CoRe-Diffusion: Bridging the Generative-Discriminative Gap in Dataset Distillation via Manifold-Aligned Contrastive Guidance

Heng Shu1,2, Qiming Yang1, Linjuan Cheng1, and Yuquan Wu1*

1 Institute of Software, Chinese Academy of Sciences, Beijing, China
{shuheng2025,yangqiming,linjuan,yuquan}@iscas.ac.cn
2 University of Chinese Academy of Sciences, Beijing, China

Abstract. Dataset Distillation aims to condense large-scale datasets into a tiny but highly informative subset, enabling efficient training while achieving performance comparable to the full dataset. To overcome the scalability bottlenecks of traditional optimization, Generative Dataset Distillation (GDD) leverages diffusion priors to reformulate the dataset condensation process as an efficient generative sampling paradigm. However, existing GDD paradigms primarily optimize intra-class likelihood while largely overlooking explicit inter-class separation, which may cause generated features to concentrate near decision boundaries. Furthermore, relying on manually designed guidance targets to steer synthesis often pushes the diffusion trajectory off the natural data manifold, inducing severe structural artifacts. To address these issues, CoRe-Diffusion is proposed as a unified framework. Specifically, it introduces Contrastive Negative Guidance, which utilizes inter-class mode centers—computed but largely unused during synthesis by existing methods—to exert zero-overhead repulsive gradients against hard negatives. To preserve synthesis fidelity, Manifold-Aligned Real-Anchor Discovery strictly constrains the guidance targets to exact real-image latents, preventing the trajectory from drifting off the empirical latent manifold. This intrinsically preserves complex semantic structures, bypassing the prohibitive computational overhead of relying on external generative priors or auxiliary modules. Alongside an Annealed Sampling Schedule for smooth trajectory evolution, CoRe-Diffusion achieves state-of-the-art performance, yielding a 3.8% absolute accuracy improvement on ImageNet-1K while introducing negligible computational overhead.

Keywords: Generative Dataset Distillation Contrastive Negative Guidance Manifold Alignment Diffusion Models

* Corresponding authors: Qiming Yang and Yuquan Wu.

1 Introduction

The rapid development of deep learning has been driven by the explosive growth of data scale, yet this progress comes at the cost of prohibitive storage overhead and training inefficiency. To mitigate these challenges, Dataset Distillation (DD) [1] has emerged as a pivotal paradigm. It aims to condense a massive training dataset into a minimal synthetic set, ensuring that a neural network trained on this compact substitute achieves performance comparable to one trained on the full dataset.

Early DD methods predominantly relied on bi-level optimization frameworks, such as Gradient Matching (GM) [2], Distribution Matching (DM) [3], and Trajectory Matching (TM) [4]. A key characteristic of these traditional methods is their inherently *multi-class perspective* : by approximating gradients across the entire dataset during each optimization epoch, they implicitly capture rich inter-class dynamics. However, calculating derivatives of pixels with respect to gradients requires unrolling complete computation graphs (Back-Propagation Through Time, BPTT). This leads to an unsustainable $O(M^2)$ memory explosion and computational intractability, limiting their application to high-resolution, large-scale datasets like ImageNet-1K.

To overcome this computational bottleneck, the field has recently witnessed a paradigm shift toward Generative Dataset Distillation (GDD). By leveraging the deep generative priors of pre-trained diffusion models (e.g., DiT [5] and Stable Diffusion [6]), GDD recasts the $O(M^2)$ optimization problem into a highly efficient latent sampling process and substantially reduces both generation time and peak memory usage. Nonetheless, this efficiency gain is partially undermined in practice: many contemporary GDD variants introduce heavyweight external mechanisms that add substantial computational and temporal overhead. For instance, prior-tuning methods like MinimaxDiffusion [7] require expensive adversarial fine-tuning of the generative model itself. Gradient-guidance methods like DD-IGD [8] depend on pre-trained surrogate models to compute trajectory gradients, while VLCP [9] relies heavily on massive Vision-Language Models (VLMs) to generate extensive text prototypes for semantic injection. These additions significantly compromise the fundamental efficiency advantage of GDD. In contrast, training-free mode-guided methods like MGD^3 [10] offer a much faster alternative by steering the latent trajectory using unsupervised mode clustering, yet they suffer from a conceptual limitation.

