

410 **Appendix: Online Feedback Efficient Active Target Discovery in Partially** 411 **Observable Environments**

412 **A Proof of Proposition 5.1**

Proof.

$$\begin{aligned}
 q_t^{\text{exp}} &= \arg \max_{q_t} [I(\hat{x}_t; x \mid Q_t, \tilde{x}_{t-1})] \\
 &= \arg \max_{q_t} [\mathbb{E}_{p(\hat{x}_t|x, Q_t)p(\tilde{x}_{t-1})} [\log p(\hat{x}_t \mid Q_t, \tilde{x}_{t-1}) - \log p(\hat{x}_t \mid x, Q_t, \tilde{x}_{t-1})]] \\
 &= \arg \max_{q_t} \left[H(\hat{x}_t \mid Q_t, \tilde{x}_{t-1}) - \underbrace{H(\hat{x}_t \mid x, Q_t, \tilde{x}_{t-1})}_{\text{remains independent of the } q_t. \text{ Therefore, this term can be disregarded when optimizing for } q_t} \right] \\
 &= \arg \max_{q_t} [H(\hat{x}_t \mid Q_t, \tilde{x}_{t-1})]
 \end{aligned}$$

416

□

417 **B Proof of Theorem 5.2**

418 *Proof.* We demonstrate that maximizing the marginal entropy of the belief distribution as defined
419 in Equation 5 does not require computing a separate set of particles for every possible choice of
420 measurement location q_t when the action corresponds to selecting a measurement location. Since
421 the measurement locations $Q_t = Q_{t-1} \cup q_t$ only differ in the newly selected indices q_t within the
422 $\arg \max$, the elements of each particle $x_t^{(i)}$ remain unchanged across all possible Q_t , except at the
423 indices specified by q_t . Exploiting this structure, we decompose the squared L_2 norm into two parts:
424 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
425 constant in the $\arg \max$ and can be disregarded. Consequently, the formulation reduces the computing
426 of the squared L_2 norms exclusively for the elements corresponding to q_t . We denote k as the set of
427 possible measurement locations at step t . Utilizing Equation 5, we can write,

$$\begin{aligned}
 q_t^{\text{exp}} &= \arg \max_{q_t} \sum_{i=0}^{N_B} \alpha_i \log \sum_{j=0}^{N_B} \alpha_j \exp \left\{ \frac{\|\hat{x}_t^{(i)} - \hat{x}_t^{(j)}\|_2^2}{2\sigma_x^2} \right\} \\
 q_t^{\text{exp}} &= \arg \max_{q_t} \sum_{i,j} \log \left(\exp \left(\frac{\|\hat{x}_t^{(i)} - \hat{x}_t^{(j)}\|_2^2}{2\sigma_x^2} \right) \right) \text{ (By assuming, } \alpha_i = \alpha_j, \forall i, j) \\
 q_t^{\text{exp}} &= \arg \max_{q_t} \sum_{i,j} \log \left(\exp \left(\frac{\sum_{a \in Q_t} ([\hat{x}_t^{(i)}]_a - [\hat{x}_t^{(j)}]_a)^2}{2\sigma_x^2} \right) \right) \\
 q_t^{\text{exp}} &= \arg \max_{q_t} \sum_{i,j} \log \left(\exp \left(\frac{\sum_{q_t \in k} ([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2 + \sum_{r \in Q_{t-1}} ([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right) \right) \\
 q_t^{\text{exp}} &= \arg \max_{q_t} \sum_{i,j} \log \left(\prod_{q_t \in k} \exp \left(\frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} \right) \prod_{r \in Q_{t-1}} \exp \left(\frac{([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2} \right) \right) \\
 q_t^{\text{exp}} &\propto \arg \max_{q_t} \sum_{i,j} \left(\sum_{q_t \in k} \frac{([\hat{x}_t^{(i)}]_{q_t} - [\hat{x}_t^{(j)}]_{q_t})^2}{2\sigma_x^2} + \underbrace{\sum_{r \in Q_{t-1}} \frac{([\hat{x}_t^{(i)}]_r - [\hat{x}_t^{(j)}]_r)^2}{2\sigma_x^2}}_{\text{We can ignore as it doesn't depend on the choice of } q_t.} \right)
 \end{aligned}$$

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

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

428

□

429 C Proof of Theorem 5.4

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

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

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

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

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

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

433 We obtain the last equality using the following identity:

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

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

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

435 By definition, it is equivalent to the following

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

436 D Proof of Theorem 5.3

437 *Proof.* We start with the definition of entropy H :

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

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

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

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:

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

Equivalently, we can write the above expression as:

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

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

E 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 [33] and report the findings in Table 6. 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.8, with additional visualizations are in appendix I.

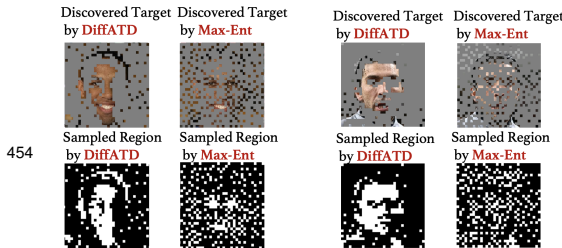


Figure 8: Active Discovery of regions with Face.

