

409 **Appendix: Active Target Discovery under Uninformative Prior: The Power of**
 410 **Permanent and Transient Memory**

411 **A Proof of Proposition 1**

Proof.

$$\theta^* = \arg \max_{\theta} \mathbb{E}_{p(y)} [\log q_{\theta}(y)] \quad (12)$$

$$= \arg \min_{\theta} \text{KL}(p(y) \parallel q_{\theta}(y)) \quad (13)$$

412 The core principle of the EM algorithm is that, for any two parameter sets θ_a and θ_b , the following
 413 identity holds:

$$\log \frac{q_{\theta_a}(y)}{q_{\theta_b}(y)} = \log \frac{q_{\theta_a}(x, y)}{q_{\theta_b}(x, y)} \cdot \frac{q_{\theta_b}(x | y)}{q_{\theta_a}(x | y)} \quad (14)$$

$$= \mathbb{E}_{q_{\theta_b}(x|y)} \left[\log \frac{q_{\theta_a}(x, y)}{q_{\theta_b}(x, y)} \right] + \text{KL}(q_{\theta_b}(x | y) \parallel q_{\theta_a}(x | y)) \quad (15)$$

$$\geq \mathbb{E}_{q_{\theta_b}(x|y)} [\log q_{\theta_a}(x, y) - \log q_{\theta_b}(x, y)] \quad (16)$$

414 This inequality remains valid when taking the expectation over $p(y)$. Consequently, starting from an
 415 initial parameter setting θ_0 , the EM update rule can be expressed as:

$$\theta_{k+1} = \arg \max_{\theta} \mathbb{E}_{p(y)} \mathbb{E}_{q_{\theta_k}(x|y)} [\log q_{\theta}(x, y) - \log q_{\theta_k}(x, y)] \quad (17)$$

$$= \arg \max_{\theta} \mathbb{E}_{p(y)} \mathbb{E}_{q_{\theta_k}(x|y)} [\log q_{\theta}(x, y)] \quad (18)$$

416 In the empirical Bayes setting, the forward model $p(y | x)$ is known and only the parameters of the
 417 prior $q_{\theta}(x)$ should be optimized. In this case, Eq. 18 becomes:

$$\theta_{k+1} = \arg \max_{\theta} \mathbb{E}_{p(y)} \mathbb{E}_{q_{\theta_k}(x|y)} [\log q_{\theta}(x) + \log p(y | x)] \quad (19)$$

$$\theta_{k+1} = \arg \max_{\theta} \mathbb{E}_{p(y)} \mathbb{E}_{q_{\theta_k}(x|y)} [\log q_{\theta}(x)] \quad (20)$$

418 □

419 **B Proof of Lemma 1**

420 *Proof.* We can express the conditional score as follows:

$$\nabla_{x_t} \ln p_t(x | y) = \int \nabla_{x_t} \ln \vec{p}_{t|0}(x_t | x_0) \overleftarrow{p}_{0|t}(x_0 | x_t, y) dx_0 \quad (21)$$

421 The minimizer is obtained by exploiting the property that the conditional expectation yields the
 422 optimal solution under the mean squared error criterion:

$$\mathbb{E} [\nabla_{H_t} \ln p_{t|0}(H_t | X_0) - \nabla_{H_t} \ln p_t(H_t) | Y = y, H_t = x] \quad (22)$$

$$h_t^*(x, y) = \left(\int [\nabla_x \ln p_{t|0}(x | x_0) - \nabla_x \ln p_t(x)] p_{0|t}(x_0 | H_t = x, Y = y) dx_0 \right) \quad (23)$$

$$= \int \nabla_x \ln p_{t|0}(x | x_0) p_{0|t}(x_0 | H_t = x, Y = y) dx_0 - \nabla_x \ln p_t(x) \quad (24)$$

$$= \nabla_x \ln p_t(x | y) - \nabla_x \ln p_t(x) \quad (25)$$

$$= \nabla_x \ln p_t(y | x) \quad (26)$$

423

□

424 C Proof of Theorem 1

Proof.

$$\arg \max_{\phi} \mathbb{E}_{p(y)} [\log q_{\phi}(y)] \quad (27)$$

425 Utilizing the result of Theorem 1, we can express the above expression as follows:

$$\phi^{\text{new}} = \arg \max_{\phi} \mathbb{E}_{p(y)} \mathbb{E}_{q_{\phi}(x|y)} [\log q_{\phi}(x)] \quad (28)$$

426 Maximizing the above objective involves computing $\nabla_x \log q_{\phi}(x)$. Following the approach in
427 denoising score matching, the above optimization can be equivalently reformulated as:

$$\zeta^{\text{new}} = \min_{\zeta} \mathbb{E}_{(X_0, Y), \varepsilon, t} \left\| \left(h_t^{\zeta}(x_t, Y) + s_t^{\theta^*}(x_t) \right) - \varepsilon \right\|^2 \quad (29)$$

428 Let $x_t = \sqrt{\bar{\alpha}_t} X_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon$, where $X_0 \sim q_{\phi}(x_0 | y)$, $Y \sim y$ and $\varepsilon \sim \mathcal{N}(0, \mathbf{I})$. Also, $\phi^{\text{new}} =$
429 $(\theta^*, \zeta^{\text{new}})$.

430 Thus, optimizing the objective in Equation 29, is equivalent to one step of EM. Hence, guarantees the
431 improvement of expected log-evidence. As a result, we can write:

$$\mathbb{E}_{p(y)} [\log q_{\phi^{\text{new}}}(y)] \geq \mathbb{E}_{p(y)} [\log q_{\phi}(y)].$$

432

□

433 D Proof of Theorem 2

Proof.

$$\mathbb{E}_{p(y_{t+k})} [\log q_{\phi_{t+k}}(y)] \quad (30)$$

434 Assuming the observations collected during the active discovery process are independent, we can
435 decompose the above expression into two parts as follows:

$$= \underbrace{\mathbb{E}_{p(y_t)} [\log q_{\phi_{t+k}}(y)]}_{\text{part1}} \times \underbrace{\mathbb{E}_{p(y_{t+k:(t+1)})} [\log q_{\phi_{t+k}}(y)]}_{\text{part2}} \quad (31)$$

436 The set of observations gathered from time step $t + 1$ to $t + k$ is denoted as $y_{t+k:(t+1)}$.

437 For *Part1*, we can write,

$$\mathbb{E}_{p(y_t)} [\log q_{\phi_{t+k}}(y)] \geq \mathbb{E}_{p(y_t)} [\log q_{\phi_t}(y)] \quad (\text{Following Theorem 2}) \quad (32)$$

438 For *Part2*, we can write,

$$\mathbb{E}_{p(y_{t+k:(t+1)})} [\log q_{\phi_{t+k}}(y)] \geq \mathbb{E}_{p(y_{t+k:(t+1)})} [\log q_{\phi_t}(y)] \quad (\text{By Definition}) \quad (33)$$

439 The above relation holds because, unlike ϕ_k , the model ϕ_{t+k} is trained on posterior samples explicitly
440 conditioned on the observations $y_{t+k:(t+1)}$. As a result, when ϕ_{t+k} is optimally trained, the left-hand
441 side of Equation 33 reaches its optimal value. In contrast, since ϕ_k is never exposed to $y_{t+k:(t+1)}$
442 during training, it cannot attain the optimal value of the left-hand side in Equation 33.

