Distribution-Aware Energy Minimization: Physical-Inspired Efficient Active Learning and Quantum Potentials

Zhicheng Yao[1], **Wenguo Yang**[1], **Yancheng Chen**[2], **Dun Ma**[2], **Shengminjie Chen**[3,4,*] and **Xiaoming Sun**[3,4]

[1]School of Mathematical Sciences, University of Chinese Academy of Sciences
[2]School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences
[3]State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences
[4]School of Computer Science and Technology, University of Chinese Academy of Sciences

Abstract

Active learning aims to maximize model performance with minimal annotation costs by selecting the most informative samples from large unlabeled pools, which often face a budget dilemma: uncertainty-based methods induce redundancy under low budgets, while representativeness-based methods struggle to mine challenging samples under high budgets. Although some heuristic parameter interpolation schemes attempt to bridge this gap, such strategies suffer from a misalignment between theoretical assumptions and real-world distributions, failing to achieve a good exploration-exploitation balance across all scenarios. In this paper, we propose a general active learning framework based on Distribution-Aware Energy Minimization, which reformulates sample selection as minimizing the energy function for the distributional discrepancy between the selected subset and the global uncertainty field. This physical-inspired perspective naturally derives a Hamiltonian comprising attractive terms and repulsive terms, mathematically achieving an intrinsic and dynamic balance between uncertainty and representativeness. Furthermore, we transform the optimization objective to the ground-state search of an Ising Model, enabling efficient solutions via Coherent Ising Machine. Extensive numerical experiments show that our method outperforms prior state-of-the-art methods across multi-budget regimes on several benchmark datasets. Validation on a real quantum hardware also demonstrates the potential for quantum computer in future large-scale selection tasks.

1 Introduction

In deep learning, unlabeled data are abundant, yet obtaining high-quality annotations remains costly, time-intensive, and often demands specialized expertise. To address this, Active Learning (AL), as a compelling framework, has emerged to maximize model performance under limited annotation budgets by iteratively selecting the most informative samples. This framework is particularly indispensable in high-stakes domains requiring scarce expert intervention, such as medical imaging [Yang *et al.*, 2017], autonomous driving [Haussmann *et al.*, 2020], and scientific discovery [Bassman Oftelie *et al.*, 2018]. By bridging the gap between data abundance and annotation scarcity, AL serves as a vital engine for real-world deployment, allowing large-scale data to be utilized efficiently through minimal yet strategic supervision.

Traditional active learning strategies generally fall into two categories: 1) Uncertainty-based Methods [Lewis and Gale, 1994; Gal *et al.*, 2017; Ash *et al.*, 2019], which prioritize challenging samples near decision boundaries to maximize information gain; 2) Representativeness-based Methods [Sener and Savarese, 2017; Yehuda *et al.*, 2022; Hacohen *et al.*, 2022], which aim to select a geometrically representative subset that comprehensively covers the underlying data distribution. However, in practice, a single strategy rarely adapts well to all scenarios. There is a significant budget dilemma: under low budgets, over-reliance on uncertainty sampling often leads to failure. This occurs not merely due to unstable estimates, but primarily because high-uncertainty samples tend to cluster in specific regions. Consequently, batch selection collapses into severe local redundancy, wasting the limited budget on structurally repetitive data. Conversely, under high annotation budgets, while representativeness-based strategies ensure global coverage, they often fail to target the fine-grained, challenging instances required to further refine an already capable model, leading to performance saturation. To bridge this gap, UHerding [Bae *et al.*, 2025] introduces an uncertainty coverage objective, which is shown to degenerate to geometric coverage when uncertainty becomes constant across all samples, and to uncertainty sampling as the kernel bandwidth approaches zero. Meanwhile, the method aims to develop a framework that can smoothly transition from coverage to uncertainty across small and large budgets. In practice, however, we observe a critical disconnect between these theoretical asymptotes and real-world behavior. Under low annotation budgets, the assumption of constant uncertainty seldom holds, as samples already exhibit significant entropy differentiation even in early training stages. Conversely, under high annotation budgets, the operational kernel bandwidth remains substantially larger than zero. Consequently, the method fails to fully realize its

*Corresponding Author: Shengminjie Chen, csmj@ict.ac.cn

limiting theoretical behaviors in practice. Perhaps, the key to resolve this inconsistency lies not in ideal asymptotic assumptions, but in returning to the fundamental goal of active learning: accurately approximating the uncertainty distribution over the true data manifold.

To this end, we propose a novel active learning framework, namely Distribution-Aware Energy Minimization (DAEM). Departing from covering uncertainty, our method treats the unlabeled pool as a physical field in which each sample carries an energy mass defined by its uncertainty. The objective is to select a subset whose empirical distribution minimizes the statistical discrepancy from the global target uncertainty distribution. From this perspective, we naturally derive a physical Hamiltonian composed of both attractive terms and repulsive terms, which mathematically achieves an intrinsic and dynamic balance between uncertainty and representativeness without relying on hand-crafted parameter annealing strategies. Our main contributions are as follows:

- **Physical-Inspired Framework for Efficient Active Learning:** We formulate an energy minimization framework based on distribution matching, which can be solved by efficient greedy solutions with some theoretical guarantees. This framework naturally avoids redundancy in low-budget regimes through repulsive terms, while accurately targeting challenging samples in high-budget regimes via attractive terms, without relying on tuned hyper-parameters. Experiments demonstrate that our method outperforms previous State-of-the-Art across multiple benchmarks including balanced, imbalanced, and transfer learning tasks while maintaining higher computational efficiency on standard hardware compared to complex baselines.
- **Quantum Scalability via Ising Model:** We mathematically establish the equivalence between the sample selection problem and identifying the ground state of an Ising Model. This physical perspective not only is efficient on classical computers but also bridges the gap to emerging quantum computing paradigms, such as Coherent Ising Machines (CIM). Our experiments demonstrate the forward-looking nature of this approach, indicating that as quantum hardware matures, this framework is poised to unlock dual potential in both computational efficiency and model performance for ultra-large-scale selection tasks.