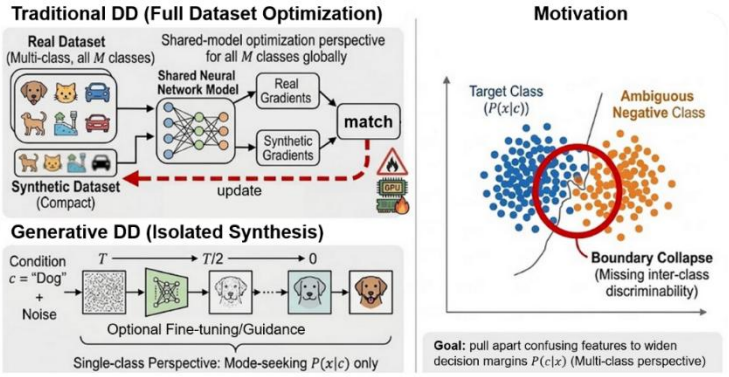

Fig. 1. Motivation of CoRe-Diffusion. Left: Current Generative DD methods focus on isolated single-class synthesis, leading to a generative-discriminative mismatch and severe feature collapse near decision boundaries. Right: CoRe-Diffusion reintroduces a multi-class perspective. By applying Contrastive Negative Guidance (a repulsive guidance signal against hard negatives), CoRe-Diffusion explicitly enforces inter-class discriminability and widens decision margins without parameter fine-tuning.

Current GDD paradigms—including MGD³—suffer from a fundamental *generative-discriminative objective mismatch*. While traditional DD captures global inter-class information, GDD strictly focuses on maximizing the single-class data likelihood P(x|c). It answers "what to generate" (intra-class representativeness) but fails to explicitly model "what to differentiate from" (inter-class discriminability). Downstream classifiers, however, require optimizing the decision boundary P(c|x). By omitting information from other classes, current GDD methods cause severe feature collapse near decision boundaries, frequently synthesizing ambiguous samples (e.g., generating a dog with fox-like ears), as illustrated in Fig. 1.

To address these issues, we propose CoRe-Diffusion (illustrated in Fig. 2), a strictly training-free framework. To bridge the generative-discriminative gap, we introduce **Contrastive Negative Guidance**. By repurposing inter-class mode centers, which are computed but not used during synthesis in existing methods, we introduce a repulsive guidance term against hard negatives. Since mode discovery is already required in the pipeline, this explicit repulsive force effectively widens decision boundaries and achieves discriminative gains without introducing additional function evaluations. As visually evidenced in Fig. 3, this mechanism successfully separates highly entangled classes and helps alleviate mode collapse during generation.