Table 6: 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

F Active Discovery of Bone Suppression

Next, we evaluate *DiffATD* on the Chest X-Ray dataset [34], with results shown in Table 7. 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. 9, with additional examples in the appendix I.

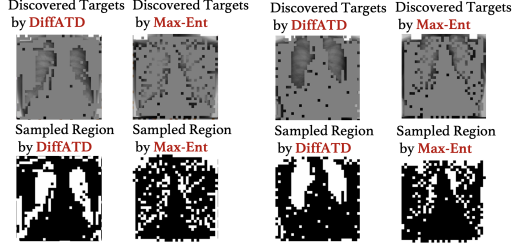


Figure 9: Active Discovery of Bone Suppression.

Table 7: 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
<i>DiffATD</i>	0.3065	0.3733	0.4142

G Qualitative Comparison with Greedy Adaptive Approach

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 10,11,12, 13. 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, misleading 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.

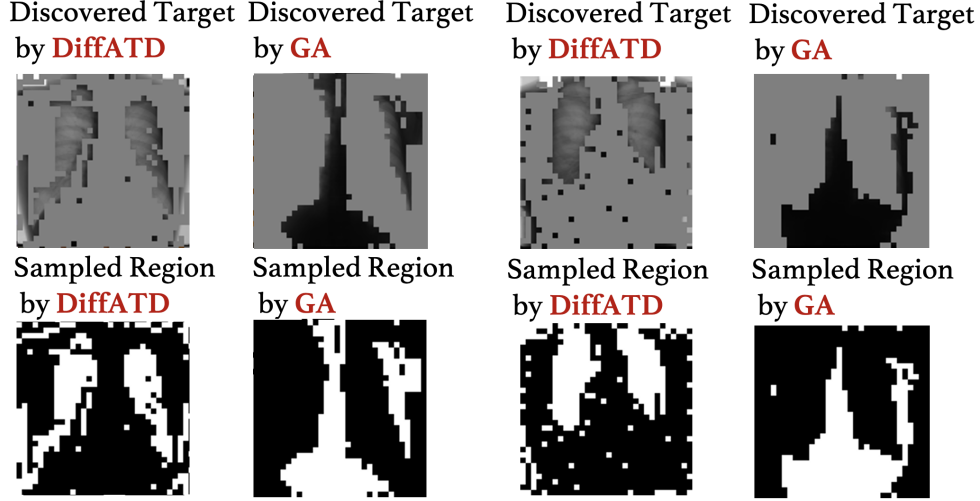


Figure 10: Visualization of Active Discovery of Lung Disease.

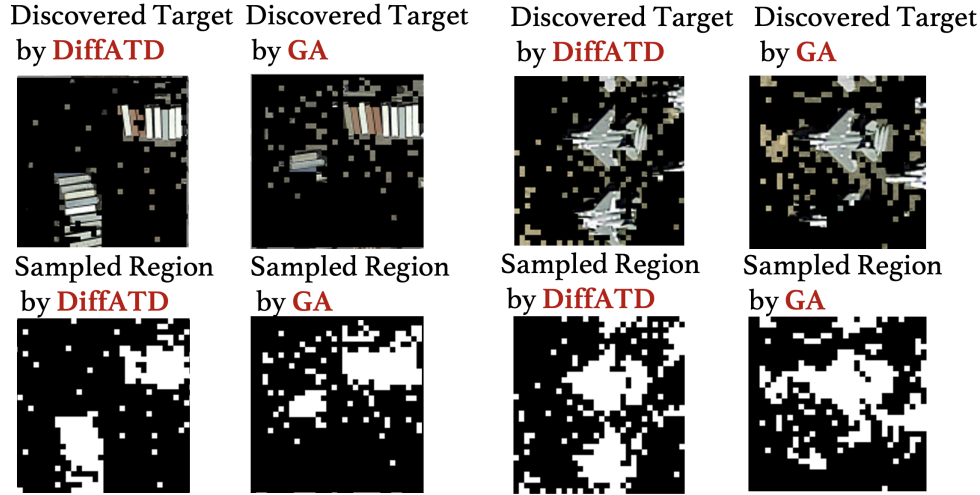


Figure 11: Visualization of Active Discovery of (left) Truck and (right) Plane

H 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 8. 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. 14, *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.