2 Related Work

Active Learning Strategies. Active learning seeks to maximize performance under limited labeling budgets. Classical methods rely on uncertainty sampling, including Entropy [1994] and Margin sampling [Scheffer *et al.*, 2001]. Bayesian approaches such as BALD [2017] and Batch-BALD [Kirsch *et al.*, 2019] measure epistemic uncertainty for individual or batch selection. To reduce inference costs, LPL [Yoo and Kweon, 2019] and DDU [Mukhoti *et al.*, 2023] introduce loss prediction or deterministic uncertainty, while [Chen *et al.*, 2024] addresses cold-start selection via gradients. Hybrid methods include BADGE [2019] and VAAL [Sinha *et al.*, 2019]. Other works exploit decision boundaries through adversarial perturbations [Ducoffe and Precioso, 2018], gradient changes [Koh and Liang, 2017], or interpolation spaces [Parvaneh *et al.*, 2022], and further extend active learning to robustness [Yuan *et al.*, 2023], open-world discovery [Schmidt *et al.*, 2025], reinforcement learning [Dukkipati *et al.*, 2025], and automated search [Wang *et al.*, 2024].

Geometric and Distributional Selection. To reduce sampling bias, geometric methods formulate active learning as coverage or distribution matching. DPPs [Kulesza and Taskar, 2012] model diversity with kernels, Core-set [2017] uses k-center clustering, and graph-based methods [Caramalau *et al.*, 2021] exploit topological coverage. Distributional methods include DAL [Gissin and Shalev-Shwartz, 2019] and WAAL [Shui *et al.*, 2020], while recent studies [Gupte *et al.*, 2024] show that robust pre-trained representations such as CLIP improve geometric coverage. TypiClust [2022] and ProbCover [2022] select typical instances or maximize probability mass. UHerding [Bae *et al.*, 2025] unifies uncertainty and coverage but often requires approximations. In contrast, DAEM takes a **distributional perspective**, directly minimizing an energy objective to reconstruct the **uncertainty distribution** without relying on asymptotic assumptions.

Combinatorial Optimization and Quantum Computing. Subset selection with cardinality constraints is combinatorial. Facility Location [Cornuejols *et al.*, 1977] and related objectives [Pinsler *et al.*, 2019; Wei *et al.*, 2015] use submodularity for greedy approximation. For quantum optimization, one line maps combinatorial problems to Ising/QUBO formulations [Lucas, 2014; Kochenberger *et al.*, 2014], which can be solved by specialized Ising machines, including the **Coherent Ising Machine (CIM)** [McMahon *et al.*, 2016; Inagaki *et al.*, 2016] and D-Wave quantum annealers [Johnson *et al.*, 2011]. Another line develops gate-based quantum optimization algorithms, such as QAOA [Farhi *et al.*, 2014; Lu *et al.*, 2023], variational quantum algorithms for combinatorial optimization [Cerezo *et al.*, 2021], and Lyapunov-based quantum algorithmic frameworks with approximation guarantees [Chen *et al.*, 2025]. Our work connects active learning with this **photonic computing paradigm** by constructing an Ising Hamiltonian for the distributional energy landscape, enabling optical parallelism for global ground-state search.

3 Distribution-Aware Energy Minimization

3.1 Problem Definition

We consider a standard pool-based active learning setup. Let $\mathcal{U} = \{x_i\}_{i=1}^{N}$ denote a large pool of unlabeled data instances, where samples are drawn i.i.d. from the marginal distribution $\mathcal{P}_X$. Our core task is to devise a query strategy that selects an optimal subset $S \subset \mathcal{U}$ with a fixed budget constraint $|S| = B$ at each round. These selected samples are then annotated by an oracle and used to update the model.

We formulate active sampling as a Constrained Risk Minimization problem for a single round. Fundamentally, we aim to select a subset S such that, after updating the model $f_{\theta|S}$ using the information from S, the expected risk on the true

underlying distribution $\mathcal{P}_{XY}$ is minimized:

$$\min_{S\subset\mathcal{U},|S|=B} \mathcal{R}(f_{\theta|S}) = \min_{S\subset\mathcal{U},|S|=B} \mathbb{E}_{(x,y)\sim\mathcal{P}_{XY}}[\ell(f_{\theta|S}(x), y)] \tag{1}$$

where $\ell(\cdot,\cdot)$ denotes the loss function. This formulation frames the goal of active learning as identifying a subset selection that yields the maximal gain in generalization performance. However, directly solving the optimization problem in Eq.(1) faces two fundamental obstacles:

- **Label Agnosticism and Distribution Inaccessibility:** The true labels y for the candidate subset S are unknown during the selection phase, making the posterior loss uncomputable. The true joint distribution $\mathcal{P}_{XY}$ is unknown; we can only approximate the population using the finite observations in $\mathcal{U}$.
- **Hard-to-Optimize:** The landscape for the objective function in Eq.(1) is highly rugged, rendering it $\mathcal{NP}$-hard. Generally, heuristic and approximation algorithms serve as alternative approaches, yet these methods tend to be time-expensive when pursuing high-quality solutions.

Given these challenges, optimizing the true generalization error directly is challenging. Alternatively, the key for active learning thus lies in constructing a computable Surrogate Objective. If the selected subset S can maximally approximate the empirical distribution of the full pool $\mathcal{U}$ in the feature space (ensuring distribution consistency) while simultaneously retaining the model's most uncertain samples (maximizing information utility), it serves as a robust proxy for the intractable risk minimization. This dual-objective strategy effectively steers the selection towards a highly efficient and high-performance active learning solution.

3.2 Optimal Subset Selection via Moment Matching

Our core premise is to reconstruct the sample selection problem in active learning as a discretized sampling problem of the *Uncertainty Field*. Let $\mathcal{U} = \{x_i\}_{i=1}^N$ be the pool of unlabeled samples, and $u(x)$ be a scoring function quantifying the uncertainty information of a sample. To effectively capture high-value regions on the data manifold, we define the *Target Information Distribution* Q as a normalized probability measure weighted by uncertainty:

$$dQ(x) \propto u(x)dP(x) \tag{2}$$

Although the ultimate goal of active learning is to minimize the generalization error of the model, directly optimizing this objective is often intractable. Therefore, we propose a surrogate strategy: selecting a cardinality-constrained subset $\mathcal{S} \subset \mathcal{U}$ (with $|\mathcal{S}| = B$) such that its weighted discrete empirical distribution $\hat{P}_{\mathcal{S}}$ statistically approximates the target distribution Q as closely as possible. The intuition behind this strategy is that if the sample set can faithfully reproduce the distributional structure of the uncertainty field, a model trained on this set will most effectively cover the ambiguous regions of the decision boundary.