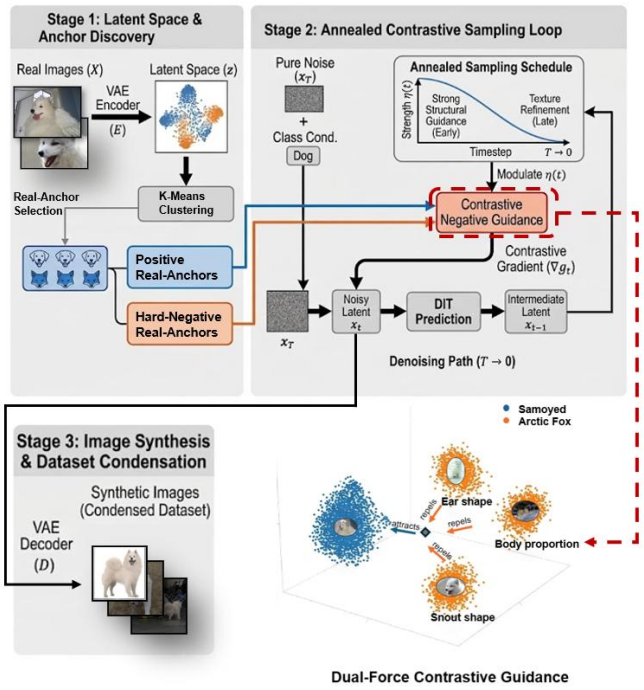

Fig. 2. The overall pipeline of CoRe-Diffusion. Our training-free framework first explicitly maps the data into the latent space to discover Manifold-Aligned Real-Anchors (Stage 1). During the diffusion sampling loop, it applies Contrastive Negative Guidance modulated by an Annealed Sampling Schedule to actively push trajectories away from confusing negative modes (Stage 2), ultimately decoding them into highly discriminative synthetic datasets (Stage 3).

However, explicitly steering trajectories using vanilla arithmetic averages (e.g., K-Means centroids) introduces a critical topological flaw. As illustrated in Fig. 3, forcing latent codes toward cluster centers distorts the natural feature distribution, collapsing samples toward local means and disrupting the rich topological structure of the underlying data manifold. To rectify this issue, we propose **Manifold-Aligned Real-Anchor Discovery**. By anchoring guidance targets strictly to the nearest empirical real-sample latents, our method constrains the generation process to remain on the natural data manifold, preserving the intrinsic feature distribution and complex semantic structures while avoiding the prohibitive computational costs of VLM-based semantic injection. Furthermore, to mitigate the severe numerical stiffness caused by the "hard truncation" strategies in prior works, we employ an **Annealed Sampling Schedule**. This is consistent with the denoising dynamics of the Ordinary Differential Equation (ODE): strong contrastive constraints are applied during the early semantic formation phase to repel confounding classes, and control is smoothly released back to the unconditional diffusion prior during the late texture-filling phase for high-fidelity detail refinement.

Our contributions are summarized as follows:

We identify the critical generative-discriminative mismatch in current GDD and are among the first to introduce *Contrastive Negative Guidance*. This explicit repulsive force successfully widens the

minimum inter-class margin and achieves the highest feature discriminability (e.g., Silhouette Score) among existing generative paradigms.

We reveal the topological collapse and off-manifold distortion issues caused by guiding diffusion with vanilla K-Means centroids. To rectify this, we propose *Manifold-Aligned Real-Anchor Discovery* coupled with an *Annealed Sampling Schedule*. This provides rigorous on-manifold structural constraints and eliminates pseudo-discontinuity artifacts, mathematically aligning with the natural ODE evolution.

We achieve a superior trade-off between synthesis accuracy and computational efficiency. Compared to heavy VLM-based methods like VLCP, CoRe-Diffusion yields significantly more discriminative latent features and achieves comprehensive state-of-the-art results, notably yielding a +3.8% absolute accuracy improvement on the full ImageNet-1K benchmark with negligible additional overhead.

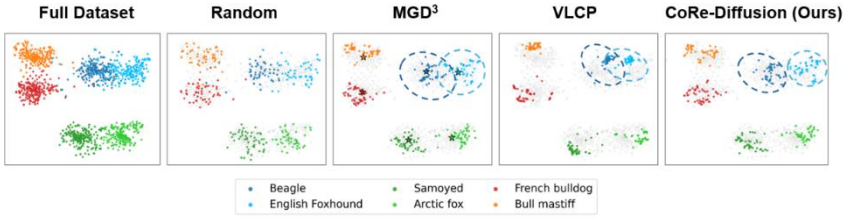

Fig. 3. Feature Distribution Analysis (t-SNE) on highly entangled classes.

2 Related Work

2.1 Optimization-based Dataset Distillation

Early optimization-based dataset distillation mimics full-dataset training dynamics via pixel-space bi-level optimization [1]. To avoid explicit unrolling, subsequent methods align surrogate objectives like gradients [2], trajectories [4], or feature distributions [3]. Despite successes on small benchmarks, they suffer from $O(M^2)$ memory explosion on large-scale datasets (e.g., ImageNet-1K). Recent decoupling [11, 12] and parameter-tuning [13] strategies improve scalability, yet synthesized samples remain highly sensitive to initialization and hyperparameters, motivating a paradigm shift toward generative priors.