443 Finally, combining the results of Equation 32 and 33, we can write:

$$\mathbb{E}_{p(y_{t+k})} [\log q_{\phi_{t+k}}(y)] \geq \mathbb{E}_{p(y_{t+k})} [\log q_{\phi_t}(y)]$$

444

□

445 **E Proof of Theorem 3**

446 *Proof.* We start with the definition of q_t^{exp} as follows:

$$q_t^{\text{exp}} = \arg \max_{q_t} -\mathbb{E}_{\hat{x}_t} [\log p(\hat{x}_t | Q_t, y_{t-1})] \quad \text{where} \quad p(\hat{x}_t | Q_t, y_{t-1}) = \sum_{i=0}^P \alpha_i \mathcal{N}(\hat{x}_t^i, \sigma_x^2 I)$$

447 According to [18],

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

448 We can write,

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

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

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

449 We decompose it into two parts: one representing the set of potential measurement locations at the
 450 query step t , and the other corresponding to the set of locations already selected in Q_{t-1} . Hence, we
 451 can express it as follows:

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

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

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

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

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

452

□

453 F Proof of Theorem 4

454 *Proof.* Using Equation 12, we can decompose the score as follows:

$$\begin{aligned} ||\text{Score}_{(\phi^*, \eta^*)}^*(q_t) - \text{Score}_{(\phi, \eta)}(q_t)|| &= \alpha(\mathcal{B}) \underbrace{(\text{expl}_{\phi^*}^{\text{score}}(q_t) - \text{expl}_{\phi}^{\text{score}}(q_t))}_A \\ &\quad + (1 - \alpha(\mathcal{B})) \underbrace{(\text{exploit}_{(\phi^*, \eta^*)}^{\text{score}}(q_t) - \text{exploit}_{(\phi, \eta)}^{\text{score}}(q_t))}_B \end{aligned} \quad (34)$$

455 By following a similar decomposition, we can write,

$$\begin{aligned} ||\text{Score}_{(\phi^*, \eta^*)}^*(q_t) - \text{Score}_{(\phi_{\text{new}}, \tilde{\eta})}(q_t)|| &= \alpha(\mathcal{B}) \underbrace{(\text{expl}_{\phi^*}^{\text{score}}(q_t) - \text{expl}_{\phi_{\text{new}}}^{\text{score}}(q_t))}_C \\ &\quad + (1 - \alpha(\mathcal{B})) \underbrace{(\text{exploit}_{(\phi^*, \eta^*)}^{\text{score}}(q_t) - \text{exploit}_{(\phi_{\text{new}}, \tilde{\eta})}^{\text{score}}(q_t))}_D \end{aligned} \quad (35)$$

456 We can write,

$$\begin{aligned} \underbrace{(\text{expl}_{\phi^*}^{\text{score}}(q_t) - \text{expl}_{\phi}^{\text{score}}(q_t))}_A &\geq \underbrace{(\text{expl}_{\phi^*}^{\text{score}}(q_t) - \text{expl}_{\phi_{\text{new}}}^{\text{score}}(q_t))}_C \\ \text{(Follows from Proposition 1 and Theorem 1)} \end{aligned} \quad (36)$$

457 The above relation holds as ϕ_{new} is obtained by applying one iteration of EM, thus guaranteeing
458 improvement from the previous iteration prior (ϕ). Hence, ensure a more accurate exploration score
459 estimation compared to the prior of previous iteration.

460 Next, we will compare the terms B and D and show that

$$\underbrace{(\text{exploit}_{(\phi^*, \eta^*)}^{\text{score}}(q_t) - \text{exploit}_{(\phi, \eta)}^{\text{score}}(q_t))}_B \geq \underbrace{(\text{exploit}_{(\phi^*, \eta^*)}^{\text{score}}(q_t) - \text{exploit}_{(\phi_{\text{new}}, \tilde{\eta})}^{\text{score}}(q_t))}_D$$

461 Utilizing the expression in 11, we compare the exploit scores computed via different parameterizations
462 of the prior as follows:

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

463 We can rewrite the above expression as follows:

$$\text{exploit}_{(\phi, \eta)}^{\text{score}}(q_t) = \text{likeli}_{\phi}^{\text{score}}(q_t) + \mathbb{E}_{[\hat{x}_t^{(i)}] \sim q_{\phi}(x|y)} [r_{\eta}([\hat{x}_t^{(i)}]_{q_t})]$$

464 Following exactly similar steps, we can also write

$$\text{exploit}_{(\phi_{\text{new}}, \tilde{\eta})}^{\text{score}}(q_t) = \text{likeli}_{\phi_{\text{new}}}^{\text{score}}(q_t) + \mathbb{E}_{[\hat{x}_t^{(i)}] \sim q_{\phi_{\text{new}}}(x|y)} [r_{\tilde{\eta}}([\hat{x}_t^{(i)}]_{q_t})]$$

465 Following the same reasoning as in 37, we can write

$$\begin{aligned} (\text{likeli}_{\phi^*}^{\text{score}}(q_t) - \text{likeli}_{\phi}^{\text{score}}(q_t)) &\geq (\text{likeli}_{\phi^*}^{\text{score}}(q_t) - \text{likeli}_{\phi_{\text{new}}}^{\text{score}}(q_t)) \\ \text{(Follows from Proposition 1 and Theorem 1)} \end{aligned} \quad (37)$$

466 Now, note that at the beginning of the active target discovery process, $r_{\tilde{\eta}}([\hat{x}_t^{(i)}]_{q_0}) = r_{\eta}([\hat{x}_t^{(i)}]_{q_0})$.

467 As ϕ_{new} is a improved parameterization of the previous iteration prior ϕ , thus posterior samples from
468 $q_{\phi_{\text{new}}}(x | y)$ are more reliable and accurate compared to the posterior samples from $q_{\phi}(x | y)$ and
469 thus reward evaluated on the posterior samples from $q_{\phi_{\text{new}}}(x | y)$ are more reliable. Furthermore,
470 the reward model $r_{\tilde{\eta}}$ benefits from training on more diverse samples, as entropy computed using the
471 updated prior ϕ_{new} is more accurate than that from ϕ . This leads to improved data diversity during
472 collection, enabling $r_{\tilde{\eta}}$ to converge faster than r_{η} . The following lemma supports this hypothesis:

Lemma 2 (Diverse Data Improves Convergence). [19, 20] Let θ_t be the parameters of a neural network trained using SGD with batch size 1 and learning rate η on dataset S . Let $\mathcal{L}_S(\theta)$ be the empirical loss. Assume the loss is L -smooth and gradients are bounded by G . Let S_1 and S_2 be two datasets of size n , with $D(S_1) > D(S_2)$. Then, for the same number of iterations T , the expected generalization gap

$$\mathbb{E}[\mathcal{L}(\theta_T^{(1)}) - \mathcal{L}(\theta_T^{(2)})] < 0 \quad \text{where, } D(S) = \frac{1}{n^2} \sum_{i,j} \|x_i - x_j\|^2 \quad \text{for } (x_i, y_i), (x_j, y_j) \in S$$

where $\theta_T^{(1)}$ and $\theta_T^{(2)}$ are trained on S_1 and S_2 respectively, assuming the data distribution \mathcal{D} has high support over \mathcal{X} .

Hence, $\theta_T^{(1)}$ is closer to optimal solution of $\mathcal{L}(\theta)$ than $\theta_T^{(2)}$ in fewer steps, assuming same training budget.

Thus, utilizing the result of the lemma 2, we can write:

$$\begin{aligned} & (\mathbb{E}_{[\hat{x}_t^{(i)}] \sim q_{\phi^*}(x|y)} [r_{\eta^*}([\hat{x}_t^{(i)}]_{q_t})] - \mathbb{E}_{[\hat{x}_t^{(i)}] \sim q_{\phi}(x|y)} [r_{\eta}([\hat{x}_t^{(i)}]_{q_t})]) \geq \\ & (\mathbb{E}_{[\hat{x}_t^{(i)}] \sim q_{\phi^*}(x|y)} [r_{\eta^*}([\hat{x}_t^{(i)}]_{q_t})] - \mathbb{E}_{[\hat{x}_t^{(i)}] \sim q_{\phi_{\text{new}}}(x|y)} [r_{\tilde{\eta}}([\hat{x}_t^{(i)}]_{q_t})]) \end{aligned} \quad (38)$$

Now, combining the results of (38) and (39), we can write

$$\underbrace{(\text{exploit}_{(\phi^*, \eta^*)}^{\text{score}}(q_t) - \text{exploit}_{(\phi, \eta)}^{\text{score}}(q_t))}_B \geq \underbrace{(\text{exploit}_{(\phi^*, \eta^*)}^{\text{score}}(q_t) - \text{exploit}_{(\phi_{\text{new}}, \tilde{\eta})}^{\text{score}}(q_t))}_D \quad (39)$$

