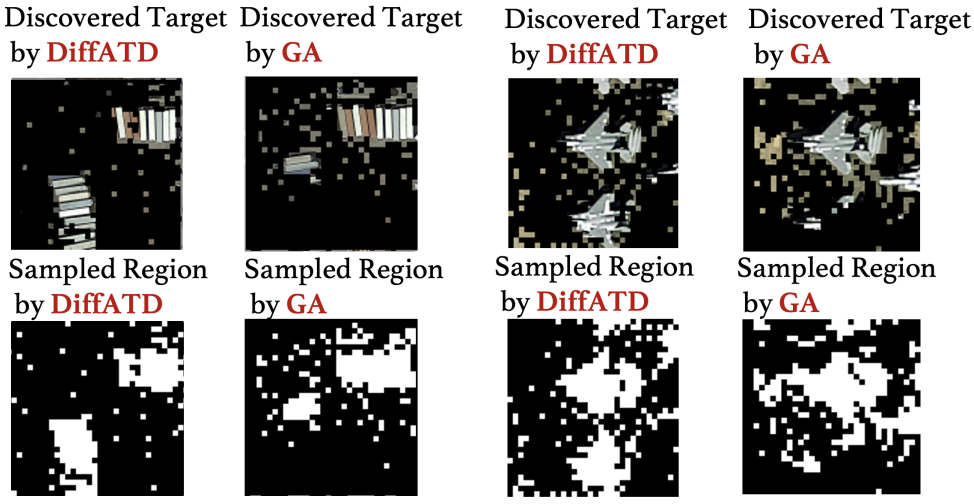

Figure 13: Visualization of Active Discovery of (left) Truck and (right) Plane

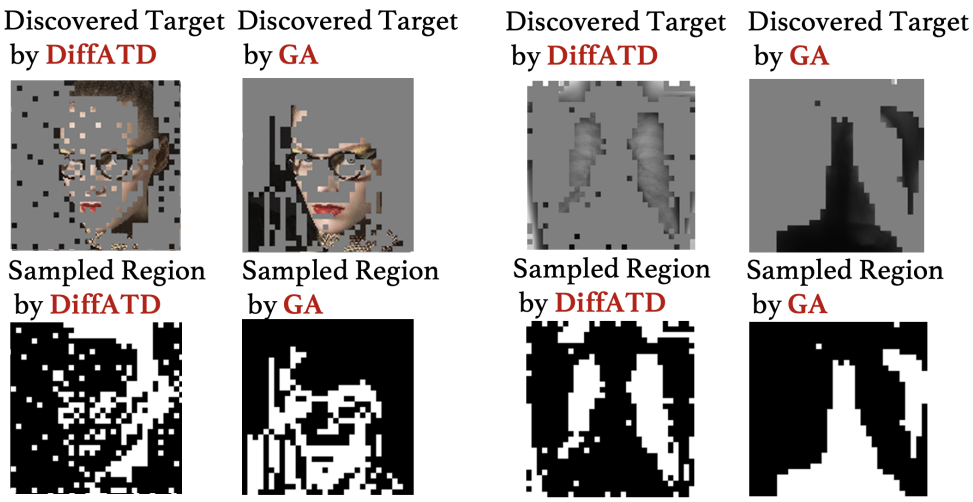

Figure 14: Visualization of Active Discovery of Eye, hair, nose, and lip.

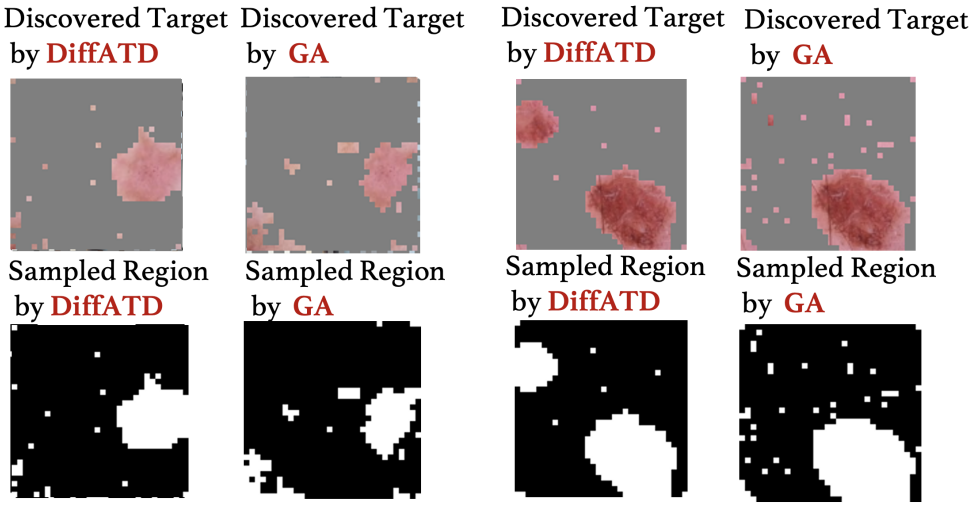

Figure 15: Visualization of Active Discovery of Skin Disease.

D Additional Visualizations of the Exploration Strategy of *DiffATD*

D.1 Comparative Visualizations using Lung Disease, Skin Disease, DOTA, CelebA Images

In this section, we provide further comparative visualizations of *DiffATD*'s exploration strategy versus *Max-Ent*, using samples from a variety of datasets, including DOTA, CelebA, Skin, and Lung images. We present the visualization in figures 16, 17, 18, 19, 20, 21. In each example, *DiffATD* demonstrates a strategic balance between exploration and exploitation, resulting in a significantly higher success rate and more effective target discovery within a predefined budget compared to strategies focused solely on maximizing information gain. These visualizations further reinforce the effectiveness of *DiffATD* in active target discovery within partially observable environments.