2.2 Generative Dataset Distillation

To bypass optimization bottlenecks, Generative Dataset Distillation (GDD) leverages the deep semantic priors of pre-trained diffusion models [14]. Current GDD forms two main streams. **1) Prior-tuning methods** adapt the generative model via fine-tuning using adversarial objectives [7], disentangled latents [15], skip-connections [16], or vision-language prototypes [9]. Related explorations also include distilling diffusion models into faster architectures like GANs [17], applying progressive distillation for rapid sampling [18], and extending generative latents to federated learning paradigms [19]. However, iterative fine-tuning is computationally prohibitive and risks prior degradation. **2) Sampling-guidance methods**, conversely, keep diffusion weights frozen and explicitly steer the inference

trajectory via trajectory influence [8], mode discovery [10], optimal transport [20], or by taming diffusion to enhance representativeness priors [21, 22]. Recent efforts also rectify generative evaluation inconsistencies [23]. Despite these strides, current sampling strategies predominantly rely on positive, prototype-seeking guidance, largely neglecting the inter-class boundaries essential for discriminative tasks [24].

2.3 Contrastive and Negative Guidance

Guidance mechanisms like Classifier-Free Guidance [25] are central to controllable generation, frequently utilizing *negative prompting* to avoid unwanted attributes in text-to-image synthesis [6]. While explicitly contrasting negatives is foundational for learning well-separated decision boundaries in general representation learning [26, 27], its potential to enhance *inter-class discriminability* within dataset distillation remains severely underexplored. Existing GDD paradigms exclusively perform positive mode-seeking, inadvertently causing feature collapse near semantic boundaries.

3 Methodology

In this section, we present **CoRe-Diffusion** (Contrastive Real-guided Diffusion), a training-free generative dataset distillation framework. As discussed in Section 1, the core challenge of dataset distillation lies not only in compressing the data volume but also in preserving the complex semantic boundaries necessary for downstream classification.

3.1 Problem Definition

The limitations of current Generative Dataset Distillation (GDD) paradigms stem from two fundamental issues: the mismatch between generative and discriminative objectives, and topological distortion during latent guidance.

Generative–Discriminative Mismatch. Let $D=(x_i,y_i)(i=1)^N$ be a real dataset. GDD synthesizes a condensed set S using a pre-trained diffusion model by approximating the conditional distribution P(x|c) through the score function:

$$s_\theta(x_t, c, t) \approx \nabla_{x_t} \log P(x_t|c). \tag{1}$$

This objective maximizes intra-class likelihood, describing "what constitutes class c". However, classification relies on maximizing the posterior P(c|x). By Bayes' rule,

$$\log \frac{P(c|x)}{P(c_{neg}|x)} = \log P(x|c) - \log P(x|c_{neg}) + \log \frac{P(c)}{P(c_{neg})}. \tag{2}$$

Existing GDD methods optimize only the positive term P(x|c) while entirely ignoring the repulsive term - P(x|c(neg)). As a result, synthesized features lack explicit inter-class separation and tend to collapse near decision boundaries.

Topological Distortion from Off-Manifold Guidance. To enhance semantic control, some methods construct guidance anchors by clustering latent representations $Z(c)=z_1,\ldots,z_n$:

$$\mu_c = \frac{1}{n}\sum_{j=1}^{n} z_j. \tag{3}$$

However, since the natural image manifold M embedded in the latent space Z is highly non-convex, such arithmetic averaging often produces deviations from the natural feature distribution:

$$\mu_c \notin \mathcal{M} \quad \text{in general.} \tag{4}$$

Guiding diffusion trajectories toward these off-manifold targets introduces out-of-distribution states, leading to structural artifacts and semantic distortions in synthesized images.

To address these issues, we reformulate diffusion inference as a boundary-aware, manifold-constrained process. CoRe-Diffusion operates entirely in the latent space and consists of three stages: (1) Manifold-Aligned Real-Anchor Discovery, (2) Contrastive Negative Guidance, and (3) an Annealed Sampling Schedule for stable trajectory evolution. The entire pipeline avoids backpropagation through the diffusion network, maintaining minimal memory overhead.