Finally, leveraging the results of (37) and (40), and utilizing the definition of (35), we can write

$$|\text{Score}_{(\phi^*, \eta^*)}^*(q_t) - \text{Score}_{(\phi, \eta)}(q_t)| \geq |\text{Score}_{(\phi^*, \eta^*)}^*(q_t) - \text{Score}_{(\phi_{\text{new}}, \tilde{\eta})}(q_t)|$$

□

G Proof of Proposition 2

Proof. We start with the definition of entropy H :

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

Following the results of [18], and by setting $\alpha_i = \alpha_j = 1$, we obtain:

$$\begin{aligned} \mathbb{E}_{\hat{x}_t} [\log p(\hat{x}_t | Q_t, y_{t-1})] & \propto - \sum_{i=0}^P \log \sum_{j=0}^P \exp \left\{ \frac{\|\hat{x}_t^{(i)} - \hat{x}_t^{(j)}\|_2^2}{2\sigma_x^2} \right\} \\ \mathbb{E}_{\hat{x}_t} [\log p(\hat{x}_t | Q_t, y_{t-1})] & \propto - \sum_{i,j} \log \left(\exp \left(\frac{\sum_{a \in Q_t} ([\hat{x}_t^{(i)}]_a - [\hat{x}_t^{(j)}]_a)^2}{2\sigma_x^2} \right) \right) \end{aligned}$$

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

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

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

491 We then compute the expected log-likelihood at a specified measurement location q_t , discarding all
 492 terms independent of q_t . This key observation allows us to simplify the expression as follows:

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

493 Equivalently, we can write the above expression as:

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

494 By definition, the left-hand side of the above expression corresponds to $\text{likeli}^{\text{score}}(q_t)$. \square

495 H Doob's h -transform as the Correction Factor

496 We demonstrate that the h -transform serves as a correction term for Tweedie's estimate. Specifically,
 497 the conditional Tweedie estimate can be expressed as:

$$\begin{aligned} \mathbb{E}[x_0 | x_t, y] \approx \hat{x}_0(x_t, y) &= \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \left(h_t^\zeta(x_t, y) + s_t^{\theta^*}(x_t) \right)}{\sqrt{\bar{\alpha}_t}} \\ &= \underbrace{\left(\frac{x_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} s_t^{\theta^*} \right)}_{\text{Unconditional Tweedie estimate}} - \underbrace{\frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} h_t^\zeta(x_t, y)}_{\text{Correction Factor (i.e., } h\text{-transform)}} \end{aligned}$$

499 In Equation (40), the first term represents the unconditional Tweedie estimate, illustrating that the
 500 h -transform can be viewed as a correction to the unconditional denoised prediction.

501 H.1 Details of h -model Update Scheduler

502 As highlighted in the main paper, the pessimistic updating of the h -model parameters plays a critical
 503 role in stabilizing the prior model's adaptability dynamics. While a straightforward heuristic might
 504 suggest updating the h -model after a fixed number of observations, this approach only provides
 505 marginal improvements. A more effective scheduling strategy requires fewer updates during the
 506 early stages of discovery, allowing the model to gather ample data and avoid the risk of erroneous
 507 updates when the observations are still sparse. However, as the discovery process progresses and the
 508 model's understanding of the search space strengthens, more frequent updates of the h -model become
 509 essential. This shift enables more effective exploitation of the environment, as the model has already
 510 gathered enough information, making the need for pessimistic updates unnecessary. Motivated by
 511 this observation, we propose the following h -model update scheduler:

$$\Delta t_i = \frac{\mathcal{B}}{U} \cdot \left(1 - \frac{i}{U+1} \right)^\gamma \quad (40)$$

512 Here, \mathcal{B} denotes the overall sampling budget, U is the total number of h -model updates throughout
 513 the active discovery process, i indicates the current sampling step, γ governs the decay rate, and Δt_i
 514 defines the interval between two successive updates of the h -model parameters at the i -th sampling
 515 step. Note that since $\gamma \geq 1$, the update frequency of the h -model naturally accelerates with increasing
 516 i , leading to more frequent updates during the later stages of the active discovery process.

517 I Training and Inference Pseudocode

Algorithm 1 H-TRANSFORM FINE-TUNING (AT THE t -TH OBSERVATION STEP)

Require: Posterior Samples drawn from $q_{\phi_{t-1}}(x \mid y_{t-1})$

Require: Noise schedule $\beta_t = \beta(t)$, $\bar{\alpha}_t = \bar{\alpha}(t)$

Require: Permanent Memory (i.e. Pre-Trained Noise predictor function) $s_t^{\theta^*}(x)$ with parameters θ^* .

Require: Current state of Transient Memory (i.e. h -transform) $h_t^\zeta(x, \hat{x}_0, y)$ with parameters ζ .

```

1: repeat
2:    $x_0 \sim P_0 = q_{\phi_{t-1}}(x \mid y_{t-1})$ 
3:    $t \sim \text{Uniform}(\{1, \dots, T\})$ 
4:    $\varepsilon_t \sim \mathcal{N}(0, I)$  ▷ Sample noise
5:    $x_t \leftarrow \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\varepsilon_t$ 
6:    $\hat{\varepsilon}_\theta \leftarrow s_t^{\theta^*}(x_t)$  ▷ Estimate noise with pretrained model
7:    $\hat{x}_0 \leftarrow \frac{x_t - \sqrt{1 - \bar{\alpha}_t}\hat{\varepsilon}_\theta}{\sqrt{\bar{\alpha}_t}}$ 
8:    $\hat{\varepsilon}_\zeta \leftarrow h_t^\zeta(x_t, \hat{x}_0, y)$  ▷ Estimate correction via h-transform
9:   Take gradient descent step w.r.t.  $\zeta$  on
10:   $\nabla_\zeta \mathcal{L}(\varepsilon_t, \hat{\varepsilon}_\theta + \hat{\varepsilon}_\zeta)$  ▷  $\mathcal{L}$  is defined in Equation 7
11: until convergence or maximum epochs reached
12: return Updated  $h$ -model Parameter  $\zeta$ .

```

518 J Exploratory Nature of EM-PTDM

519 In active target discovery under an uninformative prior, exploration isn't just helpful—it's essential.
520 Especially in the early stages, when the permanent memory offers little insight into the target domain
521 and the correction factor is large, the h -model must rapidly adapt. This demands smart, strategic
522 exploration of the search space to collect informative observations, enabling the h -model to efficiently
523 learn and calibrate the correction factor. To assess EM-PTDM's exploratory behavior, we tackle active
524 target discovery of overhead objects using ground-level imagery as prior knowledge, evaluating across
525 a wide range of observation budgets—from very sparse (200) to less sparse (350)—and benchmark
526 against baseline methods. We present the results in the following Table 4.

Table 4: Importance of Exploration

ATD of Overhead Objects with ImageNet as Prior.				
Method	$\mathcal{B} = 200$	$\mathcal{B} = 250$	$\mathcal{B} = 300$	$\mathcal{B} = 350$
DiffATD	0.3873	0.5143	0.6391	0.7348
GA	0.3479	0.4784	0.5659	0.6562
EM-PTDM	0.4127	0.5620	0.7013	0.8256

527 We observe that EM-PTDM's performance improvement over baselines grows with the observation
528 budget. When the budget is low, the performance gap is narrow, reflecting limited opportunity for
529 exploration. As the budget increases, EM-PTDM leverages richer exploration to adapt its h -model,
530 leading to significantly more effective target discovery.

531 K Species Distribution Modelling as Active Target Discovery Problem

532 We constructed our species distribution experiment using observation data of the chosen species from
533 iNaturalist. Center points were randomly sampled within North America (latitude 25.6°N to 55.0°N,
534 longitude 123.1°W to 75.0°W). Around each center, we defined a square region approximately 480
535 km \times 480 km in size (roughly 5 degrees in both latitude and longitude). Each retained region was
536 discretized into a 64 \times 64 grid, where the value of each cell represents the number of observed species.
537 To simulate the querying process, each 2 \times 2 block of grid cells was treated as a query.