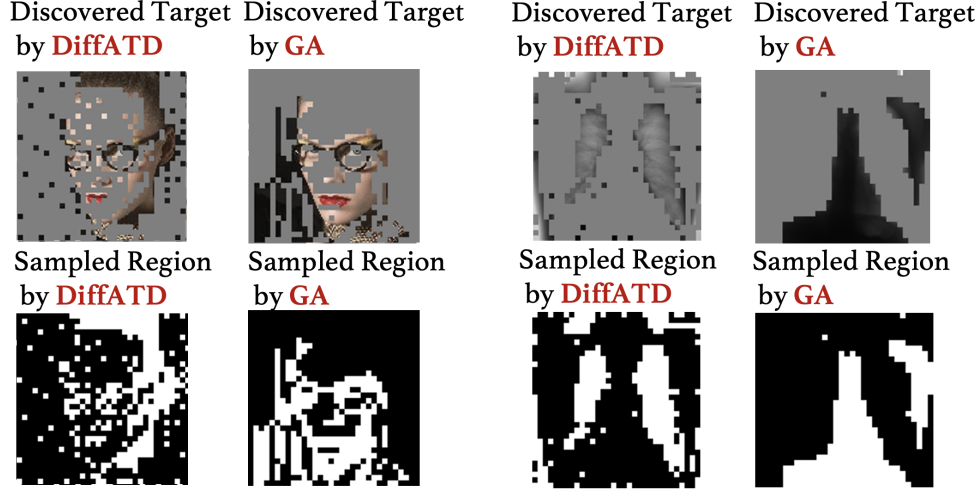


Figure 12: Visualization of Active Discovery of Eye, hair, nose, and lip.

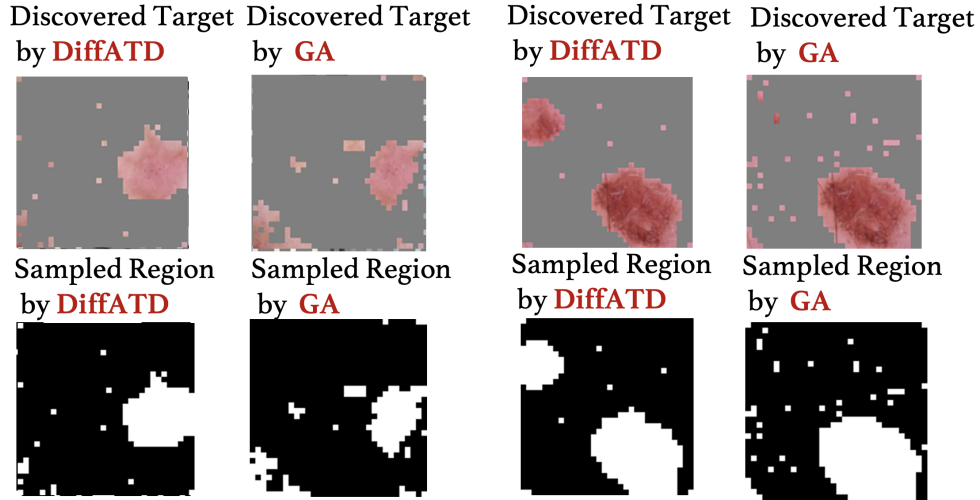


Figure 13: Visualization of Active Discovery of Skin Disease.

Table 8: *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
<i>DiffATD</i>	0.7324	0.8447	0.9682

I More Visualizations of the Exploration Strategy of *DiffATD*

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 15, 16, 17, 18, 19, 20. 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

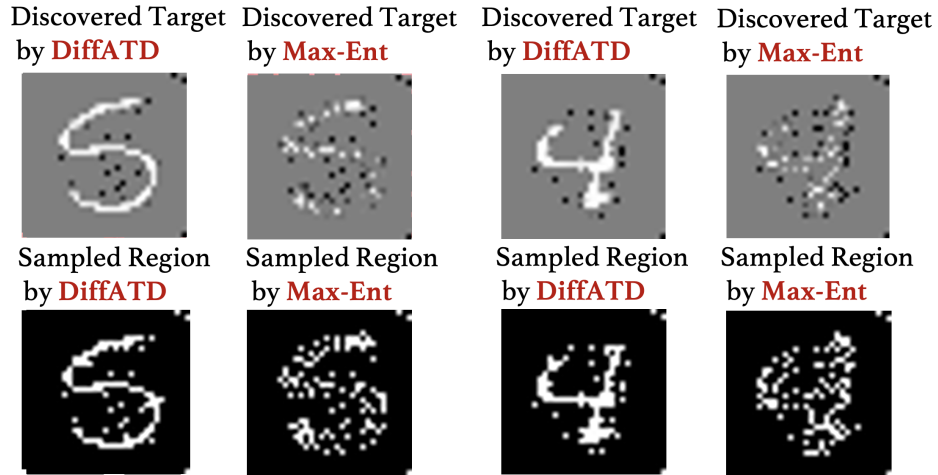


Figure 14: Visualization of Active Discovery of Handwritten Digits.

499 solely on maximizing information gain. These visualizations further reinforce the effectiveness of *DiffATD* in active target discovery within partially observable environments.