3.2 Manifold-Aligned Real-Anchor Discovery

Mode-guided frameworks have demonstrated that discovering multiple intra-class modes can significantly enhance the diversity of distilled datasets. These discovered modes serve as semantic anchors to guide the generation process. However, existing methods ubiquitously rely on standard K-Means clustering to calculate these anchors. This introduces a fundamental topological flaw. The latent space of diffusion models is highly non-linear; consequently, the vanilla arithmetic mean of multiple distinct latent vectors (i.e., a K-Means centroid) often disrupts the natural distribution in the feature space. Guiding the generative trajectory toward these off-manifold regions inevitably corrupts the local geometry, leading to synthetic images with severe structural distortions.

To enforce strict adherence to the intrinsic high-dimensional distribution, we propose *Real-Anchor Discovery*. Instead of discarding K-Means, we leverage its efficiency but correct its topological flaw by grounding the vanilla centroids to actual empirical data. Given a set of real-image VAE latents Z(c) = z_1, z_2, ..., z(N) for class c, where each z_j is encoded by the frozen VAE associated with the diffusion model, we first apply K-Means directly in this VAE latent space to obtain K vanilla cluster centroids μ(c,1), ..., μ(c,K). Then, for each centroid μ(c,i), we identify its nearest empirical neighbor in the latent space to serve as the true mode anchor m(c,i):

$$m_{c,i} = z_{j^*}, \quad j^* = \arg\min_{j \in \{1,\dots,N\}} \| z_j - \mu_{c,i} \|_2. \tag{5}$$

We define the set of these positive Real-Anchors as M(c) = m(c,1), ..., m(c,K). Because these anchors are encoded from real training samples rather than arithmetic centroids, they lie on the empirical data manifold, ensuring that the guided trajectory remains structurally sound.

3.3 Contrastive Negative Guidance

While Real-Anchors guarantee intra-class fidelity, they do not resolve the gap between generation and discrimination. Standard diffusion sampling is conditioned exclusively on the target class, lacking explicit instructions on what features to *avoid*. This bias often results in boundary collapse, where fine-grained categories share overlapping visual traits. We propose rectifying this by explicitly injecting a contrastive repulsive field into the diffusion process.

To provide informative discriminative gradients without introducing random noise, we systematically mine the *Hardest Negative Class* for each target category. Let mbar(c) = (1)/(K) (i=1)^K m(c,i)

denote the global semantic center of the discovered Real-Anchors for class c. Its semantic rival c(neg) is identified by minimizing the inter-class Euclidean distance:

$$c_{neg} = \arg\min_{k \in \mathcal{C}, k \neq c} \| \bar{m}_c - \bar{m}_k \|_2. \tag{6}$$

We adopt a class-level negative assignment because the negative guidance is designed as a stable inter-class boundary regularizer rather than a mode-specific attraction target. In contrast, the positive guidance uses mode-level anchors to preserve intra-class diversity, while the negative class mainly provides a reference-level repulsive direction for separating the target category from its most confusing semantic rival.

Once c(neg) is identified, generating a repulsive force directly from its global center is suboptimal due to the high intra-class variance. To compensate for local mode-level ambiguity within the selected negative class, we introduce *Dynamic Mode Alignment*. During the synthesis targeting a specific positive anchor m(c,i), we calculate its distance to all available modes within the negative class M(c)(neg). We dynamically select the top-k nearest negative modes and compute their localized mean to form an aligned negative prototype m^(-)(c,i):

$$m_{c,i}^{-} = \frac{1}{k} \sum_{j \in \mathcal{N}_k(m_{c,i})} m_{c_{neg},j}, \tag{7}$$

where N_k(m(c,i)) represents the indices of the k nearest modes in the selected negative class (we set k=5). This top-k aggregation provides a localized, robust repulsive field, allowing the globally selected hard negative class to adapt to the mode-level geometry of each positive anchor.