To quantify this distributional consistency, we adopt the concept of *Generalized Moment Matching*. Our optimization objective is to minimize the *Distributional Discrepancy* between the two distributions under a generalized correlation measure $\mathcal{G}(\cdot,\cdot)$:

$$\min_{\mathcal{S}} \mathcal{D}(\mathcal{S}) = \min_{\mathcal{S}} \iint \Big(dQ(x) - d\hat{P}_{\mathcal{S}}(x)\Big) \times \mathcal{G}(x,x') \Big(dQ(x') - d\hat{P}_{\mathcal{S}}(x')\Big) \tag{3}$$

where $\mathcal{G}(x,x')$ is a symmetric positive-definite function describing the pairwise correlation between samples. To solve this problem, we expand the squared term and decompose it into three independent parts:

$$\begin{aligned}\mathcal{D}(\mathcal{S}) = &\underbrace{\iint dQ(x)\mathcal{G}(x,x')dQ(x')}_{\text{Constant}} \\ &- 2\underbrace{\iint dQ(x)\mathcal{G}(x,x')d\hat{P}_{\mathcal{S}}(x')}_{\text{Cross-term}} \\ &+ \underbrace{\iint d\hat{P}_{\mathcal{S}}(x)\mathcal{G}(x,x')d\hat{P}_{\mathcal{S}}(x')}_{\text{Self-term}}\end{aligned} \tag{4}$$

Since the first term depends only on the global target distribution Q and is independent of the choice of subset $\mathcal{S}$, it can be ignored during optimization. We discretize the remaining two terms over the finite dataset $\mathcal{U}$ and subset $\mathcal{S}$. Specifically, we convert integrals to sums, where $dQ(x) \approx \frac{1}{N}\sum_{x\in\mathcal{U}} u(x)\delta_x$ and $d\hat{P}_{\mathcal{S}}(x) = \frac{1}{B}\sum_{s\in\mathcal{S}} u(s)\delta_s$.

By ignoring the constant term and transforming the minimization of discrepancy into the maximization of utility, we finally obtain the following discretized subset selection objective function $\mathcal{J}(\mathcal{S})$:

$$\begin{aligned}\mathcal{J}(\mathcal{S}) = &\underbrace{\frac{2}{BN}\sum_{s\in\mathcal{S}} u(s)\mathcal{V}_{global}(s)}_{\text{Information Coverage}} \\ &\underbrace{- \frac{1}{B^2}\sum_{s\in\mathcal{S}}\sum_{s'\in\mathcal{S}} u(s)u(s')\mathcal{G}(s,s')}_{\text{Structural Redundancy}}\end{aligned} \tag{5}$$

Here, we introduce the auxiliary variable $\mathcal{V}_{global}(s) = \sum_{x\in\mathcal{U}} u(x)\mathcal{G}(s,x)$, representing the *Global Correlation Potential* of sample s relative to the entire dataset. This derivation reveals the intrinsic mechanism of optimal sampling: the algorithm must find a balance between maximizing the correlation of samples with the global uncertainty field and minimizing the pairwise correlation within the sample set. Notably, these coefficients are not heuristic hyperparameters but emerge directly from the expansion of the distribution discrepancy. This implies that our objective effectively balances the global alignment and local redundancy in a parameter-free manner, ensuring consistent distribution matching regardless of the subset size.

The proposed objective $\mathcal{J}(\mathcal{S})$, however, is derived from a variational perspective to match the uncertainty field, its effectiveness fundamentally relies on ensuring that the selected samples physically cover the high-value regions of the data manifold. Theoretically, we establish a rigorous relationship between the statistical distributional consistency and the geometric coverage capabilities of the selected subset. The formal statement is as follows, proving that maximizing the interaction energy (analogous to the information coverage term in our objective) explicitly lifts the lower bound of the geometric coverage ratio, thereby providing a solid theoretical foundation for our energy minimization framework. To operationalize this analysis, we specifically instantiate the generalized correlation measure $\mathcal{G}$ as a kernel function $\mathcal{K}$. Since $\mathcal{G}$ is required to be symmetric and positive-definite to ensure a valid discrepancy metric, reproducing kernels naturally satisfy these geometric constraints and allow for rigorous bound derivation.

Theorem 1 (Distribution Consistency and Coverage Efficiency Bound)**.** *Let $\mathcal{U}$ denote the unlabeled data pool, and $S \subset \mathcal{U}$ be the selected subset. For any covering radius $r > 0$, define the geometric coverage ratio as:*

$$\rho(r) := \frac{1}{|\mathcal{U}|} \left| \{ x \in \mathcal{U} : \mathit{dist}(x, S) \leq r \} \right|. \tag{6}$$

*Assume the interaction between data points is defined by a Gaussian kernel function $\mathcal{K}(\cdot,\cdot)$. Let $\mathcal{E}_{inter}(S,\mathcal{U}) = \frac{1}{N}\sum_{x\in\mathcal{U}} \frac{1}{|S|}\sum_{s\in S}\mathcal{K}(s,x)$ denote the **Average Interaction Energy** between the subset S and the global pool $\mathcal{U}$. Furthermore, define the **Maximum Local Interaction Energy** $\mathcal{E}_{max}(S) = \max_{x\in\mathcal{U}} \frac{1}{|S|}\sum_{s\in S}\mathcal{K}(s,x)$ as the peak potential within the pool. The coverage ratio $\rho(r)$ satisfies the following bounds:*

$$\frac{\mathcal{E}_{inter}(S,\mathcal{U}) - \gamma(r)}{\mathcal{E}_{max}(S) - \gamma(r)} < \rho(r) < \mathcal{E}_{inter}(S,\mathcal{U}) \cdot |S| \cdot \frac{1}{\gamma(r)}, \tag{7}$$

where $\gamma(r)$ is the decay threshold associated with radius r (for a Gaussian kernel, $\gamma(r) = e^{-r^2/(2\sigma^2)}$).