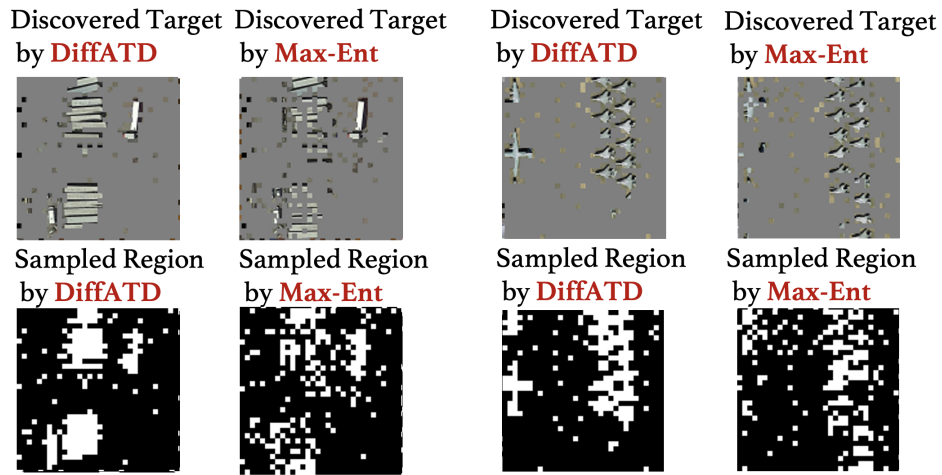


Figure 15: Visualizations of Active Discovery of (left) *large-vehicle*, and (right) *Plane*.

500

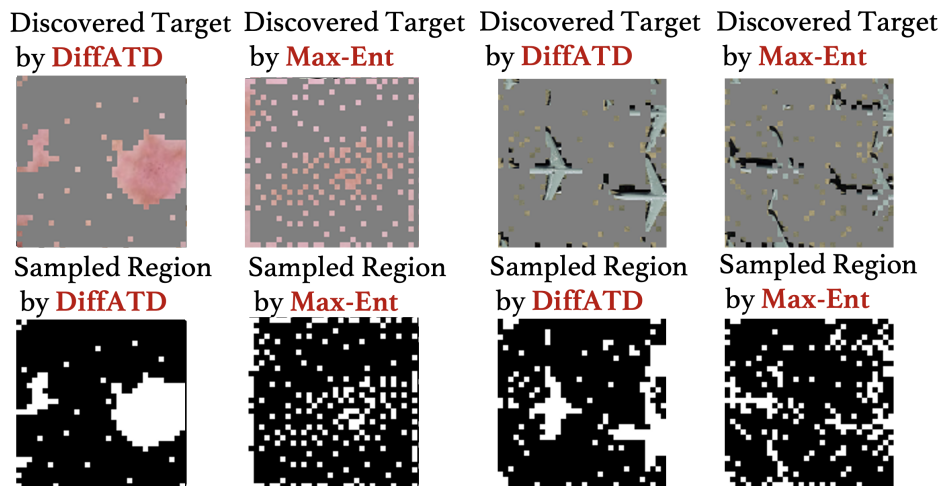


Figure 16: Visualizations of Active Discovery of (left) *Skin disease*, and (right) *Plane*.

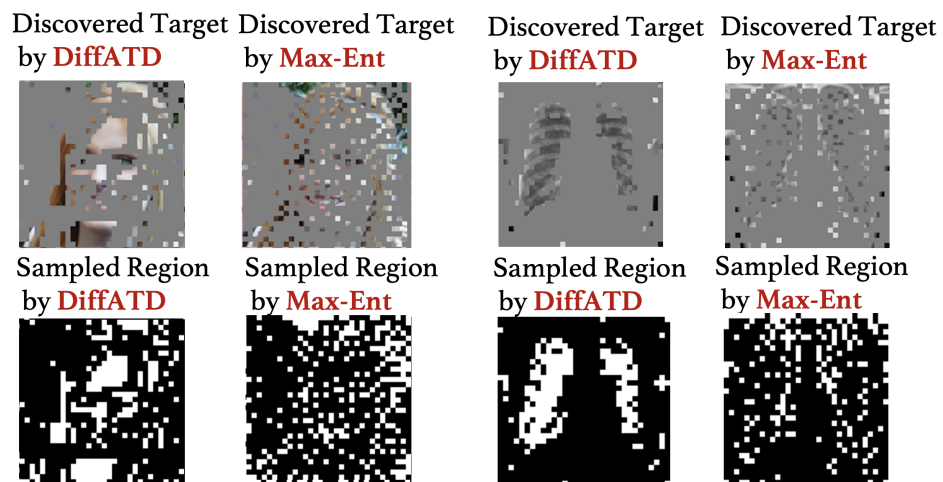


Figure 17: Visualizations of Active Discovery of (left) *human face*, and (right) *lung disease, such as TB*.