At each denoising timestep t, using the Tweedie estimate of the clean data xhat(0,t), we compute a contrastive guidance score g_t that concurrently pulls the latent toward the positive anchor and pushes it away from the negative prototype:

$$g_t = \underbrace{-\lambda_{pos}(t)(\hat{x}_{0,t} - m_{c,i})}_{\text{Positive Attraction}} + \underbrace{\lambda_{neg}(t)(\hat{x}_{0,t} - m_{c,i}^{-})}_{\text{Negative Repulsion}}. \tag{8}$$

This score dynamically modifies the predicted noise, effectively steering the ODE trajectory to enlarge the safety margin between semantic clusters.

3.4 Annealed Sampling Schedule

Previous guidance-based methods often employ a rigid "hard-truncation" strategy—abruptly terminating guidance at an intermediate step (e.g., t=25). Such truncation introduces severe numerical stiffness into the vector field of the denoising ODE, disrupting the natural evolution of the latent trajectory.

To ensure mathematical continuity, we propose an *Annealed Sampling Schedule* using a cosine decay function η(t) over the denoising timesteps t ∈ [T, 0]:

$$\eta(t) = \frac{1}{2}\left(1 + \cos\left(\pi\left(1 - \frac{t}{T}\right)\right)\right). \tag{9}$$

This dynamic factor smoothly modulates the guidance scales rather than changing the diffusion noise schedule: $\lambda_{pos}(t) = \lambda_{pos} \cdot \eta(t)$ and λ(neg)(t) = λ(neg) · η(t). This strictly aligns with the physics of diffusion: strong contrastive constraints are applied during the early semantic formation phase (t ≈ T) to

repel confounding classes, while control is seamlessly released back to the unconditional generative prior as $t \rightarrow 0$, ensuring artifact-free texture refinement without pseudo-discontinuity artifacts.

4 Experiments

4.1 Datasets

We evaluate our method on large-scale benchmarks and specific subsets. For large-scale evaluation, we utilize the full ImageNet-1K [28] and ImageNet-100 (a 100-class subset). To thoroughly investigate synthesis capabilities across diverse distributions, we conduct extensive experiments on three high-resolution (256 x 256) 10-class subsets: ImageNette [29], ImageWoof [29], and ImageIDC [30]. Specifically, ImageNette consists of visually distinct objects, serving as a standard baseline. ImageWoof is a highly challenging fine-grained subset comprising 10 visually similar dog breeds, making it an ideal testbed for evaluating our contrastive negative guidance in pushing apart entangled decision boundaries. ImageIDC, derived from the first 10 classes of ImageNet-100, exhibits significant intra-class diversity.

4.2 Implementation Details

Synthesis Settings. We employ the pre-trained DiT-XL/2 as our base generative model, synthesizing images at a 256 x 256 resolution with 50 denoising steps. For Manifold-Aligned Real-Anchor Discovery, real training images are first encoded into the VAE latent space of the frozen diffusion model. Following standard K-Means clustering in this latent space, we identify the nearest empirical VAE latent neighbor for each vanilla centroid to serve as the authentic mode anchor (Real-Anchor). The initial positive (λ_{pos}) and contrastive negative (λ(neg)) guidance scales are set to 0.1 and 0.01, respectively.

Evaluation on Subsets. For ImageNette, ImageWoof, ImageIDC, and ImageNet-100, we strictly follow the standard evaluation protocol established by MinimaxDiffusion [7]. Distilled images are used to train randomly initialized networks (ConvNet-6, ResNet-18, and ResNetAP-10 with instance normalization) from scratch. We utilize identical data augmentations, optimizers, and training schedules to ensure fair comparisons, reporting the average Top-1 accuracy over multiple trials.

Evaluation on ImageNet-1K. For the full ImageNet-1K dataset, we adopt the rigorous RDED [31] evaluation protocol. To guarantee a scalable and fair assessment, we employ the exact soft-label supervision and optimization hyperparameters prescribed by RDED across deeper architectures, including ResNet-18, ResNet-50, and ResNet-101.