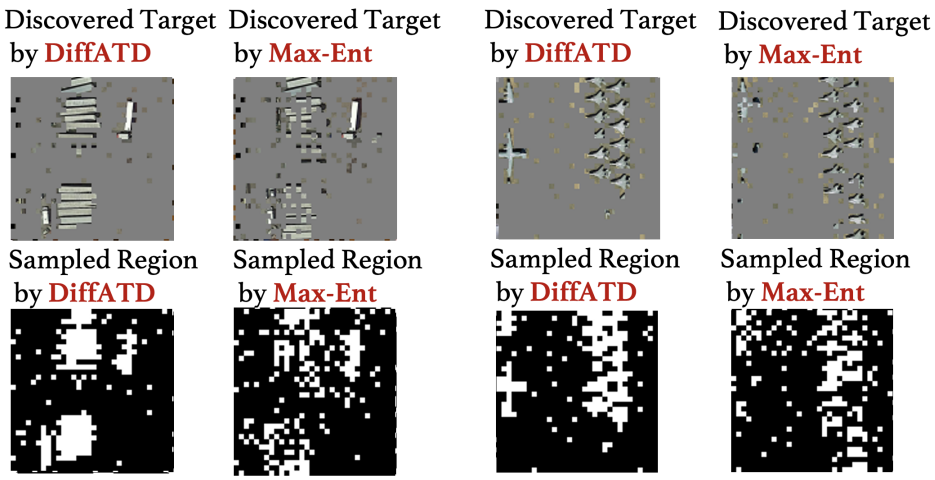

Figure 16: Visualizations of Active Discovery of (left) *large-vehicle*, and (right) *Plane*.

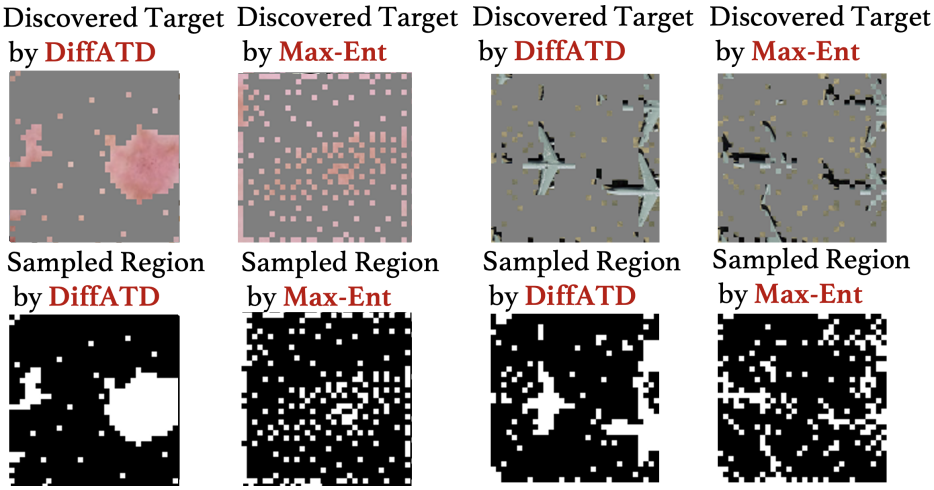

Figure 17: Visualizations of Active Discovery of (left) *Skin disease*, and (right) *Plane*.

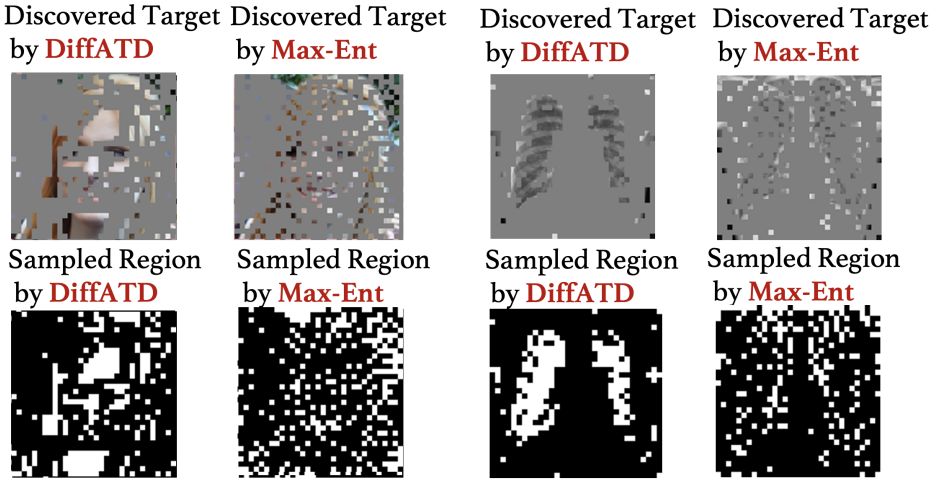

Figure 18: Visualizations of Active Discovery of (left) *human face*, and (right) *lung disease, such as TB*.