Algorithm 2 EM-PTDM SAMPLING STRATEGY (AT THE t -TH OBSERVATION STEP)

Require: Current State of Transient Memory: Trained h -transform $h_t^\zeta(x, \hat{x}_0, y)$ with parameters ζ .

Require: Permanent Memory as Unconditionally trained noise predictor $s_t^{\theta^*}(x_t)$

Require: Noise schedule $\beta_t = \beta(t)$, $\bar{\alpha}_t = \bar{\alpha}(t)$

Require: Sampling schedule $\sigma_t = \sigma(t)$

Require: Observation y , Posterior samples list $ps = []$, Success = $R = 0$.

```
1:  $x_T \sim P_T = \mathcal{N}(0, 1)$  ▷ Sample a starting point
2: for  $i = P$  to 1 do
3:    $\hat{x}^i = 0$ 
4:   for  $t$  in  $(T, T - 1, \dots, 1)$  do
5:      $\hat{\epsilon}_\theta \leftarrow s_t^{\theta^*}(x_t)$  ▷ Predict unconditional noise
6:      $\hat{x}_0 \leftarrow \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}_\theta}{\sqrt{\bar{\alpha}_t}}$ 
7:      $\hat{\epsilon}_\zeta \leftarrow h_t^\zeta(x_t, \hat{x}_0, y)$  ▷ Predict correction noise via  $h$ -transform
8:      $\hat{\epsilon} \leftarrow \hat{\epsilon}_\theta + \hat{\epsilon}_\zeta$  ▷ Estimate posterior noise
9:     if  $t > 1$  then
10:      Sample  $\epsilon_t \sim \mathcal{P}_{\text{noise}}$ 
11:     else
12:       $\epsilon_t \leftarrow 0$ 
13:     end if
14:      $x_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}} \left( \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \hat{\epsilon} + \sigma_t \epsilon_t$ 
15:     if  $t = 1$  then
16:       $\hat{x}^i = x_{t-1}$ 
17:     end if
18:   end for
19:    $ps.append(\hat{x}^i)$ 
20: end for
21: Utilize Posterior samples in  $ps$  to compute  $\text{expl}^{\text{score}}(q_t)$  and  $\text{exploit}^{\text{score}}(q_t)$  using Eqn. 9 and 10
    respectively for each  $q_t \in k$ .
22: Compute  $\text{score}(q_t)$  using Eqn. 11 for each  $q_t \in k$  and sample a location  $q_t$  with the highest
    score.
23:  $\mathcal{B} \leftarrow \mathcal{B} - 1$ ,  $\{k\} \leftarrow \{k\} \setminus q_t$ 
24: Update:  $Q_t \leftarrow Q_{t-1} \cup q_t$ ,  $y_t \leftarrow y_{t-1} \cup [x]_{q_t}$ .
25: Update:  $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{[x]_{q_t}, y^{(q_t)}\}$ ,  $R += y^{(q_t)}$ 
26: Train  $r_\eta$  with updated  $\mathcal{D}_t$  and optimize  $\eta$  with Cross-Entropy loss.
27: return  $R$ 
```

538 **L Analyzing the Role of Permanent and Transient Memory for Enhancing**
539 **In-Domain Target Discovery**

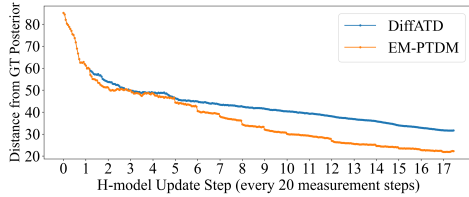
540 Since we’ve seen that inside the EM-PTDM framework, updating permanent memory with posteriors
541 from prior in-domain ATD tasks boosts performance, it raises a key question: do we still need the
542 h -model? To investigate, we turn to DiffATD—a state-of-the-art baseline that uses only permanent
543 memory. To update the permanent memory of DiffATD, we apply the same continual memory
544 update strategy as in EM-PTDM, incorporating accumulated posteriors after each task, to see how far
545 performance can go without the h -model. For this comparison, we examine active target discovery of
546 overhead objects using ground-level ImageNet images as prior knowledge. While updating only the
547 permanent memory leads to improved discovery rates compared to DiffATD with fixed permanent
548 memory across different observation budgets, a clear and consistent performance gap remains when
549 compared to EM-PTDM, particularly when EM-PTDM updates its permanent memory after each
550 in-domain ATD task. These results, summarized in Table 5, highlight the added value of the h -model
551 in driving more effective exploration and adaptation irrespective of whether permanent memory is
552 being updated or not.

Table 5: Importance of h -model with or without Permanent Memory (PM) Update

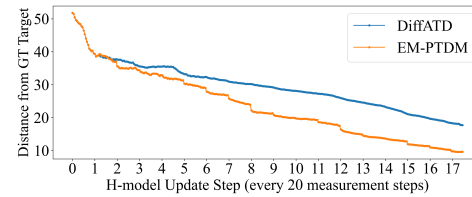
ATD of Overhead Objects with ImageNet as Prior.			
Method	$\mathcal{B} = 250$	$\mathcal{B} = 300$	$\mathcal{B} = 350$
DiffATD	0.5143	0.6391	0.7348
DiffATD w/ PM Update	0.5294	0.6623	0.7589
EM-PTDM	0.5620	0.7013	0.8256
EM-PTDM w/ PM Update	0.5859	0.7194	0.8461

M Efficacy of h -model in Estimating The Search Space with only Few Observations

We analyze the adaptability of the h -model using quantitative visualizations in the context of an active target discovery of overhead objects, where ground-level ImageNet images serve as the prior. Specifically, we assess the h -model’s role by computing L2-based semantic similarity between predicted and ground-truth targets (right 7b), and between predicted and ground-truth posteriors (left 7a). As depicted in Figure 7, the inclusion of the h -model leads to significantly faster convergence to the true posterior and the true targets, thus enabling more informed target discovery. The rapid drop in dissimilarity compared to using permanent memory alone highlights the h -model’s ability to quickly correct the prior with minimal observations.



(a) L2 Distance between predicted and ground-truth posterior.



(b) L2 Distance between predicted and ground-truth targets.

Figure 7: Quantitative analysis of h -model’s adaptability in Active Target Discovery.

N An Important Observation: ATD is not just About Reconstruction

One might ask: if the primary goal is for the posterior samples to accurately reconstruct the search space, isn’t that sufficient for efficient target discovery? Interestingly, in our setting, a precise reconstruction of the entire search space is not strictly necessary, as long as the model effectively identifies and reconstructs the target regions, efficient discovery can still be achieved. To validate this hypothesis, we conduct an experiment and visualize the posterior at intermediate stages of active target discovery. We compare posterior samples from EM-PTDM and DiffATD using a representative task where EM-PTDM significantly outperforms DiffATD, allowing us to understand how the posterior contributes to improved target discovery. We present the visualization in the Figure 8. Our observations reveal that, while EM-PTDM’s posterior samples exhibit lower overall reconstruction quality compared to those from DiffATD, they more effectively focus on target-rich regions (For example, see the highlighted Red box in the Figure 8), leading to significantly improved target discovery performance. This highlights a key insight: successful target discovery relies more on accurately modeling the regions of interest than on reconstructing the entire search space.

ATD is NOT just Reconstruction!!



DiffATD Posterior at Intermediate Active Target Discovery Phase

EM-PTDM Posterior at Intermediate Active Target Discovery Phase

Ground Truth Posterior

Figure 8: ATD is about Discovering Targets, NOT just Reconstructing the Search Space.