This theorem indicates that a higher degree of distributional approximation (larger $\mathcal{E}_{inter}$) raises the theoretical lower bound of the geometric coverage $\rho(r)$. Crucially, minimizing the local energy peaks (reducing $\mathcal{E}_{max}$) via repulsive mechanisms further tightens this lower bound. Thus, minimizing distributional discrepancy is equivalent to explicitly optimizing the coverage efficiency of the core set on the manifold. Leveraging Theorem 1, we can derive the following upper bound on the global error loss:

Corollary 1 (Energy-based Generalization Error Bound)**.** *Given N i.i.d. samples drawn from P_μ as $\mathcal{U} = \{(x_i, y_i)\}_{i=1}^N$, where $y_i \in [C]$ is the class label for x_i. For any covering radius $r > 0$, let $S \subset \mathcal{U}$ be a selected subset with average interaction energy $\mathcal{E}_{inter}(S,\mathcal{U})$ and maximum local interaction energy $\mathcal{E}_{max}(S)$. For any $\epsilon > 1 - \mathcal{B}_{energy}(r)$, suppose: (i) the loss $l(\cdot, y, w)$ is λ_l-Lipschitz continuous for all y, w and bounded by L; (ii) the class-specific regression function $\eta_c(x) := \mathbb{P}(y = c|x)$ is λ_η-Lipschitz for all $c \in [C]$; and (iii) $l(x, y; h_S) = 0$ for all $(x, y) \in S$. Then, with probability at least $1 - \epsilon$:*

$$\left| \frac{1}{N} \sum_{(x,y)\in\mathcal{U}} l(x, y; h_S) \right| \leq r(\lambda_l + \lambda_\eta LC) + L\sqrt{\frac{\log\left(\frac{\mathcal{B}_{energy}(r)}{\mathcal{B}_{energy}(r)+\epsilon-1}\right)}{2N}}, \tag{8}$$

where $\mathcal{B}_{energy}(r) := \frac{\mathcal{E}_{inter}(S,\mathcal{U}) - \gamma(r)}{\mathcal{E}_{max}(S) - \gamma(r)}$.

The detailed proofs are provided in the supplementary material.

3.3 Physical Realization: Quantum Hamiltonian and Antiferromagnetic System

To find the global optimum within the complex high-dimensional energy landscape, we map the subset selection problem into an interacting spin system in statistical physics. See supplementary material for details of the specific transformation of the Ising model.

From Energy Minimization to Ising Model. We model the selection state as a binary variable $z_i \in \{0, 1\}$. Following the principle of energy minimization derived from distribution matching, the system's Hamiltonian $H(z)$ (ignoring constant terms) is given by:

$$H(z) = \sum_{i,j} \left(\frac{1}{B^2} u_i u_j \mathcal{G}(i,j) \right) z_i z_j - \sum_i \left(\frac{2}{BN} u_i \mathcal{V}_{\text{global}}(i) \right) z_i \tag{9}$$

To reveal the deeper physical mechanisms governing this competition, we transform the binary variables into physical spins $s_i \in \{-1, +1\}$ via the mapping $z_i = (s_i + 1)/2$. To strictly enforce the budget constraint $\sum z_i = B$, we introduce a quadratic penalty term parameterized by λ. In this framework, the cardinality constraint B acts as a Chemical Potential controlling particle density. Expanding terms yields the fully explicit **Ising Hamiltonian**:

$$H(s) = \sum_{i<j} J_{ij} s_i s_j + \sum_i h_i s_i \tag{10}$$

This formulation provides a profound physical interpretation for Active Learning:

1. Geometric Frustration via Antiferromagnetism (J_{ij}): The coupling strength $J_{ij} = \frac{1}{2}\left(\frac{1}{B^2} u_i u_j \mathcal{G}_{ij} + \lambda\right)$ is strictly non-negative, creating an **Information-Weighted Antiferromagnetic Interaction**. Physically, adjacent spins (similar samples) strive to align anti-parallel ($s_i s_j = -1$) to minimize energy. When multiple high-information samples cluster locally, strong antiferromagnetic coupling induces Geometric Frustration, forcing the system to break symmetry and spontaneously form a mutually exclusive lattice, thereby enforcing sparsity.

2. Competition via Effective Field (h_i): The effective field h_i determines the spin orientation of sample i, composed of three competing forces:

$$h_i = \underbrace{\frac{1}{2}\left(-\frac{2}{BN}u_i\mathcal{V}_i\right)}_{\text{Information Attraction}} + \underbrace{\frac{1}{2}\sum_j\left(\frac{1}{B^2}u_iu_j\mathcal{G}_{ij}\right)}_{\text{Redundancy Screening}} + \underbrace{\frac{\lambda}{2}(N-2B)}_{\text{Constraint Bias}} \tag{11}$$

Whether sample i ends up in the "excited state" ($s_i = +1$) depends on whether its **Attraction** (high uncertainty and representativeness) can overcome the **Screening Effect** generated by its neighbors. This screening cloud (the summation term) effectively cancels out the attractiveness of redundant samples in high-density regions, explaining why our global energy minimization is superior to greedy strategies that ignore these collective interactions.

Truncated Potential and Manifold Localization. Finally, to ensure the robustness of the aforementioned physical interactions on high-dimensional data manifolds, we instantiate the general potential $\mathcal{G}$ as a *Truncated Gaussian Potential*:

$$\mathcal{G}(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)\cdot\mathbb{I}(\|x_i - x_j\| \leq \epsilon(x_i)) \tag{12}$$

where $\epsilon(x_i)$ is an adaptive truncation radius based on local density. Introducing truncation serves not only to sparsify computation but also to eliminate *Long-tail Noise* in high-dimensional space. Due to the concentration of measure, distant samples often exhibit spurious weak correlations. The truncation operation enforces Ising interactions to be effective only within local topological neighborhoods, ensuring that the energy minimization process focuses on resolving the *local fine-grained structure* of the manifold. This rugged energy landscape caused by short-range interactions is also the scenario where the Coherent Ising Machine can better exploit its optical bifurcation dynamics compared to greedy.

3.4 Algorithm

We encapsulate the proposed approach into the Distribution-Aware Energy Minimization (DAEM) algorithm. In this section, we outline the classical greedy solution based on submodularity and further discuss the adoption of Coherent Ising Machine for addressing the Ising mapping problem.

Classical Greedy Solver and Submodularity: Submodular maximization has been widely studied, from classical greedy methods for monotone objectives [Nemhauser *et al.*, 1978] to recent continuous approaches for non-monotone DR-submodular maximization [Chen *et al.*, 2023]. Given that the objective function $\mathcal{J}(\mathcal{S})$ exhibits submodularity due to non-positive elements in Hessian matrix, the greedy strategy is an efficient approximation method on classical computers. At each iteration, the algorithm selects the sample x^* that maximizes the marginal gain:

$$x^* = \underset{x\in\mathcal{U}\setminus S_{t-1}}{\arg\max}\ \Delta J(x|S_{t-1}). \tag{13}$$

Although the computational complexity scales quadratically with the pool size, i.e., $O(N^2)$, the greedy algorithm remains a robust and effective choice for our standard tasks.