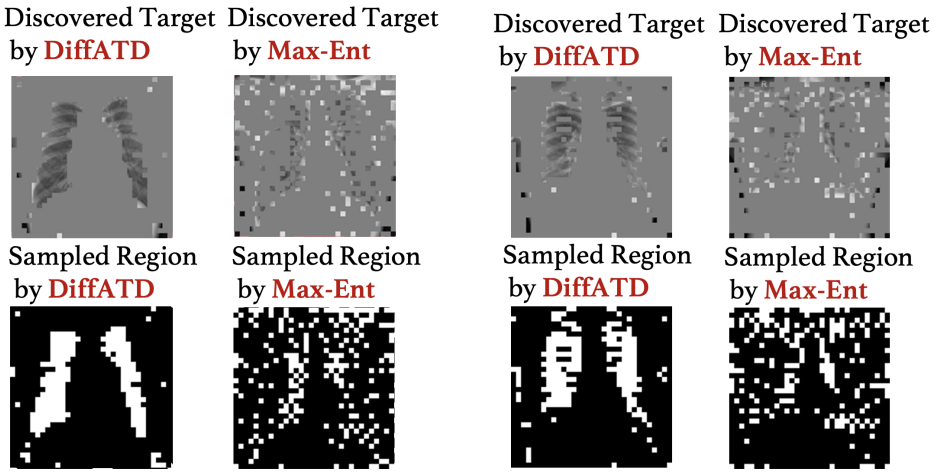

Figure 19: Visualization of Active Discovery of Lung Disease.

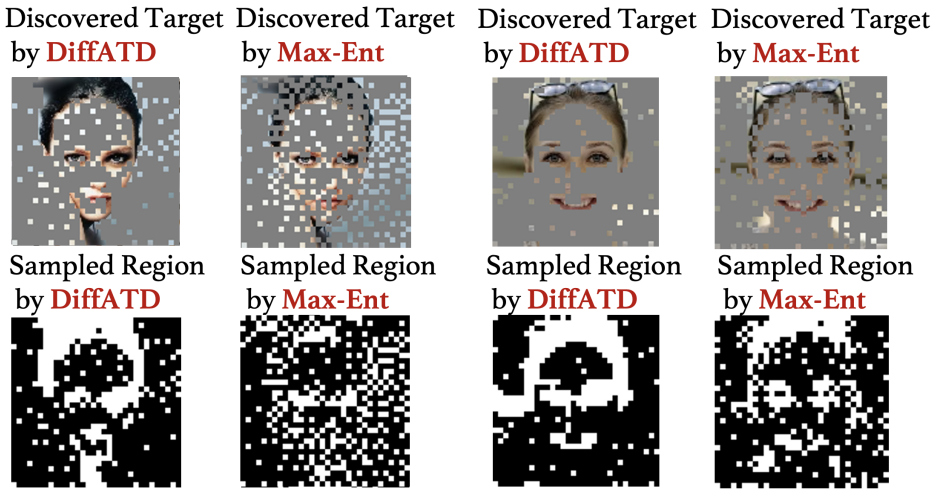

Figure 20: Visualization of Active Discovery of hair, eye, nose, and lip.

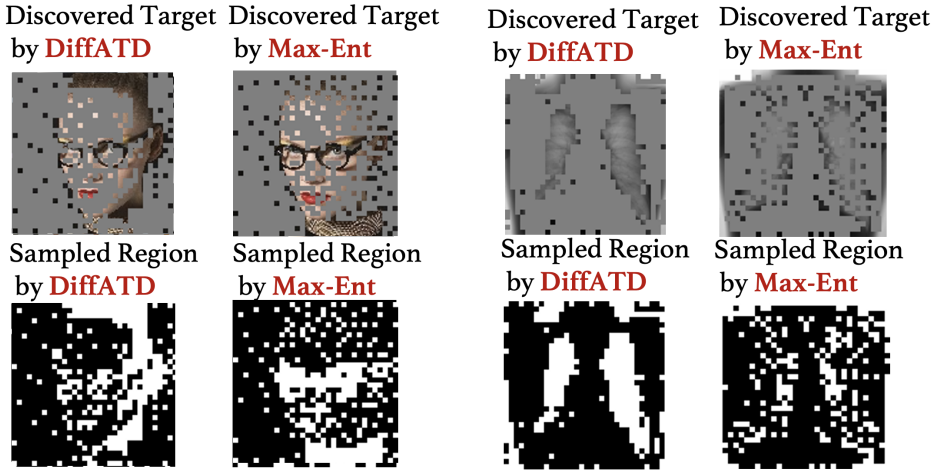

Figure 21: Viz. of Active Discovery of Lung Disease (right) and hair, eye, nose, and lip (left).

E Details of Training and Inference

E.1 Details of Computing Resources, Training, and Inference Hyperparameters

This section provides the training and inference hyperparameters for each dataset used in our experiments. We use DDIM [35] as the diffusion model across datasets. The diffusion models used in different experiments are based on widely adopted U-Net-style architecture. For the MNIST dataset, we use 32-dimensional diffusion time-step embeddings, with the diffusion model consisting of 2 residual blocks. We select the time-step embedding vector dimension to match the input feature size, ensuring the diffusion model can process it efficiently. The block widths are set to $[32, 64, 128]$, and training involves 30 diffusion steps. DOTA, CelebA, and Skin imaging datasets share the same input feature size of $[128, 128, 3]$ and architecture, featuring 128-dimensional time-step embeddings and a diffusion model with 2 residual blocks of width $[64, 128, 256, 256, 512]$. For these datasets, we perform training with 100 diffusion steps. We use 128-dimensional time-step embeddings for the Bone dataset and a diffusion model with 2 residual blocks (each block width: [64, 128, 256, 256]). We use 100 diffusion steps during training. We set the learning rate and weight decay factor to $1e^{-4}$ for all experimental settings. We set a measurement schedule (M) of 100 for a measurement budget ($\mathcal{B}$) of 200, ensuring that $\mathcal{B} \approx \frac{T}{M}$, where T is the total number of reverse diffusion steps, set to approximately 2000 for the DOTA, CelebA, and Skin datasets. Finally, all experiments are implemented in Tensorflow and conducted on NVIDIA A100 40G GPUs. Our training and inference code will be made public.