O Effect of $\kappa(\mathcal{B})$

We conduct experiments to assess the impact of $\kappa(\mathcal{B})$ on EM-PTDM’s active discovery performance. Specifically, we investigate how amplifying the exploration weight, by setting $\kappa(\mathcal{B}) =$

Table 6: Effect of $\kappa(\mathcal{B})$

Performance across varying α with $\mathcal{B} = 250$				
Target	Prior	$\alpha = 0.2$	$\alpha = 1.0$	$\alpha = 5.0$
Balls	MNIST	0.7416	0.7875	0.9272

$\max\{0, \kappa(\alpha \cdot \mathcal{B})\}$ with $\alpha > 1$, and enhancing the exploitation weight by setting $\alpha < 1$, influence the overall effectiveness of the approach. We report results for $\alpha \in \{0.2, 1, 5\}$ in two settings: using overhead objects as targets with ground-level images as the prior (first row), and discovering target balls using MNIST digit images as the prior (second row). The results are summarized in Table 6. The best performance is achieved with $\alpha = 5$, and the results suggest that higher values of α boosts performance, reinforcing the fact that exploration is key to success in active target discovery under an uninformative prior.

P EM-PTDM’s Capability of Discovering Isolated Targets within Observation Budget

To evaluate EM-PTDM’s ability to uncover disjoint target regions within a limited sampling budget, we design a series of controlled toy experiments. In each task, the goal is to discover a varying number of balls—positioned differently and with different radii—using a diffusion model pretrained on MNIST digits as the permanent memory. We systematically increase task difficulty by varying the number of target balls from 5 to 10. Notably, tasks with more targets demand effective exploration of the search space to successfully locate all disjoint regions within the budget constraints. We present comparative visualizations of the exploration behavior of EM-PTDM and DiffATD with an active discovery task involving uncovering 10 target balls with MNIST Images as the prior, shown in Fig. 10. As the number of disjoint targets increases, making the task more challenging and exploration-intensive, EM-PTDM consistently succeeds in discovering most, if not all, targets within the given budget. In contrast, the baseline (i.e., DiffATD) relying solely on permanent memory struggles as task complexity rises, as seen in Plot 9. These visualizations clearly demonstrate EM-PTDM’s superior exploration capabilities, which are crucial for efficient target discovery in settings with multiple disjoint targets.

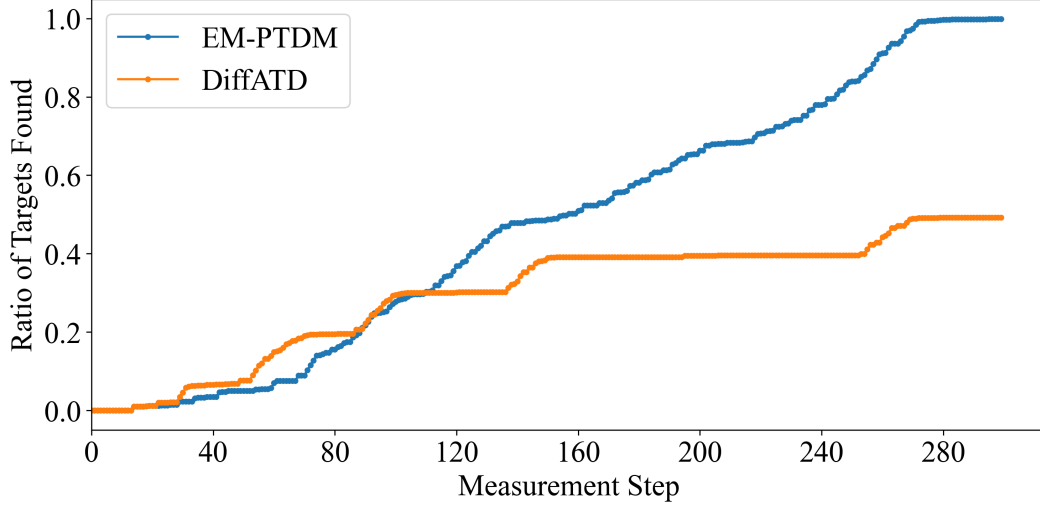


Figure 9: Comparison of the Discovery Process: EM-PTDM vs. DiffATD. In this experiment, we evaluate the active discovery of 10 target balls using a diffusion model trained on MNIST images as the prior. As the search budget increases, EM-PTDM consistently uncovers more disjoint target balls, thanks to its inherently exploratory behavior. In contrast, DiffATD struggles to discover as many targets as it lacks the same efficiency as EM-PTDM in exploring the search space.

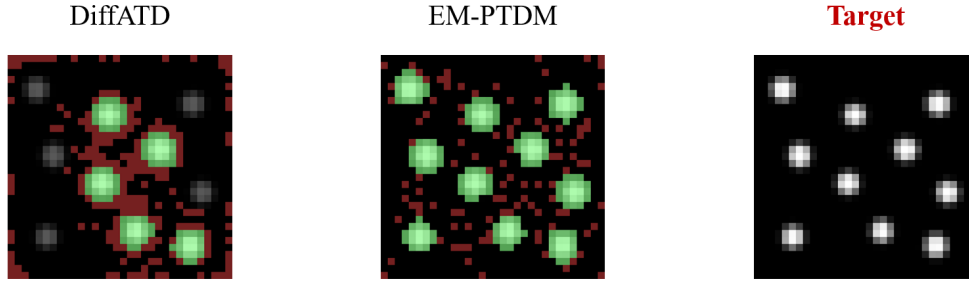


Figure 10: Comparison between EM-PTDM and DiffATD’s Discovery. **Green** patches correspond to successful target discovery, and **Red** Patches correspond to unsuccessful observations. In this example, the task is to discover the balls with MNIST images as the prior. The exploratory behavior of EM-PTDM, in contrast to DiffATD, is evident in the visualization.

Q Additional Results on Active Discovery of Unknown Species from Known Species Distribution

In the main paper, we explored active discovery of *Coccinella septempunctata* (CS) using the known species distribution of *Gladicosa* and *Gonioctena* (GG) as the prior. This section extends our analysis to a different species from the iNaturalist dataset. Specifically, we evaluate EM-PTDM on a task that involves discovering Species Cedar Waxwing from the known distribution of Species Black-capped Chickadee. Figure 11 compares the exploration behavior of EM-PTDM and DiffATD at various stages of target discovery. As shown, EM-PTDM—starting from the same prior as DiffATD—progressively and efficiently adapts the prior toward the true target distribution with only a few task-specific observations. In contrast, DiffATD, which relies solely on static permanent memory, struggles to approximate the ground-truth distribution within the given observation budget.

We further compare the performance of EM-PTDM against baseline methods using Success Rate (SR) as the evaluation metric, with results summarized in Table 7. The task involves actively discovering

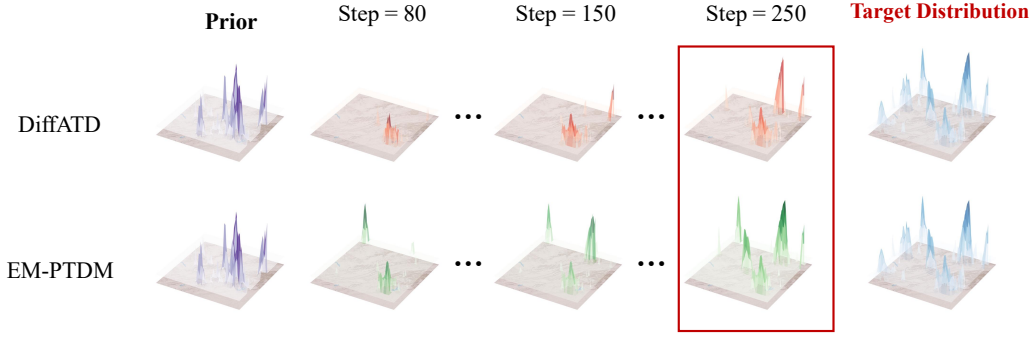


Figure 11: Exploration Behavior of Different Approaches. We visualize the explored Regions at Different Active Target Discovery Phases. We consider the task of Active Discovery of the species Cedar Waxwing with a known distribution of the species Black-capped Chickadee. EM-PTDM discovers most target regions (i.e., more accurately discovers the existence of the species Cedar Waxwing).

Table 7: *SR* Comparison with Baselines.