Coherent Ising Machine via Ising Hamiltonian: We leverage the derivation in Section 3.3 to map the optimization objective to an Ising Hamiltonian $H(s)$. The significance of this mapping lies in transforming the combinatorial optimization problem into a ground-state search of a physical system, thereby paving the way for quantum potentials.

By encoding the problem into a physical energy model, we facilitate the application of the Coherent Ising Machine, which exploits optical bifurcation phenomena and the minimum-gain principle to naturally select modes with minimal loss, allowing for a dynamic evolution toward ground states within complex energy landscapes. This physics-driven optimization offers an efficient way in handling highly non-convex landscapes. While current photonic hardware scale is still evolving, the establishment of this mathematical formulation demonstrates that our method is ready to seamlessly integrate with future optical computing resources, providing a novel computational perspective for solving complex distribution matching problems.

Algorithm Framework: The complete workflow of the DAEM framework is outlined in Algorithm 1. In each iteration, the algorithm computes the uncertainty field and kernel matrix for the global unlabeled pool to construct the distribution-aware optimization objective. Subsequently, depending on the available computational backend (Classical Greedy or Quantum), the objective is optimized to select the optimal subset.

Algorithm 1 DAEM: Distribution-Aware Energy Minimization

Require: Unlabeled pool $\mathcal{U}$, Query budget B, Initial model θ_0
Ensure: Updated model θ_T

```
Initialize labeled set L ← ∅ (or initial set)
for round t = 1 to T do
    Compute uncertainty scores u(x) and feature embeddings φ(x) for all x ∈ U
    Construct Optimization Objective J(S) in Eq. (5)
    Solve Subset Selection:
        S_t ← arg max_{S⊂U,|S|=B} J(S) {Via Greedy Algorithm or Coherent Ising Machine}
    Query labels for S_t: L ← L ∪ S_t, U ← U \ S_t
    Train model θ_t on L
end for
return θ_T
```

4 Experiments

In this section, we conduct a comprehensive evaluation of our **DAEM** algorithm across multiple image classification benchmarks. To ensure the rigor and reproducibility of our results, our experimental design strictly adheres to methodological standards outlined in [Lüth *et al.*, 2023]. The detailed experimental setup can be found in the supplementary material [1].

[1]For specific code and the supplementary material of the paper, please refer to: https://github.com/yzc060202/DAEM.

Method	CIFAR-10		CIFAR-100		Tiny-ImageNet		BloodMNIST		Imb-Tiny	
	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
Entropy	$0.33_{\pm 0.68}$	$1.86_{\pm 0.34}$	$-1.85_{\pm 0.35}$	$0.03_{\pm 0.67}$	$-3.12_{\pm 0.19}$	$-2.35_{\pm 0.13}$	$-16.08_{\pm 4.21}$	$-15.26_{\pm 13.14}$	$-2.03_{\pm 0.26}$	$-1.41_{\pm 0.05}$
Margin	$-0.48_{\pm 1.43}$	$1.64_{\pm 0.85}$	$-1.05_{\pm 0.36}$	$0.29_{\pm 0.61}$	$-1.26_{\pm 0.48}$	$-1.20_{\pm 0.25}$	$-2.50_{\pm 4.31}$	$-0.51_{\pm 2.07}$	$-0.87_{\pm 0.04}$	$-1.07_{\pm 0.48}$
CoreSet	$-15.66_{\pm 1.27}$	$-7.72_{\pm 1.60}$	$0.20_{\pm 0.20}$	$0.28_{\pm 0.13}$	$-5.44_{\pm 0.13}$	$-6.31_{\pm 0.15}$	$0.17_{\pm 2.01}$	$0.56_{\pm 0.54}$	$-3.81_{\pm 0.10}$	$-3.95_{\pm 0.45}$
BADGE	$0.18_{\pm 0.83}$	$1.64_{\pm 0.46}$	$-0.06_{\pm 0.29}$	$0.43_{\pm 0.13}$	$-0.38_{\pm 0.14}$	$-0.75_{\pm 0.25}$	$0.69_{\pm 2.68}$	$1.87_{\pm 0.60}$	$0.33_{\pm 0.11}$	$-0.51_{\pm 0.53}$
wk-means	$-0.13_{\pm 1.35}$	$-0.64_{\pm 0.14}$	$0.17_{\pm 0.31}$	$0.39_{\pm 0.38}$	$0.00_{\pm 0.29}$	$0.11_{\pm 0.18}$	$2.42_{\pm 2.68}$	$0.46_{\pm 0.19}$	$0.35_{\pm 0.30}$	$0.05_{\pm 0.21}$
TypiClust	$2.62_{\pm 0.37}$	$-0.07_{\pm 0.41}$	$0.17_{\pm 0.29}$	$0.06_{\pm 0.24}$	$2.14_{\pm 0.20}$	$1.81_{\pm 0.22}$	$1.76_{\pm 0.49}$	$0.64_{\pm 1.01}$	$2.25_{\pm 0.30}$	$1.08_{\pm 0.73}$
uHerding	$2.70_{\pm 0.71}$	$1.84_{\pm 0.42}$	$-0.28_{\pm 0.24}$	$0.32_{\pm 0.19}$	$1.03_{\pm 0.33}$	$-1.85_{\pm 0.18}$	$-2.38_{\pm 2.38}$	$-0.37_{\pm 1.47}$	$1.00_{\pm 0.09}$	$-1.22_{\pm 0.24}$
DAEM(Ours)	$\mathbf{5.05_{\pm 0.83}}$	$\mathbf{3.65_{\pm 0.58}}$	$\mathbf{0.77_{\pm 0.34}}$	$\mathbf{0.53_{\pm 0.23}}$	$\mathbf{3.59_{\pm 0.41}}$	$\mathbf{2.80_{\pm 0.46}}$	$\mathbf{4.02_{\pm 2.62}}$	$\mathbf{2.70_{\pm 1.20}}$	$\mathbf{3.05_{\pm 0.05}}$	$\mathbf{1.23_{\pm 0.31}}$

Table 1: Experimental results (Randomization-based Relative Test Accuracy %) of different Active Learning methods across five datasets under Small(Low) and Large(High) budgets. The best results are highlighted in **bold**.