E.2 Details of Reward Model r_ϕ

Our proposed method, DiffATD, utilizes a parameterized reward model, r_ϕ, to steer the exploitation process. To this end, we employ a neural network consisting of two fully connected layers, with non-linear ReLU activations as the reward model (r_ϕ). The reward model's goal is to predict a score ranging from 0 to 1, where a higher score indicates a higher likelihood that the measurement location corresponds to the target, based on its semantic features. Note that the size of the input semantic feature map for a given measurement location can vary depending on the downstream task. For instance, when working with the MNIST dataset, we use a 1×1 pixel as the input feature, while for other datasets like CelebA, DOTA, Bone, and Skin imaging, we use an 4×4 patch as the input feature size. After each measurement step, we update the model parameters (ϕ) using the objective function outlined in Equation 8. Additionally, the training dataset is updated with the newly observed data point, refining the model's predictions over time. Naturally, as the search advances, the reward model refines its predictions, accurately identifying target-rich regions, which makes it progressively more dependable for informed decision-making. The reward model architecture is consistent across datasets, including Chest X-ray, Skin, CelebA, and DOTA. It consists of 1 convolutional layer with a 3×3 kernel, followed by 5 fully connected (FC) layers, each with its own weights and biases. The first FC layer maps an input of size $\frac{(input\ size)^2}{4}$ to an output of size 4 with weights and biases of size $[\frac{(input\ size)^2}{4}, 4]$ and $[4]$ respectively. The second FC layer transforms an input of dimension 4 to an output of size 32 with a 2-dimensional weight of size $[4, 32]$ and a bias of size $[32]$. The third FC layer maps 32 inputs to 16 outputs via a weight matrix of shape $[32, 16]$ and a bias vector of size $[16]$. The pre-final FC layer transforms inputs of size 16 to outputs of size 8 with $[16, 8]$ weights, and a bias of shape $[8]$. The final FC layer produces an output of size 2, with weights of size $[8, 2]$ and a bias of size $[2]$, representing the target and non-target scores. The reward model uses the leaky ReLU activation function after each layer. We update the reward model parameters after each measurement step based on the objective in Equation 8. The reward model is trained incrementally for 3 epochs after each measurement step using the gathered supervised dataset resulting from sequential observation, with a learning rate of 0.01.

F Challenges in Active Target Discovery for Rare Categories

To highlight the challenges in Active Target Discovery for rare categories, such as skin disease, we conduct an experiment comparing the performance of *DiffATD* with a fully supervised state-of-the-art semantic segmentation model, SAM, which operates under full observability of the search space (referred to as *FullSEG*). During inference, *FullSEG* selects the top $\mathcal{B}$ most probable target regions for measurement in a single pass. We present the results for Skin Disease as the target, with varying

Table 10: Comparison with supervised and fully observable method

Active Discovery of Targets on Skin Images			
Method	$\mathcal{B} = 150$	$\mathcal{B} = 200$	$\mathcal{B} = 250$
FullSEG	0.7304	0.6623	0.6146
DiffATD	**0.9061**	**0.8974**	**0.8752**

budgets $\mathcal{B}$, in Table 10. A significant performance gap is observed across different measurement budgets. These results underscore the challenges in Active Target Discovery for rare categories, as state-of-the-art segmentation models like SAM struggle to efficiently discover rare targets, like skin disease, further demonstrating the effectiveness of *DiffATD*.

G Scalability of DiffATD on Larger Search Spaces

To empirically validate DiffATD, we conduct experiments across diverse domains, including remote sensing (DOTA) and medical imaging (Lung Disease dataset), and explore targets of varying complexity, from structured MNIST digits to spatially disjoint human face parts. These cases require strategic exploration, highlighting DiffATD's adaptability. To assess scalability, we compare DiffATD and the baseline approaches with a larger search space size of 256×256, and present the result in Table 11. Our findings reinforce DiffATD's effectiveness in complex tasks.

Table 11: Results With Larger Search Space (256×256) using DOTA Dataset

Active Discovery of Objects, e.g. Plane, Truck, etc.			
Method	$\mathcal{B} = 100$	$\mathcal{B} = 150$	$\mathcal{B} = 200$
RS	0.1990	0.2487	0.2919
Max-Ent	0.3909	0.4666	0.5759
GA	0.4780	0.5632	0.6070
DiffATD	**0.5251**	**0.6106**	**0.7576**

H Additional Results to Access the Effect of $\kappa(\beta)$

In the main paper, we analyze how the exploration-exploitation tradeoff function impacts search performance, using both the DOTA and Skin imaging datasets (see Table 12 for the result). Additionally, in this section, we have included sensitivity analysis results on the exploration-exploitation tradeoff function for the CelebA and datasets in the following Table 12. The observed trends are consistent with those reported in Table 5 of the main paper. These additional findings further strengthen our hypothesis regarding the role of the exploration-exploitation tradeoff function.