Active Discovery of Cedar Waxwing Species with Species Black-capped Chickadee as Prior.			
Method	$\mathcal{B} = 150$	$\mathcal{B} = 200$	$\mathcal{B} = 250$
DiffATD	0.2319	0.3309	0.4453
GA	0.2232	0.3098	0.3116
EM-PTDM	0.3079	0.3854	0.6347

Species Cedar Waxwing from the known distribution of Species Black-capped Chickadee. Consistent with other experimental settings, EM-PTDM significantly outperforms all baselines across varying observation budgets. These results further highlight the effectiveness of EM-PTDM in tackling active target discovery under an uninformative prior.

R Effect of h -model Update Scheduler

In this section, we evaluate the effectiveness of the h -model update scheduler by comparing two variants of EM-PTDM: one using a uniform update schedule and the other employing the adaptive scheduler defined in Equation 40. For the uniform scheduler, the h -model is updated at fixed intervals—specifically, every 20 update steps. For the adaptive scheduler, we set $\gamma = 1$, $U = 30$, and an observation budget of $\mathcal{B} = \{200, 250\}$. The comparative results under this configuration are presented in Table 8. For this analysis, we consider active discovery of balls with a diffusion model trained on MNIST data as the prior model. Our empirical results show that EM-PTDM with an adaptive h -model update scheduler consistently outperforms its uniform counterpart across various measurement budgets. This improvement can be attributed to fewer updates in the early stages, allowing the model to collect more informative observations before updating and thus reducing the risk of premature or noisy updates. As the discovery progresses and the model gains a stronger understanding of the search space, the increased update frequency of h -model in later stages proves beneficial for accelerating active target discovery.

Table 8: Effect of Adaptive h -model Update Scheduler.

Active Discovery of Balls with MNIST Digit Images as the Prior.		
h -model update Schedule	$\mathcal{B} = 200$	$\mathcal{B} = 250$
Uniform	0.6856	0.7875
Adaptive	0.7364	0.8268

S More Visualizations on Efficiency of h -model's Adaptability From Very Sparse Observations

In this section, we provide additional visualizations of posterior samples generated by EM-PTDM and DiffATD across different stages of active target discovery. As shown in Figures (12, 13, 14), EM-PTDM produces samples that are more semantically aligned with the ground-truth posterior compared to DiffATD. Notably, even with sparse observations, EM-PTDM effectively simulates the search space, enabling more informed exploration and leading to improved target discovery under an uninformative prior.

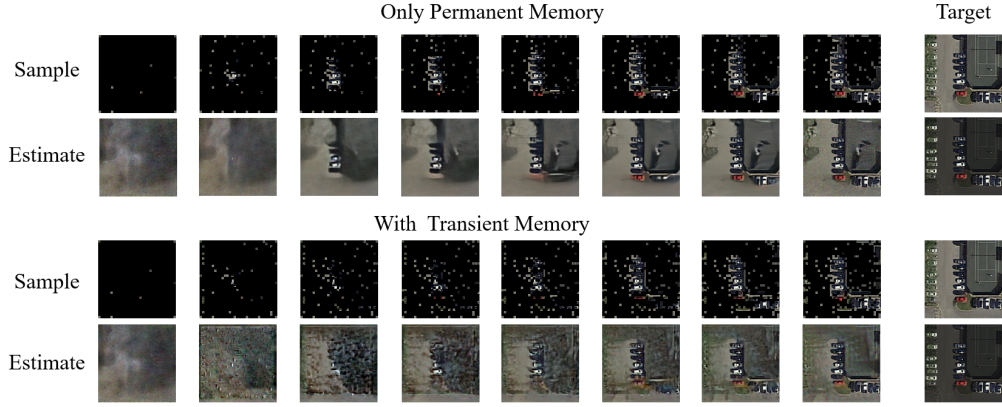


Figure 12: h -model's Adaptability From Very Sparse Observations. Posterior Samples at Different Active Target Discovery Phases. Overhead Object Discovery (i.e., **Car**) with Ground Level images from ImageNet as the Prior. Observation Budget of 300.

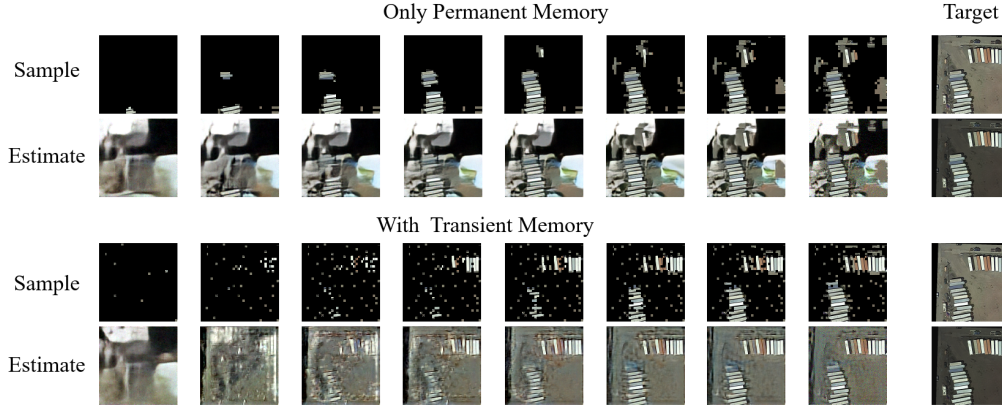


Figure 13: h -model's Adaptability From Very Sparse Observations. Posterior Samples at Different Active Target Discovery Phases. Overhead Object Discovery (i.e., **Truck**) with Ground Level images from ImageNet as the Prior. Observation Budget of 300.

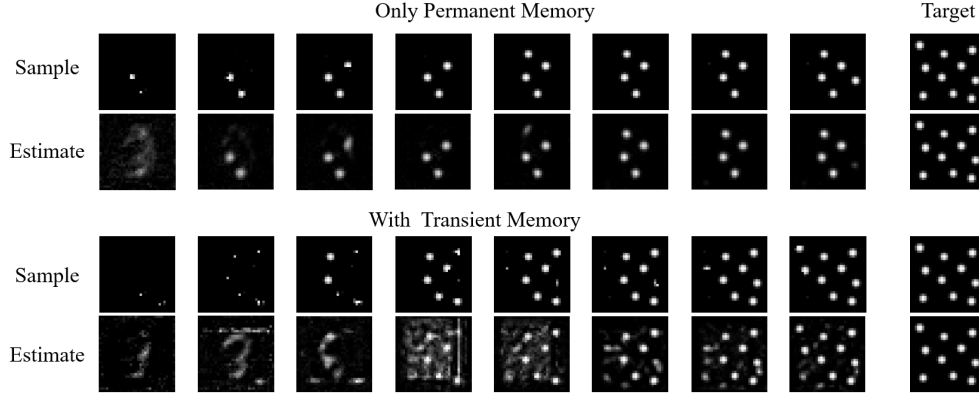


Figure 14: h -model’s Adaptability From Very Sparse Observations. Posterior Samples at Different Active Target Discovery Phases. Uncovering Disjoint Balls with MNIST digit images as the Prior.

T More Visualizations of the Exploration Behavior of EM-PTDM at Different Active Target Discovery Phases

In this section, we present additional exploration behavior of EM-PTDM at different active target discovery phases. We also provide a similar exploration behavior of the baseline approaches, including DiffATD and Greedy Adaptive, for the comparison. These visualizations are provided in Figures 15, 16, 17.