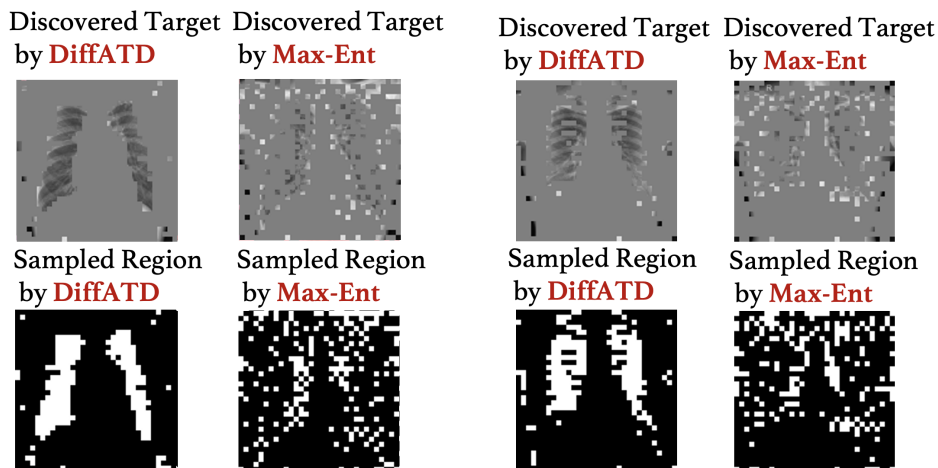


Figure 18: Visualization of Active Discovery of Lung Disease.

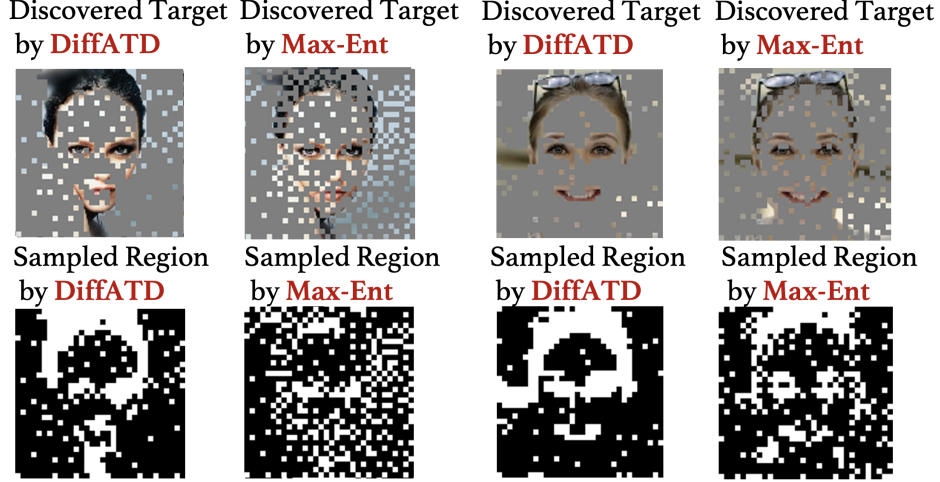


Figure 19: Visualization of Active Discovery of hair, eye, nose, and lip.

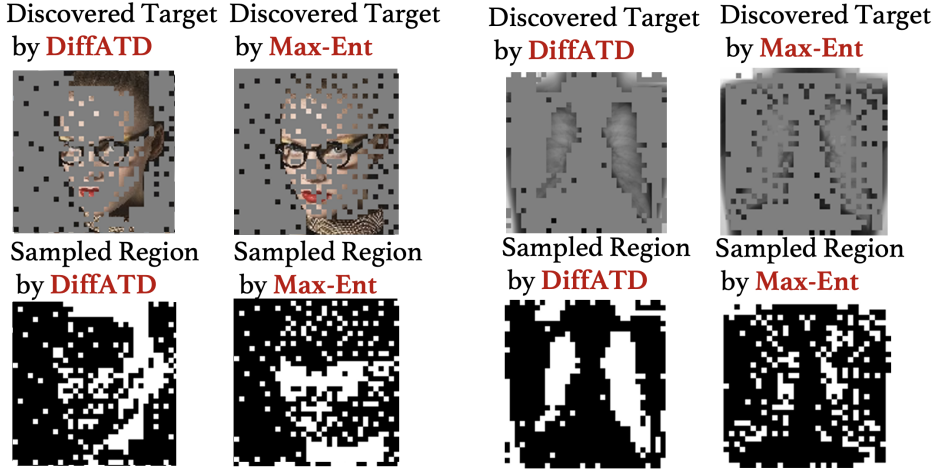


Figure 20: Visualization of Active Discovery of Lung Disease (right) and hair, eye, nose, and lip (left).