Table 12: *DiffATD's* Performance Across Varying α with $\mathcal{B} = 200$

Active Discovery of Handwritten Digits			
Dataset	$\alpha = 0.2$	$\alpha = 1.0$	$\alpha = 5.0$
CelebA	0.4193	**0.4565**	0.4258
MNIST	0.9227	**0.9682**	0.9459

I Rationale Behind Uniform Measurement Schedule Over the Number of Reverse Diffusion Steps

We adopt a uniform measurement schedule across all denoising steps, and this design choice is deliberate. A non-uniform schedule — with fewer measurements at higher noise levels and more frequent measurements at lower noise levels — might seem intuitive, but is less effective. By the time the noise level is low, the image structure is largely formed, and overly frequent measurements at that

stage do not provide additional value, as they limit the diffusion model's ability to adapt meaningfully based on the measurements. Instead, a uniform schedule ensures that measurements are distributed evenly, giving the diffusion model adequate time to adapt and refine its predictions based on each measurement. This balance ultimately leads to more stable and effective reconstructions across diverse settings. On the other hand, a non-uniform schedule — with more frequent measurements at higher noise levels and fewer at lower noise levels — can be counterproductive. At higher noise levels, the reconstruction is predominantly governed by random noise, making frequent measurements less informative and potentially wasteful.

J Performance of DiffATD Under Noisy Observations

We evaluate the robustness of DiffATD in the presence of noisy observations by introducing varying levels of noise into the observation space. As shown in Tables 13 and 14, DiffATD consistently maintains strong performance across noise levels across different settings, demonstrating its resilience and reliability under imperfect data conditions. To further understand DiffATD's robustness, we visualize posterior reconstructions of the search space from sparse, partially observable, and noisy inputs (Figure 22). Despite the noise, DiffATD effectively reconstructs a clean representation of the underlying space, explaining its consistent performance across varying noise levels in the active target discovery task.

Table 13: DiffATD's Performance with MNIST dataset Under Noisy Observation.

Active Discovery of Handwritten Digits			
Noise Level	$\mathcal{B} = 100$	$\mathcal{B} = 150$	$\mathcal{B} = 200$
μ=0 and σ = 20	0.7318	0.8420	0.9674
μ=0 and σ = 30	0.7307	0.8401	0.9641
μ=10 and σ = 30	0.7298	0.8393	0.9624
No Noise	**0.7324**	**0.8447**	**0.9682**

Table 14: DiffATD's Performance with DOTA dataset Under Noisy Observation.

Active Discovery of objects, e.g, plane, truck, etc.			
White Noise Level	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$
μ=0 and σ = 20	0.5410	0.6150	0.7297
μ=0 and σ = 30	0.5139	0.6335	0.7384
μ=10 and σ = 30	0.5292	0.6319	0.7269
No Noise	**0.5422**	**0.6379**	**0.7309**

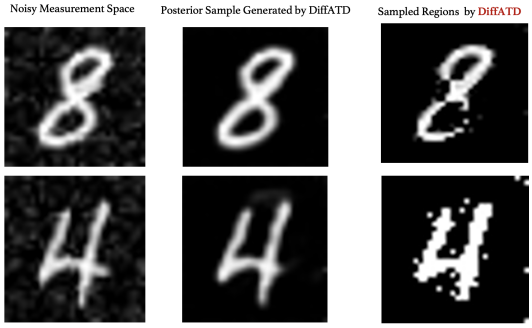

Figure 22: Visualizations of Reconstructed Search Space from Noisy Observations.

K Species Distribution Modelling as Active Target Discovery Problem

We constructed our species distribution experiment using observation data of lady beetle species from iNaturalist. Center points were randomly sampled within North America (latitude 25.6°N to 55.0°N, longitude 123.1°W to 75.0°W). Around each center, we defined a square region approximately 480 km × 480 km in size (roughly 5 degrees in both latitude and longitude). Each retained region was discretized into a 64×64 grid, where the value of each cell represents the number of observed lady beetles. To simulate the querying process, each 2×2 block of grid cells was treated as a query. We evaluated our method on real Coccinella Septempunctata observation data without subsampling. Our goal is to find as many Coccinella Septempunctata as possible within a region.

L Statistical Significance Results of DiffATD

In order to strengthen our claim on DiffATD's superiority over the baseline methods, we have included the statistical significance results with the DOTA and CelebA datasets, and present the results in Tables 15, 16. These results are based on 5 independent trials and further strengthen our empirical findings, reinforcing the effectiveness of DiffATD across diverse domains.

Table 15: Statistical Significance Results with DOTA Dataset

Active Discovery of Objects like Plane, Truck, etc.			
Method	$\mathcal{B} = 100$	$\mathcal{B} = 150$	$\mathcal{B} = 200$
RS	0.1990 ± 0.0046	0.2487 ± 0.0032	0.2919 ± 0.0036
Max-Ent	0.4625 ± 0.0150	0.5524 ± 0.0131	0.6091 ± 0.0188
GA	0.4586 ± 0.0167	0.5961 ± 0.0119	0.6550 ± 0.0158
DiffATD	$\mathbf{0.5422 \pm 0.0141}$	$\mathbf{0.6379 \pm 0.0115}$	$\mathbf{0.7309 \pm 0.0107}$

Table 16: Statistical Significance Results with CelebA Dataset

Active Discovery of Different Parts of Human Faces.			
Method	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$
RS	0.1938 ± 0.0017	0.2441 ± 0.0021	0.2953 ± 0.0002
Max-Ent	0.2399 ± 0.0044	0.3510 ± 0.0060	0.4498 ± 0.0118
GA	0.2839 ± 0.0106	0.3516 ± 0.0110	0.4294 ± 0.0121
DiffATD	$\mathbf{0.4565 \pm 0.0089}$	$\mathbf{0.5646 \pm 0.0086}$	$\mathbf{0.6414 \pm 0.0081}$