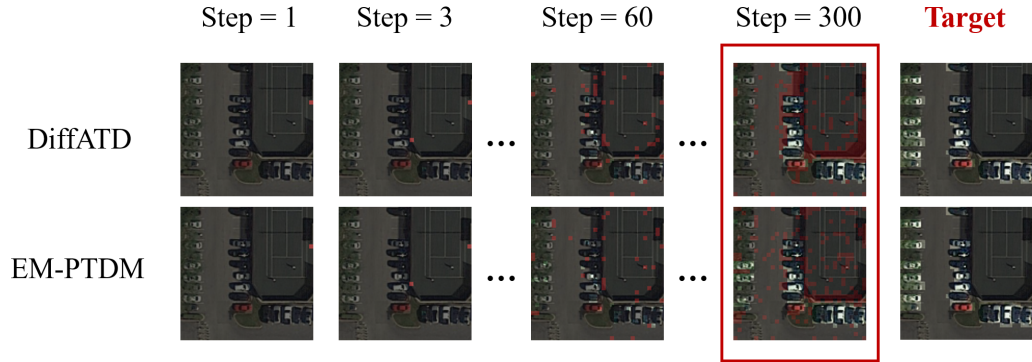


Figure 15: Visualizing the regions explored by each method at various stages of the ATD process. In these visualizations, patches corresponding to **successful queries are unmasked**, while those resulting in **unsuccessful queries are highlighted in Red**. The task focuses on discovering overhead objects (**cars**), using ground-level images from ImageNet as the prior. The results demonstrate that EM-PTDM effectively identifies and explores most of the target regions containing **cars**.

These additional visualizations further reinforce the effectiveness of EM-PTDM in addressing active target discovery under an uninformative prior.

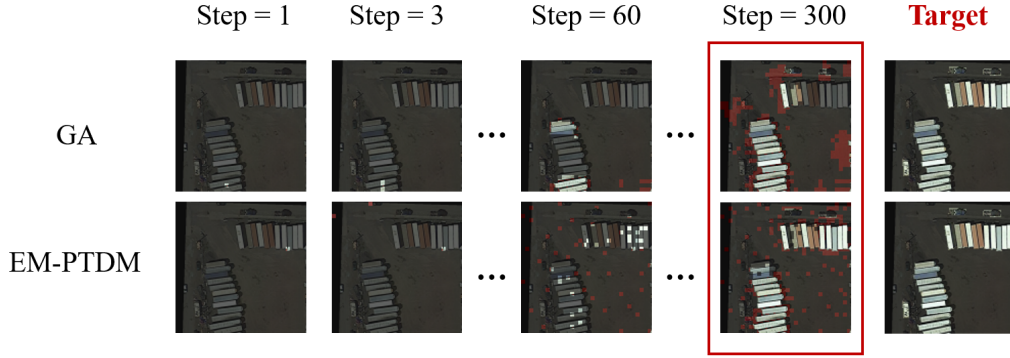


Figure 16: Visualizing the regions explored by each method at various stages of the ATD process. In these visualizations, patches corresponding to **successful queries are unmasked**, while those resulting in **unsuccessful queries are highlighted in Red**. The task focuses on discovering overhead objects (**trucks**), using ground-level images from ImageNet as the prior. The results demonstrate that EM-PTDM effectively identifies and explores most of the target regions containing **trucks**.

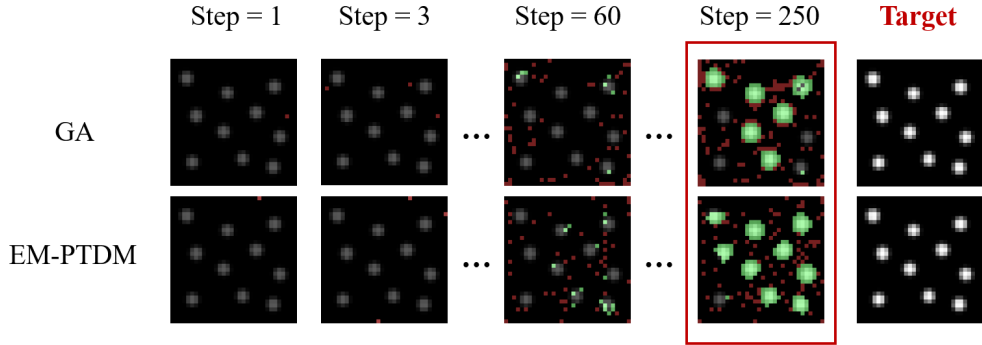


Figure 17: Visualizing the regions explored by each method at various stages of the ATD process. In these visualizations, patches corresponding to **successful queries are Green**, while those resulting in **unsuccessful queries are highlighted in Red**. The task focuses on uncovering Disjoint Balls from MNIST digit images as the Prior. EM-PTDM discovers most target regions (i.e., Disjoint Balls).

650 U Active Target Discovery of Balls Using MNIST Images as the Prior

651 To enable a comprehensive analysis of EM-PTDM, we introduce a custom-designed dataset tailored
652 for this task. It simulates active discovery scenarios involving an unknown number of balls with
653 unknown locations and radii, with MNIST images as the prior. Success in this setting requires effective
654 exploration of the search space to accurately localize the targets, capturing the core challenge of the
655 problem. The details of the proposed dataset are provided below.

656 **Dataset Creation Procedure** In this dataset, we generate each sample by randomly placing 5 to 10
657 identical balls within a 32×32 2D grid. The radius of all balls in a given sample is either 3 or 4 pixels,
658 randomly selected per sample. All placements are performed uniformly at random, subject to the
659 non-overlapping constraint and the boundary condition that each ball lies entirely within the 32×32
660 space.

661 **SR Comparisons with Baseline Approaches** As in previous settings, we quantitatively evaluate
662 EM-PTDM and baseline methods using the Success Rate (SR) metric. In this experiment, the task
663 involves actively discovering target balls using a diffusion model trained on the MNIST dataset as the

prior. The results, summarized in Table 9, show a consistent trend: EM-PTDM significantly outperforms all baselines across different measurement budgets. This further reinforces the effectiveness of EM-PTDM in handling active target discovery under an uninformative prior.

Table 9: *SR* Comparison with Baselines.

Active Discovery of balls with MNIST Digit Images as Prior.			
Method	$\mathcal{B} = 150$	$\mathcal{B} = 200$	$\mathcal{B} = 250$
RS	0.1458	0.1826	0.2187
DiffATD	0.4362	0.4432	0.4929
GA	0.3250	0.5170	0.6257
EM-PTDM	0.5561	0.6856	0.7875

Analyzing the Exploration Strategies of EM-PTDM and DiffATD Under Increasing Task Complexity In this section, we provide additional visualizations highlighting the exploration behavior of EM-PTDM and DiffATD across different stages of the active target discovery task. Using the task of discovering target balls with MNIST images as the prior, the visualizations in Figures 18, 19 clearly show that EM-PTDM consistently explores more effectively and identifies targets with higher accuracy, even under increasing task complexity while adhering to a strict budget and under an uninformative prior. These results further underscore the robustness and adaptability of EM-PTDM in challenging discovery scenarios, where efficient exploration of the search space is the key. A striking emergent behavior is observed across both examples: in the early stages of the active discovery process, EM-PTDM engages in broader exploration of the search space compared to DiffATD. This initially results in fewer target discoveries (e.g., at step 60). However, this strategic exploration enables EM-PTDM to build a richer understanding of the environment, which it later exploits to surpass DiffATD, ultimately identifying a greater number of target regions before the observation budget is depleted (see step 250).

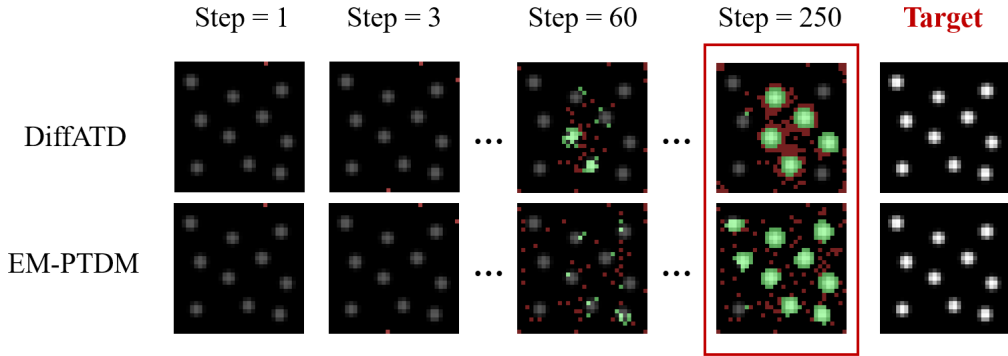


Figure 18: Visualizing the regions explored by each method at various stages of the ATD process. In these visualizations, patches corresponding to **successful queries are highlighted in Green**, while those resulting in **unsuccessful queries are highlighted in Red**. The task focuses on uncovering Disjoint Balls from MNIST digit images as the Prior. The results demonstrate that EM-PTDM discovers most target regions (i.e., Disjoint Balls).