4.3 Comparison with the SOTA Methods

To comprehensively evaluate the effectiveness of CoRe-Diffusion, we compare it against a diverse set of representative state-of-the-art (SOTA) methods. These baselines span various dataset distillation paradigms, including: (1) traditional optimization-based methods, such as Distribution Matching (DM) [3], SRe^{2}L [11], G-VBSM [12], RDED [31], and EDC [13]; (2) parameterization-efficient optimization methods like IDC [30]; and (3) cutting-edge generative prior-based methods (GDD), including MinimaxDiffusion [7], D^{4}M [15], VLCP [9], and MGD^3 [10]. Furthermore, we benchmark against standard

data selection strategies such as Random selection and Herding [32], alongside the vanilla generative prior DiT [5]. By comparing against these established approaches across different synthesis subsets and architectures, we systematically demonstrate our performance advantages.

Table 1. Comparison of state-of-the-art methods on ImageWoof (256x256). Bold indicates the best result and underline indicates the second-best. For reference, training on the full dataset yields upper-bound performances of 86.4±0.2, 87.5±0.5, and 89.3±1.2 for ConvNet-6, ResNetAP-10, and ResNet-18, respectively.

IPC (Ratio)	Test Model	Random	DiT	IDC-1	Minimax	VLCP	MGD³	Ours
10 (0.8%)	ConvNet-6	24.3±1.1	34.2±1.1	33.3±1.1	33.3±1.7	**37.0±1.0**	34.7±1.1	35.6±0.8
	ResNetAP-10	29.4±0.8	34.7±0.5	39.1±0.5	36.2±3.2	39.2±1.3	40.4±1.9	**40.8±1.0**
	ResNet-18	27.7±0.9	34.7±0.4	37.3±0.2	35.7±1.6	37.6±0.9	38.5±2.5	**39.2±1.1**
20 (1.6%)	ConvNet-6	29.1±0.7	36.1±0.8	35.5±0.8	37.3±0.1	37.6±0.2	39.0±3.5	**40.0±1.2**
	ResNetAP-10	32.7±0.4	41.1±0.8	43.4±0.3	43.3±2.7	**45.8±0.5**	43.6±1.6	44.3±0.9
	ResNet-18	29.7±0.5	40.5±0.5	38.6±0.2	41.8±1.9	42.5±0.6	41.9±2.1	**45.2±1.0**
50 (3.8%)	ConvNet-6	41.3±0.6	46.5±0.8	43.9±1.2	50.9±0.8	53.9±0.6	54.5±1.6	**54.6±1.1**
	ResNetAP-10	47.2±1.3	49.3±0.2	48.3±1.0	53.9±0.7	56.3±1.0	56.5±1.9	**57.8±0.8**
	ResNet-18	47.9±1.8	50.1±0.5	48.3±0.8	53.7±0.6	57.1±0.6	58.3±1.4	**59.7±0.9**
70 (5.4%)	ConvNet-6	46.3±0.6	50.1±1.2	48.9±0.7	51.3±0.6	55.7±0.9	55.1±2.5	**57.5±1.1**
	ResNetAP-10	50.8±0.6	54.3±0.9	52.8±1.8	57.0±0.2	58.3±0.2	60.2±2.4	**61.4±1.0**
	ResNet-18	52.1±1.0	51.5±1.0	51.1±1.7	56.5±0.8	58.8±0.7	59.7±2.7	**61.4±1.2**
100 (7.7%)	ConvNet-6	52.2±0.4	53.4±0.3	53.2±0.9	57.8±0.9	61.1±0.7	60.1±1.2	**62.9±1.3**
	ResNetAP-10	59.4±1.0	61.7±0.9	56.4±0.8	62.7±1.4	64.5±0.2	66.5±1.0	**68.2±1.4**
	ResNet-18	61.5±1.3	59.3±0.7	60.2±1.0	62.7±0.4	65.7±0.4	68.8±0.7	**69.2±0.6**