M Comparison With Multi-Armed Bandit Based Method

We compare the performance with MAB-based methods across varying measurement budgets, with results summarized in the Table below 17, 18. These analyses are conducted in the context of active target discovery on the MNIST digits and DOTA overhead images, respectively. **These MAB-based methods struggle in the active target discovery setting due to their lack of prior knowledge and structured guidance compared to the Diffusion model, resulting in significantly weaker performance than DiffATD.**

Table 17: *SR* comparison with MNIST Dataset

Active Discovery of Handwritten Digits			
Method	$\mathcal{B} = 100$	$\mathcal{B} = 150$	$\mathcal{B} = 200$
UCB	0.0026	0.0194	0.0923
ϵ-greedy	0.0151	0.0394	0.0943
DiffATD	**0.7324**	**0.8447**	**0.9682**

Table 18: *SR* comparison with DOTA Dataset

Active Discovery of Objects in Overhead Images			
Method	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$
UCB	0.1132	0.2487	0.2146
DiffATD	**0.5422**	**0.6379**	**0.7309**

N Segmentation and Active Target Discovery are Fundamentally Different Tasks

Segmentation and Active Target Discovery (ATD) serve fundamentally different purposes. Segmentation techniques operate under the assumption that the full or partial image is already available, focusing on labeling observed regions. In contrast, ATD is centered around reasoning about unobserved regions using only sparse, partial observations. This distinction is crucial—when only a few pixels are known, as in our setting (An illustrative example of this scenario is shown in Figure 23), segmentation models offer little value, since the primary challenge lies not in interpreting the visible content, but in strategically exploring and inferring the hidden parts of the search space to maximize the target object discovery.

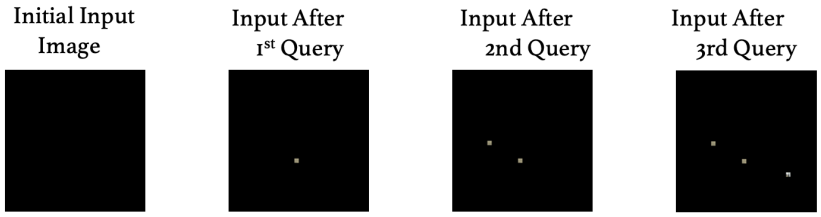

Figure 23: Visualization of Search Space at initial Active Target Discovery Phase.

O More Details on Computational Cost across Search Space

We have conducted a detailed evaluation of sampling time and computational requirements of *DiffATD* across various search space sizes. We present the results in 19. Our results show that *DiffATD* remains efficient even as the search space scales, with sampling time per observation ranging from 0.41 to 3.26 seconds, which is well within practical limits for most downstream applications. This further reinforces DiffATD's scalability and real-world applicability.

Table 19: Details of Computation and Sampling Cost Across Varying Search Space Sizes

Active Discovery of Handwritten Digits		
Search Space	Computation Cost	Sampling Time (Seconds)
28×28	726.9 MB	0.41
128×128	1.35 GB	1.48
256×256	3.02 GB	3.26

P More Visualizations on *DiffATD*'s Exploration vs Exploitation strategy at Different Stage

In this section, we provide additional visualizations that illustrate how *DiffATD* balances exploration and exploitation at various stages of the active discovery process. These visualizations are shown in figures 24, 25, 26, 27, 28. In each example, we show the exploration score ($\text{expl}^{\text{score}}()$), likelihood score ($\text{likeli}^{\text{score}}()$), and the confidence score of the reward model (r_ϕ) across measurement space

at two distinct stages of the active discovery process. The top row represents the initial phase (measurement step 10), while the bottom row corresponds to a near-final stage (measurement step 490). We observe a similar trend across all examples, i.e., *DiffATD* prioritizes the exploration score when selecting measurement locations during the early phase of active discovery. Thus, *DiffATD* aims to maximize information gain during the initial search stage. As the search approaches its final stages, the rankings of the measurement locations shift to being primarily driven by the exploitation score, as defined in Equation 7. As the search progresses, the confidence score of r_ϕ becomes more accurate, making the $\mathrm{expl}^{\mathrm{score}}()$ more reliable. This explains *DiffATD*'s growing reliance on the exploitation score in the later phases of the process. These additional visualizations illustrate *DiffATD*'s exploration strategy, particularly how it dynamically balances exploration and exploitation. These visualizations also underscore the effectiveness of *DiffATD* in addressing active target discovery within partially observable environments.