501 J Details of Computing Resources, Training, and Inference Hyperparameters

502 This section provides the training and inference hyperparameters for each dataset used in our
503 experiments. We use DDIM [35] as the diffusion model across datasets. The diffusion models
504 used in different experiments are based on widely adopted U-Net-style architecture. For the MNIST
505 dataset, we use 32-dimensional diffusion time-step embeddings, with the diffusion model consisting
506 of 2 residual blocks. We select the time-step embedding vector dimension to match the input feature
507 size, ensuring the diffusion model can process it efficiently. The block widths are set to [32, 64, 128],
508 and training involves 30 diffusion steps. DOTA, CelebA, and Skin imaging datasets share the same
509 input feature size of [128, 128, 3] and architecture, featuring 128-dimensional time-step embeddings
510 and a diffusion model with 2 residual blocks of width [64, 128, 256, 256, 512]. For these datasets,
511 we perform training with 100 diffusion steps. We use 128-dimensional time-step embeddings for
512 the Bone dataset and a diffusion model with 2 residual blocks (each block width: [64, 128, 256,
513 256]). We use 100 diffusion steps during training. We set the learning rate and weight decay factor to
514 $1e^{-4}$ for all experimental settings. We set a measurement schedule (M) of 100 for a measurement
515 budget (B) of 200, ensuring that $B \approx \frac{T}{M}$, where T is the total number of reverse diffusion steps,
516 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.

K 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.

L Challenges in Active Target Discovery for Rare Categories

Table 9: Comparison with supervised and fully observable method

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

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 budgets \mathcal{B} , in Table 9. 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*.

M 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 10. Our findings reinforce DiffATD’s effectiveness in complex tasks.

Table 10: 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

N Additional Results to Assess 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 11 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 11. 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 11: 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

O 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.

P 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 12 and 13, 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 21). 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 12: DiffATD’s Performance with MNIST dataset Under Noisy Observation.

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

Table 13: DiffATD’s Performance with DOTA dataset Under Noisy Observation.

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

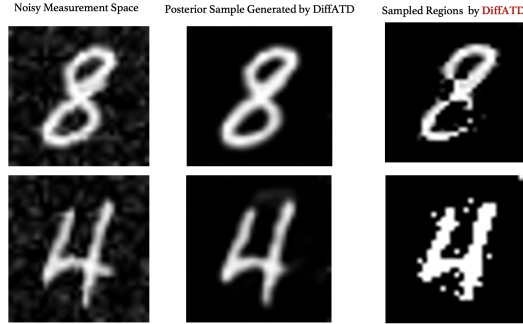


Figure 21: Visualizations of Reconstructed Search Space from Noisy Observations.

593 Q Comparison With Vision-Language Model

594 We conduct comparisons between DiffATD and two State-of-the-art VLMs: GPT-4o and Gemini, on
 595 the DOTA dataset. The corresponding empirical results are summarized in the following Table 14.
 596 We observe that DiffATD consistently outperforms VLM baselines across different measurement
 597 budgets. These findings underscore DiffATD’s advantage over alternatives like VLMs in the ATD
 598 setting, driven by its principled balance of exploration and exploitation based on the maximum
 599 entropy framework. Additionally, in Figure 22, we present comparative exploration strategies for
 600 different targets from DOTA to provide further intuition.

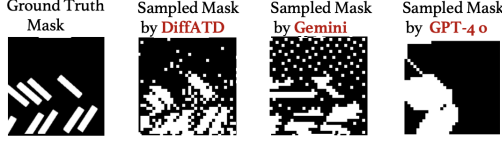


Table 14: Comparison with VLM with DOTA Dataset

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	
<i>DiffATD</i>	0.4537	0.5092	0.7309	

Figure 22: Active Discovery of Plane, Truck.

R 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 \times 480 km in size (roughly 5 degrees in both latitude and longitude). Each retained region was discretized into a 64 \times 64 grid, where the value of each cell represents the number of observed lady beetles. To simulate the querying process, each 2 \times 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.

S 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 \pm 0.0046	0.2487 \pm 0.0032	0.2919 \pm 0.0036
Max-Ent	0.4625 \pm 0.0150	0.5524 \pm 0.0131	0.6091 \pm 0.0188
GA	0.4586 \pm 0.0167	0.5961 \pm 0.0119	0.6550 \pm 0.0158
<i>DiffATD</i>	0.5422 \pm 0.0141	0.6379 \pm 0.0115	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 \pm 0.0017	0.2441 \pm 0.0021	0.2953 \pm 0.0002
Max-Ent	0.2399 \pm 0.0044	0.3510 \pm 0.0060	0.4498 \pm 0.0118
GA	0.2839 \pm 0.0106	0.3516 \pm 0.0110	0.4294 \pm 0.0121
<i>DiffATD</i>	0.4565 \pm 0.0089	0.5646 \pm 0.0086	0.6414 \pm 0.0081

T 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. This analysis is conducted in the context of active target

discovery on the MNIST digits. 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 compared to 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
<i>DiffATD</i>	0.7324	0.8447	0.9682

U 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.

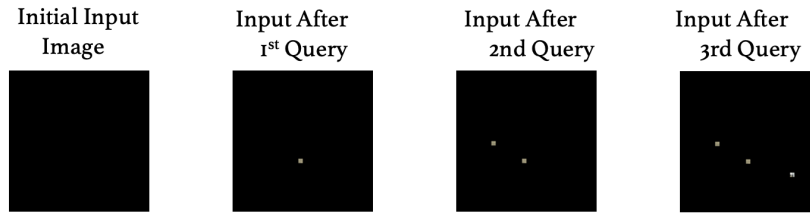


Figure 23: Visualization of Search Space at initial Active Target Discovery Phase.