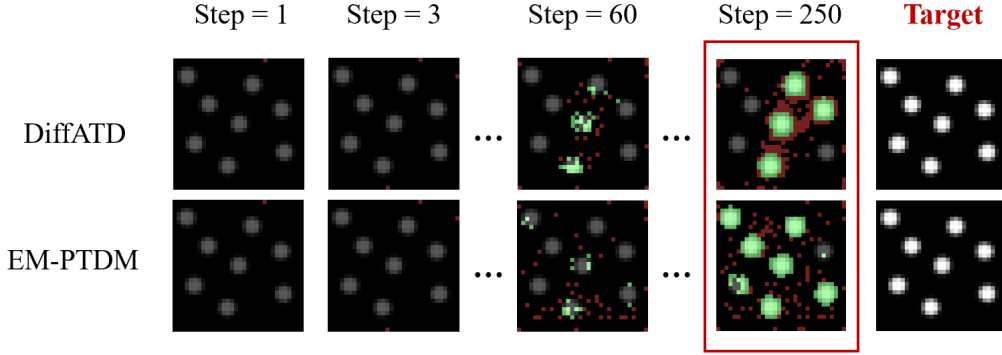


Figure 19: Visualizing the regions explored by each method at various stages of the ATD process. In these visualizations, patches corresponding to **successful queries are highlighted in Green**, while those resulting in **unsuccessful queries are highlighted in Red**. The task focuses on uncovering Disjoint Balls from MNIST digit images as the Prior. The results demonstrate that EM-PTDM discovers most target regions (i.e., Disjoint Balls).

V Architecture, Training Details: h -model, Pretrained Diffusion Model, and the Reward Model; and Computing Resources

Details of h -model For the MNIST-to-Balls tasks, we employ 32-dimensional diffusion time-step embeddings and a single-block U-Net h -model with layer widths of [32, 64]. In the species discovery tasks, we also use 32-dimensional time-step embeddings, but with a single-block h -model featuring wider layers [32, 64, 128]. For the ImageNet-to-DOTA tasks, we increase the embedding dimensionality to 128, using a similar single-block h -model with widths [32, 64, 128]

Details of Reward Model Our proposed method, EM-PTDM, utilizes a parameterized reward model, r_η , to steer the exploitation process. To this end, we employ a neural network consisting of a series of convolutional and fully connected layers, with non-linear ReLU activations as the reward model (r_η). 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 DOTA, we use an 4×4 patch as the input feature size. After each measurement step, we update the model parameters (η) using the binary cross-entropy loss. 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 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 $\frac{(\text{input size})^2}{4}$ to an output of size 4 with weights and biases of size $[\frac{(\text{input size})^2}{4}, 4]$ 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 binary cross-entropy loss. 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.

Details of Primary Memory as Pretrained Diffusion Model We use DDIM [21] as the diffusion model across datasets. The diffusion models used in different experiments are based on widely

714 adopted U-Net-style architecture. For the MNIST dataset, we use 32-dimensional diffusion time-step
715 embeddings, with the diffusion model consisting of 2 residual blocks. We select the time-step
716 embedding vector dimension to match the input feature size, ensuring the diffusion model can process
717 it efficiently. The block widths are set to [32, 64, 128], and training involves 30 diffusion steps. For
718 DOTA, we use the input feature size of [128, 128, 3], the architecture featuring 128-dimensional
719 time-step embeddings and a diffusion model with 2 residual blocks of width [64, 128, 256, 256, 512].
720 Finally, all experiments are implemented in Tensorflow and conducted on NVIDIA A100 40G GPUs.
721 Our training and inference code will be made public.

722 W Visual Illustration of EM-PTDM Sampling Strategy

We present a pictorial illustration of our proposed EM-PTDM approach in Figure 20.

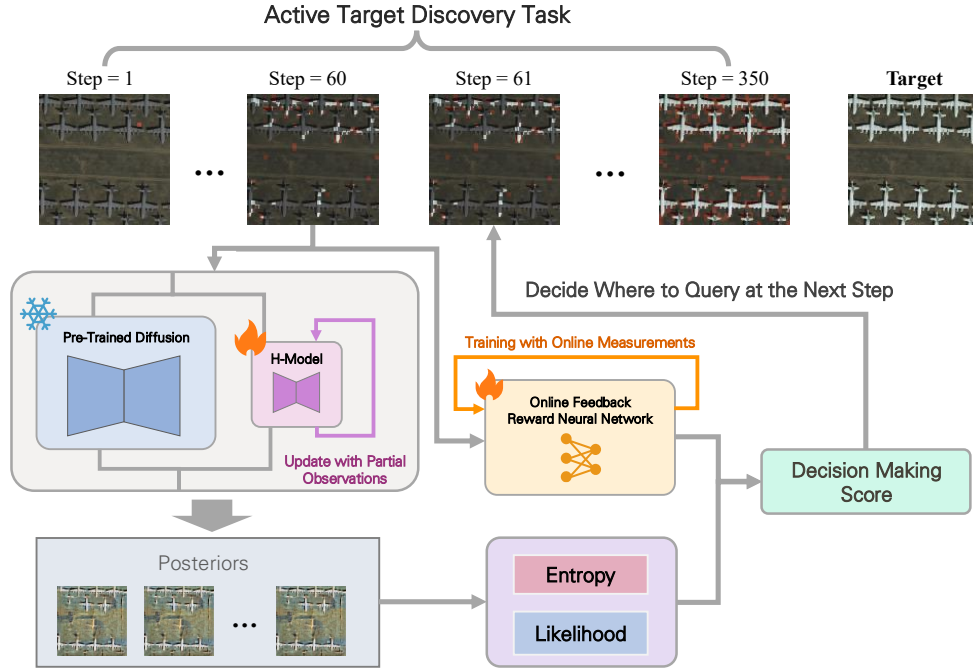


Figure 20: Overview of EM-PTDM Framework.

723

724 X Statistical Significance Results of EM-PTDM

725 In order to strengthen our claim on EM-PTDM’s superiority over the baseline methods, we have
726 included the statistical significance results with different active target discovery settings, and present
727 the results in Tables 10, 11. These results are based on 5 independent trials and further strengthen our
728 empirical findings, reinforcing the stability and effectiveness of EM-PTDM in tackling active target
729 discovery under an uninformative prior across diverse domains.

Table 10: Statistical Significance Results for Unknown Overhead Object Discovery.

Active Discovery of Overhead Objects with Ground Level ImageNet Images as the Prior.			
Method	$\mathcal{B} = 250$	$\mathcal{B} = 300$	$\mathcal{B} = 350$
RS	0.2325 ± 0.0190	0.2852 ± 0.0137	0.3207 ± 0.0168
DiffATD	0.5143 ± 0.0067	0.6391 ± 0.0102	0.7348 ± 0.0041
GA	0.4784 ± 0.0122	0.5659 ± 0.0096	0.6562 ± 0.0054
EM-PTDM	0.5620 ± 0.0073	0.7013 ± 0.0038	0.8256 ± 0.0093

Table 11: Statistical Significance Results for Unknown Species Discovery Task.

Active Discovery of Species CS with Species GG as the Prior.			
Method	$\mathcal{B} = 150$	$\mathcal{B} = 200$	$\mathcal{B} = 250$
RS	0.1624 ± 0.0133	0.2327 ± 0.0201	0.2775 ± 0.0154
DiffATD	0.3420 ± 0.0115	0.4365 ± 0.0057	0.4808 ± 0.0063
GA	0.4061 ± 0.0047	0.5067 ± 0.0079	0.5567 ± 0.0085
<i>EM-PTDM</i>	0.4983 ± 0.0060	0.6495 ± 0.0108	0.6989 ± 0.0056