4.1 Performance Comparison Across Datasets

Tab.1 summarizes the average accuracy gains of various active learning strategies relative to random sampling across five benchmark datasets, categorized into Small Budget and Large Budget regimes. Quantitative results demonstrate that our DAEM framework achieves the best performance across all experimental settings (highlighted in **bold**), exhibiting superior label efficiency. To provide a more intuitive analysis, we visualize the detailed performance curves on CIFAR-10 compare the computational time costs in Fig.1, Right.

Mitigating Redundancy and Sustaining Improvements. Uncertainty-based strategies are prone to *local redundancy* due to sample clustering. As shown in Tab.1, Entropy and Margin perform poorly in small budget scenarios across multiple datasets (e.g., -3.12% for Entropy on Tiny-ImageNet). Fig.1 (Left) further confirms this; DAEM establishes a significant lead right from the start. This is attributed to our energy minimization-based repulsion mechanism, which enforces orthogonality among selected samples, effectively dispersing sampling points. As the budget increases (in the middle part of Fig.1), many clustering-based methods (e.g., UHerding) suffer from performance saturation due to excessive distribution smoothing. In contrast, DAEM dynamically balances global distribution matching with local uncertainty attraction, maintaining the highest gains in the high-budget regimes of CIFAR-100 and Tiny-ImageNet, demonstrating its adaptability throughout the entire training lifecycle.

Computational Efficiency. Beyond accuracy, DAEM demonstrates exceptional computational efficiency. Fig.1 illustrates the time consumption of various algorithms on CIFAR-10 as the number of samples increases (log scale). While TypiClust and UHerding outperform traditional methods in some performance metrics, their computational costs exhibit a significant upward trend with data scale, potentially becoming a bottleneck. In contrast, DAEM (red line) maintains a time cost in the order of 10^1 seconds with a remarkably flat growth rate. This efficiency is attributed to our optimized energy minimization strategy, which enables the rapid identification of high-performance subsets from large-scale pools without incurring prohibitive computational overheads. These results indicate that DAEM successfully overcomes the trade-off between performance and computational cost, making it highly suitable for large-scale real-world applications.

Additional Results. Detailed curves, setups, ablations, and sensitivity analyses are provided in the supplementary material.

4.2 Quantum vs. Classical: A Comparative Analysis

To investigate the potential of quantum computing in active learning, we conducted a head-to-head comparison between the classical greedy solver and the Coherent Ising Machine (CIM) solver implemented on QBoson's 550-qubit coherent optical quantum computer, CPQC-550, using CIFAR-10 and BloodMNIST. The results are visualized in Fig.2 and analyzed from two perspectives: optimization quality and computational scalability.

1. Optimization Quality: The Trade-off between Noise and Global Search. In Fig.2 (Left), under the low-budget regime ($B = 50$), the classical greedy strategy generally outperforms the quantum solver. For instance, on BloodMNIST, the greedy approach leads by a clear margin. We attribute this to the hardware constraints of current devices, such as the limited precision (e.g., 8-bit) for parameters J_{ij} and h_i. In small-scale combinatorial problems where the energy landscape is relatively simple, the intrinsic 'physical noise' and embedding overhead of the hardware can disrupt the precise identification of the ground state, whereas the deterministic greedy algorithm robustly finds high-quality local optima.

However, a critical transition is observed as combinatorial complexity increases. In the high-budget regime ($B = 100$), as shown in Fig.2 (Right), the quantum solver begins to demonstrate its utility. Notably on CIFAR-10, the quantum trajectory (blue dashed line) intersects and eventually surpasses the classical greedy baseline. This reversal suggests that as the solution space expands exponentially ($\binom{450}{100} \gg \binom{450}{50}$), classical greedy heuristics are more prone to entrapment in sub-optimal local minima. In contrast, the CIM, leveraging search dynamics in continuous phase space, exhibits a stronger potential to escape local optima and approximate the global minimum. This indicates that photonic computing offers a promising avenue for high-complexity regimes where classical heuristics falter.

2. Time-to-Solution: The Shift in Computational Bottleneck. Beyond accuracy, we analyze the *scalability* of the solver to evaluate their future potential. To quantify the computational overhead, we recorded the runtime of the classical greedy solver and the total DAEM algorithm latency across increasing budget sizes. The averaged results on CIFAR-10 are summarized in Tab.2.

Our empirical data reveals a fundamental divergence in how computational costs scale with the query budget B:

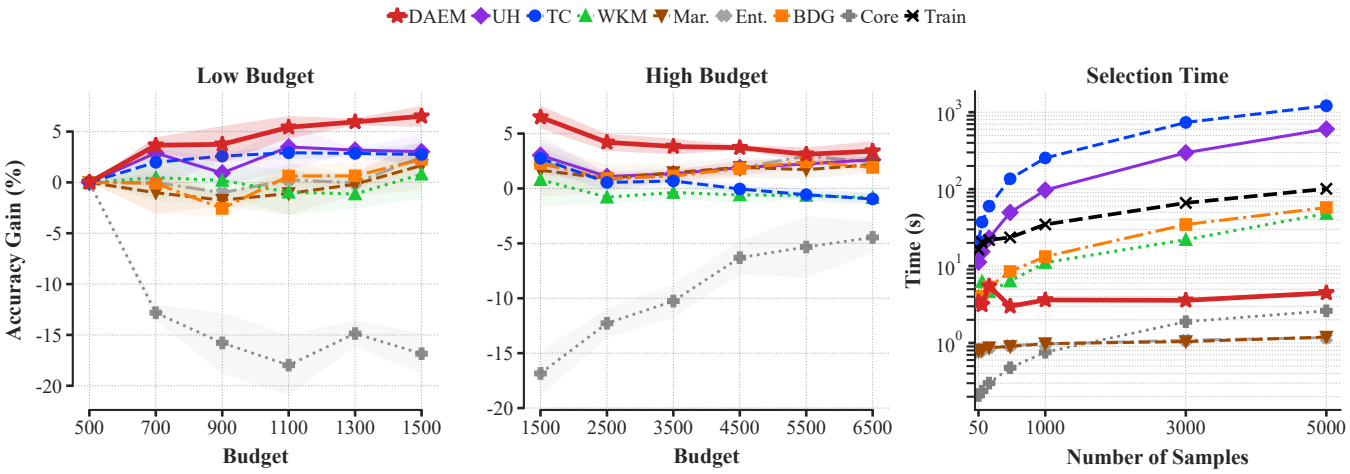

Figure 1: CIFAR-10 performance and efficiency analysis. Left and middle panels compare algorithm results with Random baseline under distinct budget constraints, while the right one shows selection latency and corresponding 100-epoch training expenses.