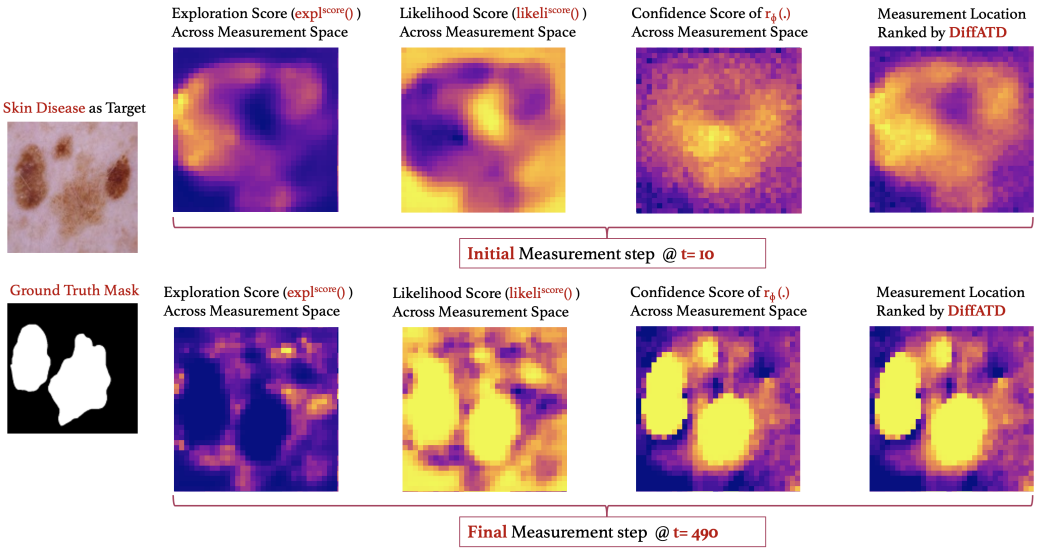

Figure 24: Explanation of *DiffATD*. Brighter indicates a higher value and higher rank.

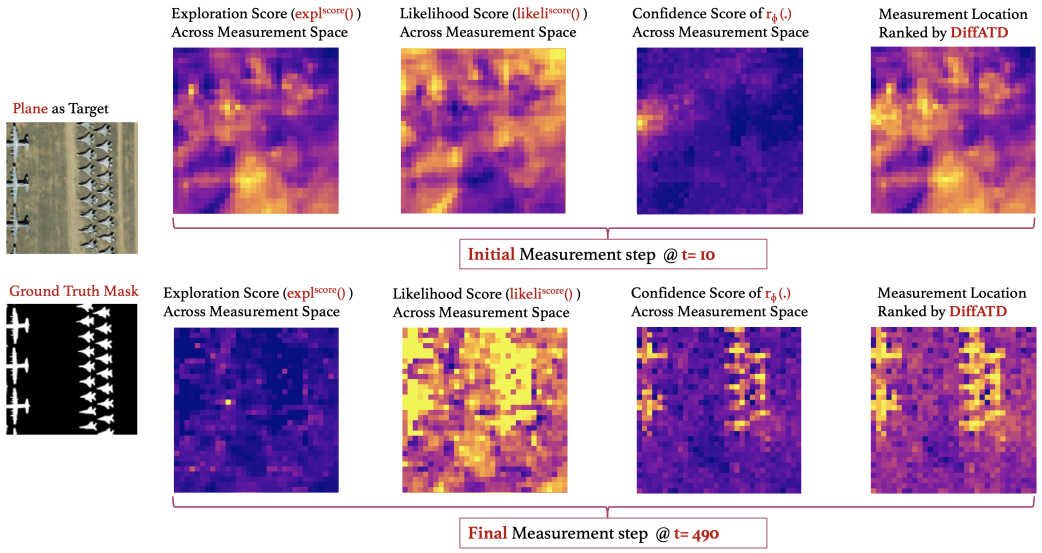

Figure 25: Explanation of *DiffATD*. Brighter indicates a higher value and higher rank.

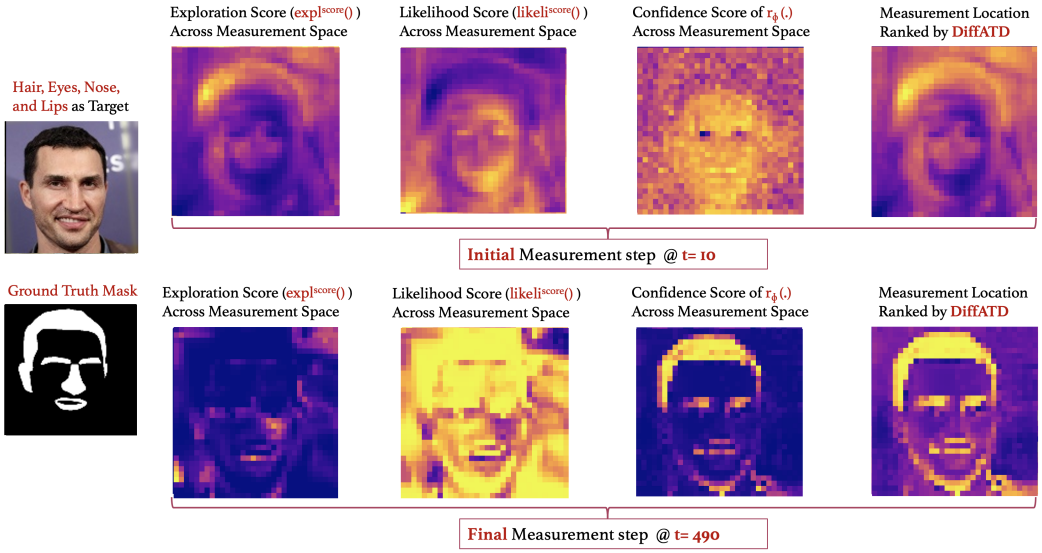

Figure 26: Explanation of *DiffATD*. Brighter indicates a higher value, and higher rank.