V 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 18. 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 18: 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

W 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 $\text{expl}^{\text{score}}()$ 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.

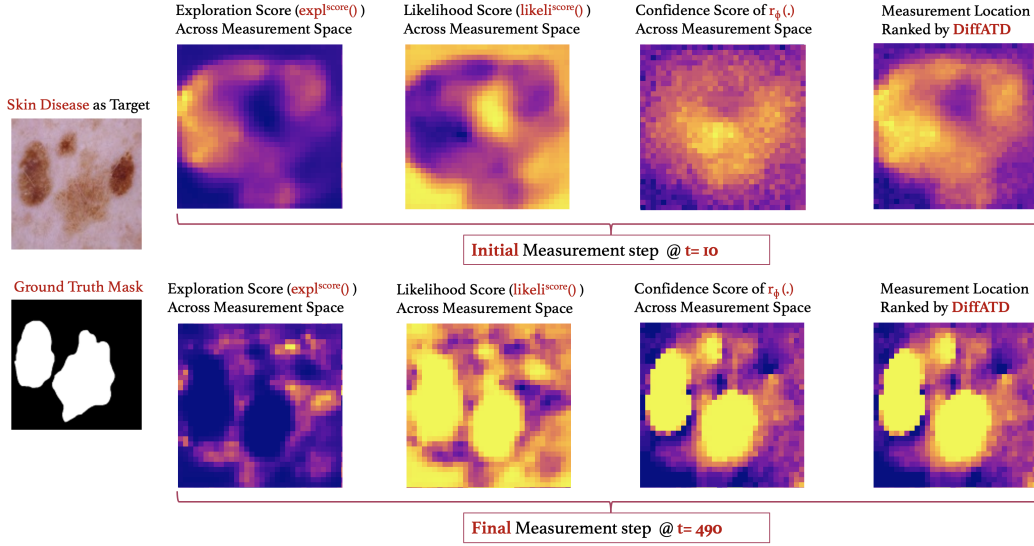


Figure 24: Explanation of *DiffATD*. Brighter indicates a higher value and higher rank.

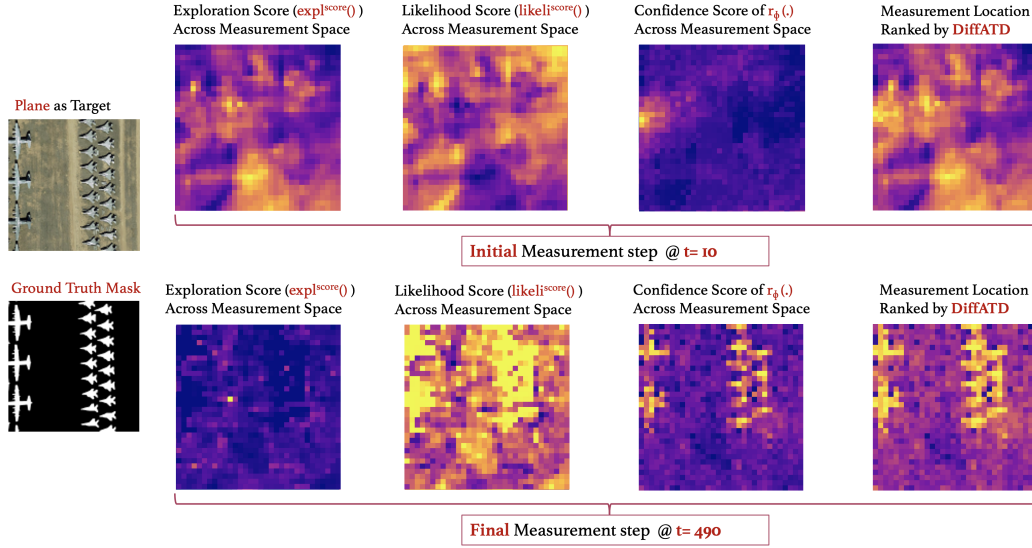


Figure 25: Explanation of *DiffATD*. Brighter indicates a higher value and higher rank.