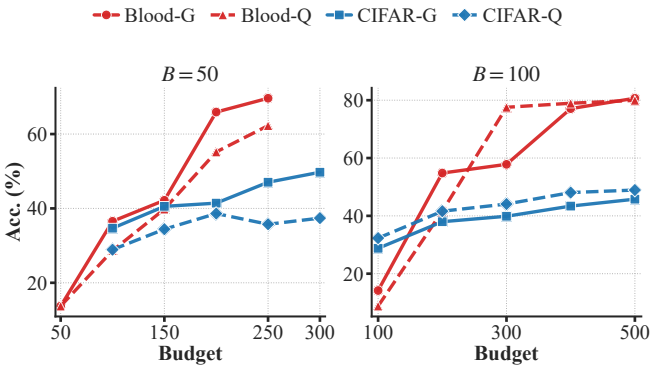

Figure 2: Greedy and quantum DAEM accuracy on BloodMNIST and CIFAR-10 under $B = 50$ and $B = 100$.

- **Classical Scalability Wall:** The computational cost of the classical greedy solver grows significantly with the budget size B, creating a distinct *bottleneck shift* as shown in Tab.2. In the low-budget regime (e.g., $B = 50$), the greedy solver consumes only **0.13s**, accounting for a negligible **3.5%** of the total runtime. Here, the latency is dominated by fixed pre-processing tasks such as kernel construction. However, as the scale expands to $B = 10,000$, the solver latency surges to **4.26s**. Crucially, the solver's contribution to the total runtime escalates dramatically to **54.8%**, overtaking pre-processing as the primary time sink. This trend confirms that for ultra-large-scale real-time applications, the classical optimization step inevitably evolves from a minor overhead into the dominant constraint.
- **CIM Speedup Potentials:** In contrast, the computation time of the CIM exhibits weak dependence on the budget size. The solution latency is governed by the temporal evolution dynamics of optical pulses. In our numerical experiments, the physical evolution times remains constant at approximately **7ms**, which illustrate the considerable speedup potential for ultra-large-scale instances.

Budget (B)	T_{solver} **(s)**	T_{total} **(s)**	**Ratio (%)**
50	0.13(0.00697)	3.69(3.5091)	3.5(0.002)
100	0.15(0.00674)	3.60(3.4385)	4.2(0.002)
1000	0.62	4.10	15.1
10000	4.26	7.77	**54.8**

Table 2: Comparison of computational time costs for the DAEM algorithm under different budgets (B) on CIFAR-10. The table lists solver time (T_{solver}), total time (T_{total}), and the ratio. Note: Values outside parentheses represent the Greedy algorithm, while values inside parentheses represent CIM.

Crucially, the value of DAEM extends beyond immediate runtime comparisons. A **distinguishing feature** of our framework is that its optimization objective is formulated to be **mathematically equivalent to an Ising model**. Unlike standard active learning methods restricted to classical heuristics, this specific mathematical structure allows DAEM to directly leverage quantum computing substrates. Consequently, our framework possesses distinct **'quantum potentials'**: as quantum hardware matures, offering higher qubit counts and lower noise floors, DAEM is inherently positioned to harness these characteristics, promising continuous improvements in both optimization performance and time efficiency that are inaccessible to purely classical algorithms.

5 Conclusion

We propose Distribution-Aware Energy Minimization, a physically inspired active learning framework that reformulates distribution matching as energy minimization. The resulting objective dynamically balances uncertainty and representativeness without hand-crafted parameters. Experiments show that DAEM outperforms prior SOTA methods across multiple budget regimes with higher computational efficiency. We further demonstrate the potential of quantum computing for future large-scale active learning.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants No. 92465202, No. 12501450, and No. 12571382. This work was also sponsored by the CCF-QBoson Quantum Computing Application Innovation Fund (NO. CCF-QBoson202407).

References

[Ash *et al.*, 2019] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*, 2019.

[Bae *et al.*, 2025] Wonho Bae, Danica Sutherland, and Gabriel Oliveira. Uncertainty herding: One active learning method for all label budgets. In *International Conference on Learning Representations*, volume 2025, pages 788–805, 2025.

[Bassman Oftelie *et al.*, 2018] Lindsay Bassman Oftelie, Pankaj Rajak, Rajiv K Kalia, Aiichiro Nakano, Fei Sha, Jifeng Sun, David J Singh, Muratahan Aykol, Patrick Huck, Kristin Persson, et al. Active learning for accelerated design of layered materials. *npj Computational Materials*, 4(1):74, 2018.

[Caramalau *et al.*, 2021] Razvan Caramalau, Binod Bhattarai, and Tae-Kyun Kim. Sequential graph convolutional network for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9583–9592, 2021.

[Cerezo *et al.*, 2021] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

[Chen *et al.*, 2023] Shengminjie Chen, Donglei Du, Wenguo Yang, Dachuan Xu, and Suixiang Gao. Continuous non-monotone dr-submodular maximization with down-closed convex constraint. *arXiv preprint arXiv:2307.09616*, 2023.

[Chen *et al.*, 2024] Liangyu Chen, Yutong Bai, Siyu Huang, Yongyi Lu, Bihan Wen, Alan Yuille, and Zongwei Zhou. Making your first choice: to address cold start problem in medical active learning. In *Medical Imaging with Deep Learning*, pages 496–525. PMLR, 2024.

[Chen *et al.*, 2025] Shengminjie Chen, Ziyang Li, Hongyi Zhou, Jialin Zhang, Wenguo Yang, and Xiaoming Sun. A lyapunov framework for quantum algorithm design in combinatorial optimization with approximation ratio guarantees. *arXiv preprint arXiv:2512.21716*, 2025.

[Cornuejols *et al.*, 1977] Gerard Cornuejols, Marshall L Fisher, and George L Nemhauser. Exceptional paper—location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management science*, 23(8):789–810, 1977.

[Ducoffe and Precioso, 2018] Melanie Ducoffe and Frederic Precioso. Adversarial active learning for deep networks: a margin based approach. *arXiv preprint arXiv:1802.09841*, 2018.