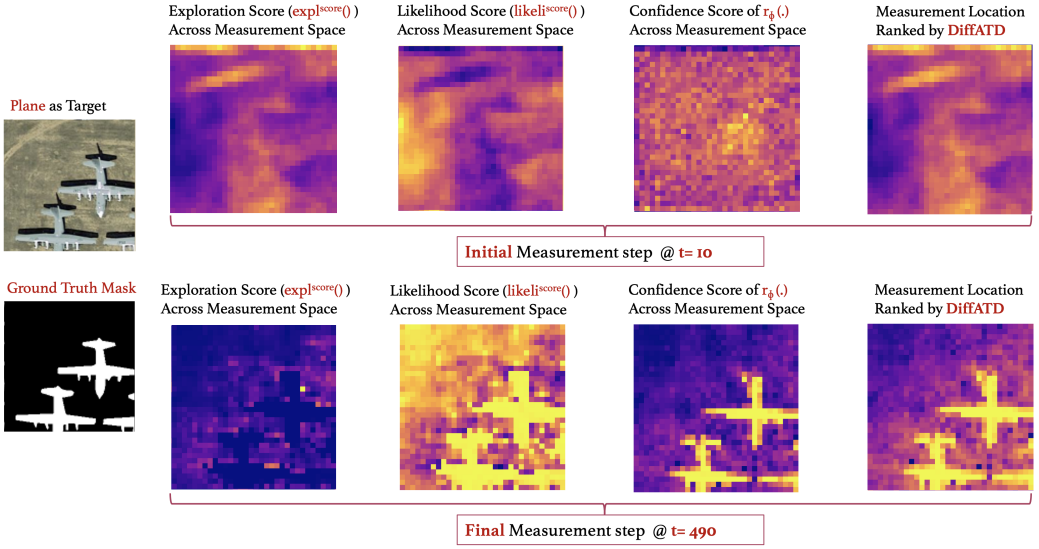

Figure 27: Explanation of *DiffATD*. Brighter indicates a higher value, and higher rank.

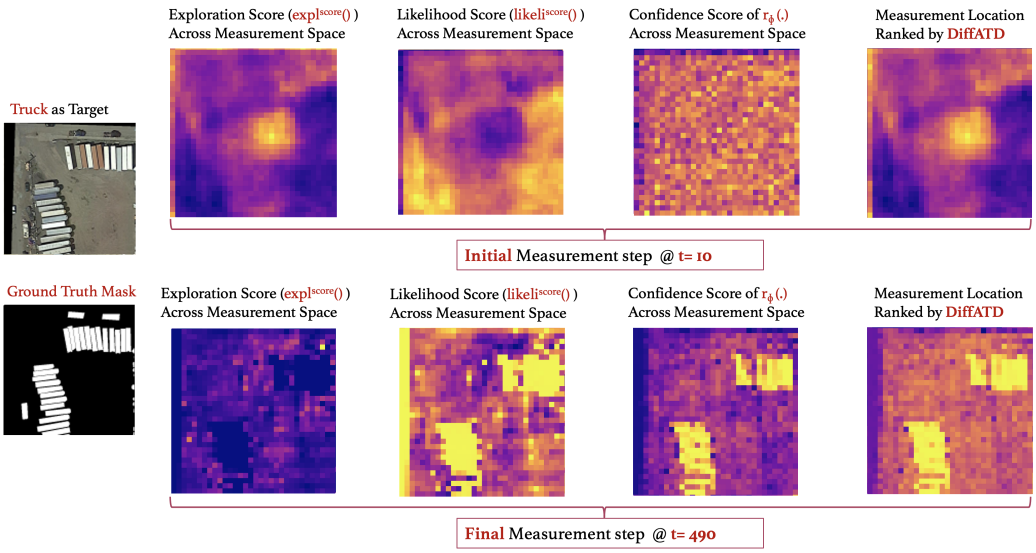

Figure 28: Explanation of *DiffATD*. Brighter indicates a higher value, and higher rank.

Q Importance of Utilizing a Strong Prior in Tackling Active Target Discovery

The purpose of the prior belief model (e.g., a pretrained Diffusion model) in DiffATD is to construct a dynamically updated, uncertainty-aware belief distribution over unobserved regions—critical for guiding the exploration-exploitation tradeoff in target discovery. Thus, it is important to utilize a belief model that is capable of understanding the hidden structure of the underlying search space. In order to validate this, we conduct an experiment with a different choice of the prior belief model that is trained on a different domain (i.e., ImageNet), and compare its performance with DiffATD where the prior belief model is trained on the same domain (i.e., DOTA). We present the result in the following Table 20. Our empirical observation further reinforces the fact that leveraging a strong prior belief model is crucial for efficiently tackling active target discovery.

Table 20: *SR* comparison with DOTA Dataset

Active Discovery of Objects from Overhead Imagery		
Method	$\mathcal{B} = 250$	$\mathcal{B} = 300$
DiffATD with a weak prior Model	0.5143	0.6391
DiffATD	**0.6379**	**0.7309**

R Notation Table

To prevent ambiguity, we include below a compact notation table summarizing all symbols and variables used in describing the DiffATD methodology.

Table 21: Notation summary for the DiffATD framework.

Notation	Description
$\hat{x}_t$	Predicted or estimated search space x at the t-th observation step.
$\tilde{x}_t$	Set of previously observed partial glimpses of the search space up to step t.
$\hat{x}_t^{(i)}$	i-th element in the batch of particles (superscript (i) denotes particle index).
q_t	Measurement location at the t-th time step.
Q_t	Set of measurement locations up to time step t.
τ	Time step index in the reverse diffusion process.

S Limitations and Future Work

While DiffATD demonstrates strong efficiency in addressing active target discovery tasks across diverse settings, it currently depends on the availability of a robust domain-specific prior model to perform effectively. This requirement can be restrictive in real-world applications such as rare species identification or rare disease discovery, where data scarcity makes it difficult to establish a reliable prior. Consequently, the method's success is inherently linked to the quality and expressiveness of the prior model. In future work, we aim to extend DiffATD toward settings with weak or no prior knowledge, enabling active target discovery under an uninformative prior.