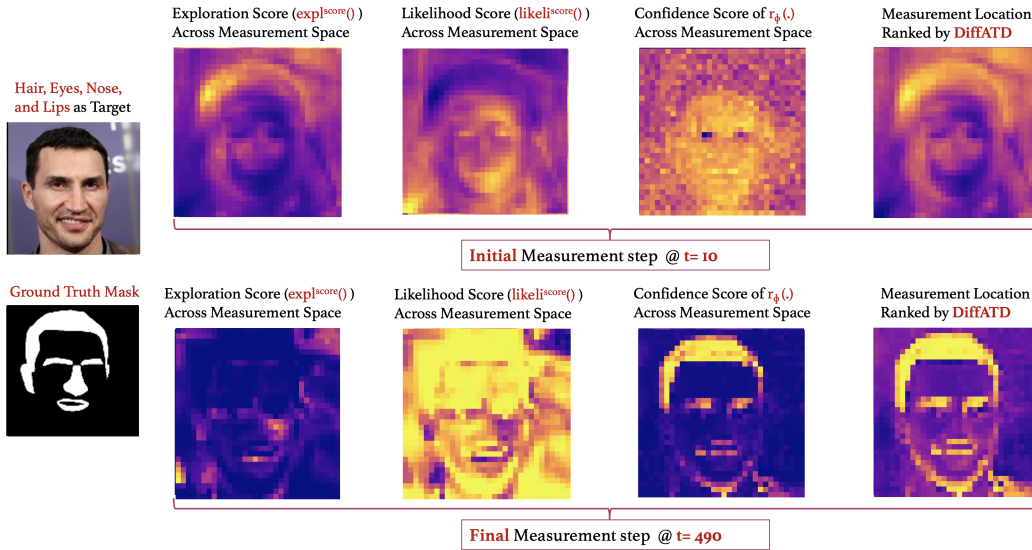


Figure 26: Explanation of *DiffATD*. Brighter indicates a higher value, and higher rank.

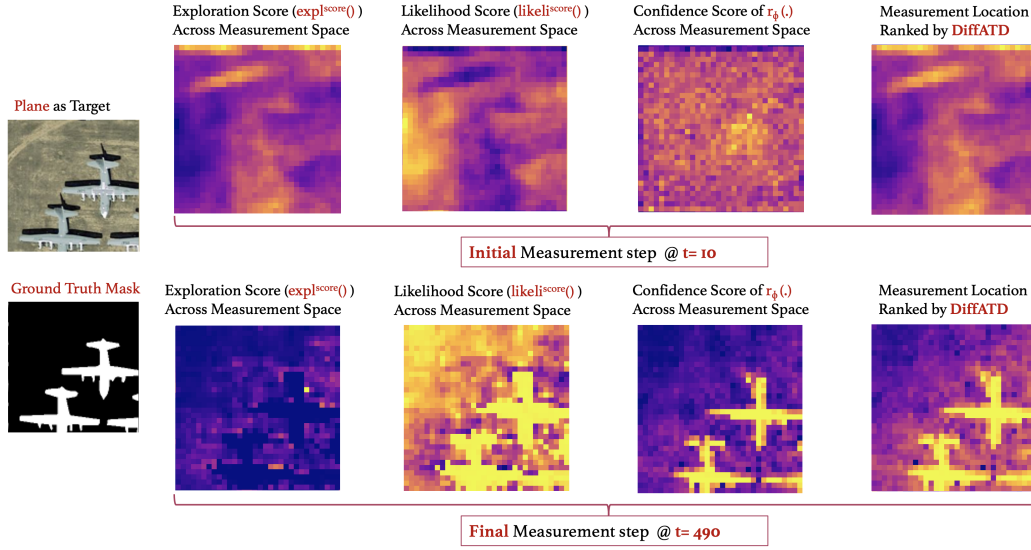


Figure 27: Explanation of *DiffATD*. Brighter indicates a higher value, and higher rank.

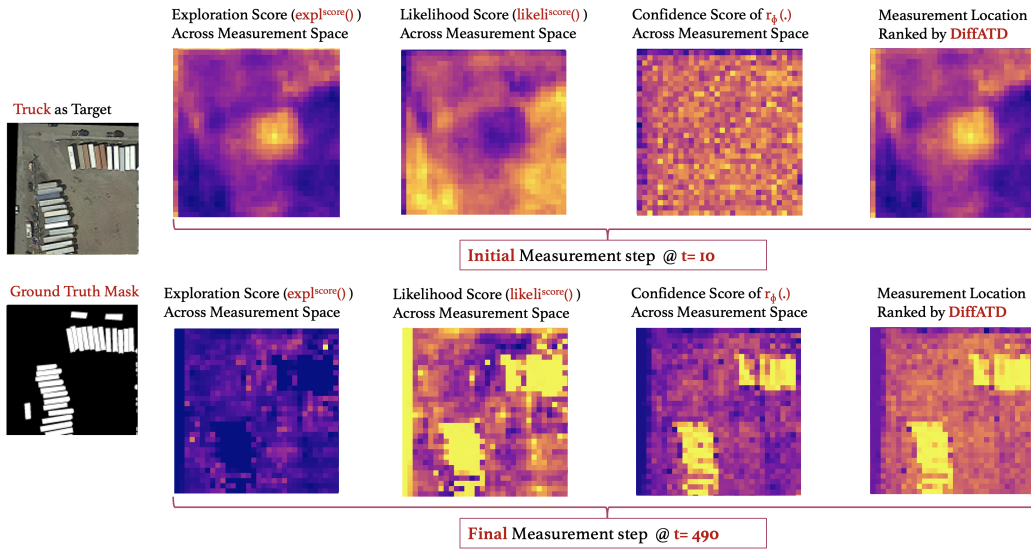


Figure 28: Explanation of *DiffATD*. Brighter indicates a higher value, and higher rank.