[Dukkipati *et al.*, 2025] Ambedkar Dukkipati, Ranga Shaarad Ayyagari, Bodhisattwa Dasgupta, Parag Dutta, and Prabhas Reddy Onteru. Active reinforcement learning strategies for offline policy improvement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 16418–16425, 2025.

[Farhi *et al.*, 2014] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

[Gal *et al.*, 2017] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International conference on machine learning*, pages 1183–1192. PMLR, 2017.

[Gissin and Shalev-Shwartz, 2019] Daniel Gissin and Shai Shalev-Shwartz. Discriminative active learning. *arXiv preprint arXiv:1907.06347*, 2019.

[Gupte *et al.*, 2024] Sanket Rajan Gupte, Josiah Aklilu, Jeffrey J Nirschl, and Serena Yeung-Levy. Revisiting active learning in the era of vision foundation models. *arXiv preprint arXiv:2401.14555*, 2024.

[Hacohen *et al.*, 2022] Guy Hacohen, Avihu Dekel, and Daphna Weinshall. Active learning on a budget: Opposite strategies suit high and low budgets. *arXiv preprint arXiv:2202.02794*, 2022.

[Haussmann *et al.*, 2020] Elmar Haussmann, Michele Fenzi, Kashyap Chitta, Jan Ivanecky, Hanson Xu, Donna Roy, Akshita Mittel, Nicolas Koumchatzky, Clement Farabet, and Jose M Alvarez. Scalable active learning for object detection. In *2020 IEEE intelligent vehicles symposium (iv)*, pages 1430–1435. IEEE, 2020.

[Inagaki *et al.*, 2016] Takahiro Inagaki, Yoshitaka Haribara, Koji Igarashi, Tomohiro Sonobe, Shuhei Tamate, Toshimori Honjo, Alireza Marandi, Peter L McMahon, Takeshi Umeki, Koji Enbutsu, et al. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.

[Johnson *et al.*, 2011] Mark W Johnson, Mohammad HS Amin, Suzanne Gildert, Trevor Lanting, Firas Hamze, Neil Dickson, Richard Harris, Andrew J Berkley, Jan Johansson, Paul Bunyk, et al. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, 2011.

[Kirsch *et al.*, 2019] Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32, 2019.

[Kochenberger *et al.*, 2014] Gary Kochenberger, Jin-Kao Hao, Fred Glover, Mark Lewis, Zhipeng Lü, Haibo Wang, and Yang Wang. The unconstrained binary quadratic programming problem: a survey. *Journal of combinatorial optimization*, 28(1):58–81, 2014.

[Koh and Liang, 2017] Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions.

In *International conference on machine learning*, pages 1885–1894. PMLR, 2017.

[Kulesza and Taskar, 2012] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2-3):123–286, 2012.

[Lewis and Gale, 1994] David D. Lewis and William A. Gale. A sequential algorithm for training text classifiers. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, page 3–12, Berlin, Heidelberg, 1994. Springer-Verlag.

[Lu *et al.*, 2023] Yiren Lu, Guojing Tian, and Xiaoming Sun. Qaoa with fewer qubits: a coupling framework to solve larger-scale max-cut problem. *arXiv preprint arXiv:2307.15260*, 2023.

[Lucas, 2014] Andrew Lucas. Ising formulations of many np problems. *Frontiers in physics*, 2:74887, 2014.

[Lüth *et al.*, 2023] Carsten Lüth, Till Bungert, Lukas Klein, and Paul Jaeger. Navigating the pitfalls of active learning evaluation: A systematic framework for meaningful performance assessment. *Advances in Neural Information Processing Systems*, 36:9789–9836, 2023.

[McMahon *et al.*, 2016] Peter L McMahon, Alireza Marandi, Yoshitaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsunomiya, Kazuyuki Aihara, et al. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.

[Mukhoti *et al.*, 2023] Jishnu Mukhoti, Andreas Kirsch, Joost Van Amersfoort, Philip HS Torr, and Yarin Gal. Deep deterministic uncertainty: A new simple baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24384–24394, 2023.

[Nemhauser *et al.*, 1978] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 1978.

[Parvaneh *et al.*, 2022] Amin Parvaneh, Ehsan Abbasnejad, Damien Teney, Gholamreza Reza Haffari, Anton Van Den Hengel, and Javen Qinfeng Shi. Active learning by feature mixing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12237–12246, 2022.

[Pinsler *et al.*, 2019] Robert Pinsler, Jonathan Gordon, Eric Nalisnick, and José Miguel Hernández-Lobato. Bayesian batch active learning as sparse subset approximation. *Advances in neural information processing systems*, 32, 2019.

[Scheffer *et al.*, 2001] Tobias Scheffer, Christian Decomain, and Stefan Wrobel. Active hidden markov models for information extraction. In *International symposium on intelligent data analysis*, pages 309–318. Springer, 2001.

[Schmidt *et al.*, 2025] Sebastian Schmidt, Leonard Schenk, Leo Schwinn, and Stephan Günnemann. Joint out-of-distribution filtering and data discovery active learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 25677–25687, 2025.

[Sener and Savarese, 2017] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

[Shui *et al.*, 2020] C. Shui, F. Zhou, C. Gagné, et al. Deep active learning: Unified and principled method for query and training. In *International Conference on Artificial Intelligence and Statistics*, pages 1308–1318. PMLR, 2020.

[Sinha *et al.*, 2019] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5972–5981, 2019.

[Wang *et al.*, 2024] Yifeng Wang, Xueying Zhan, and Siyu Huang. Autoal: Automated active learning with differentiable query strategy search. *arXiv preprint arXiv:2410.13853*, 2024.

[Wei *et al.*, 2015] Kai Wei, Rishabh Iyer, and Jeff Bilmes. Submodularity in data subset selection and active learning. In *International conference on machine learning*, pages 1954–1963. PMLR, 2015.

[Yang *et al.*, 2017] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z Chen. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 399–407. Springer, 2017.

[Yehuda *et al.*, 2022] Ofer Yehuda, Avihu Dekel, Guy Hacohen, and Daphna Weinshall. Active learning through a covering lens. *Advances in Neural Information Processing Systems*, 35:22354–22367, 2022.

[Yoo and Kweon, 2019] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 93–102, 2019.

[Yuan *et al.*, 2023] Jiakang Yuan, Bo Zhang, Xiangchao Yan, Tao Chen, Botian Shi, Yikang Li, and Yu Qiao. Bi3d: Bi-domain active learning for cross-domain 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15599–15608, 2023.