ImageWoof. Table 1 compares performance on ImageWoof. Our method achieves notable performance gains over previous SOTA methods. For instance, evaluated on ResNet-18, our approach outperforms the second-best results by 2.7%, 1.4%, and 1.7% at IPC 20, 50, and 70, respectively. Similar improvements are consistently observed at higher IPCs and across different architectures, making our method the best-performing approach in the vast majority of scenarios. Although VLCP occasionally gains a slight edge at low IPCs (e.g., on ConvNet-6 at IPC 10) through complex semantic information injection, its significant computational overhead for model fine-tuning and prototype generation cannot be ignored. Since ImageWoof is widely recognized as a hard-to-classify subset, our consistent performance advantage strongly demonstrates that the incorporation of negative guidance in our method effectively enhances the discriminative ability of the synthesized data.

Table 2. Comparison of the state-of-the-art methods on ImageNette and ImageIDC under various IPC settings. All the results are obtained on ResNetAP-10. The best results are marked as bold, and the second-best are underlined.

	IPC	Random	DiT	DM	Minimax	D⁴M	VLCP	MGD³	Ours
Nette	10	54.2±1.6	59.1±0.7	60.8±0.6	57.7±1.2	60.9±1.7	64.8±3.6	66.2±2.4	**68.9±0.9**
	20	63.5±0.5	64.8±1.2	66.5±1.1	64.7±0.8	66.3±1.3	71.4±0.5	71.2±0.5	**71.8±0.7**
	50	76.1±1.1	73.3±0.9	76.2±0.4	73.9±0.3	77.7±1.1	81.2±0.8	79.5±1.3	**81.5±0.7**
IDC	10	48.1±0.8	54.1±0.4	52.8±0.5	51.9±1.4	50.3±1.0	57.0±1.4	55.9±2.1	**57.4±1.0**
	20	52.5±0.9	58.9±0.2	58.5±0.4	59.1±3.7	55.8±0.2	**63.3±1.2**	61.9±0.9	62.0±0.9
	50	68.1±0.7	64.3±0.6	69.1±0.8	69.4±1.4	69.1±2.4	71.9±0.4	72.1±0.8	**75.4±0.8**

ImageNette and ImageIDC. As shown in Table 2, we compare our method against previous SOTA baselines on the ImageNette and ImageIDC datasets using the ResNetAP-10 architecture. On the ImageNette dataset, our approach consistently surpasses prior methods, outperforming MGD³ by 2.7%, 0.6%, and 2.0% at IPC 10, 20, and 50, respectively. Furthermore, even when compared to VLCP-a method characterized by its high computational demands and lower efficiency-our approach still secures a clear performance advantage in the vast majority of cases. Similarly, on the ImageIDC dataset, our method establishes new SOTA results for IPC 10 and 50 with improvements of 1.5% and 3.3% over MGD³, while maintaining highly competitive performance at IPC 20. Overall, these results highlight the effectiveness and robustness of our proposed approach across various IPC settings.

Table 3. Performance comparison on ImageNet-100. Bold indicates the best result.

IPC (Ratio)	Test Model	Random	Herding	IDC-1	Minimax	MGD³	Ours
10 (0.8%)	ConvNet-6	17.0±0.3	17.2±0.3	**24.3±0.5**	22.3±0.5	23.4±0.9	24.0±0.3
	ResNetAP-10	19.1±0.4	19.8±0.3	25.7±0.1	24.8±0.2	25.8±0.5	**27.6±0.5**
	ResNet-18	17.5±0.5	16.1±0.2	25.1±0.2	22.5±0.3	23.6±0.4	**25.2±0.6**
20 (1.6%)	ConvNet-6	24.8±0.2	24.3±0.4	28.8±0.3	29.3±0.4	30.6±0.4	**32.4±0.2**
	ResNetAP-10	26.7±0.5	27.6±0.1	29.9±0.2	32.3±0.1	33.9±1.1	**35.1±0.6**
	ResNet-18	25.5±0.3	24.7±0.1	30.2±0.2	31.2±0.1	32.6±0.4	**34.2±0.2**

ImageNet-100. Table 3 presents the performance comparison on ImageNet-100 under IPC 10 and 20 settings across various target architectures. For IPC 20, our method consistently surpasses the strongest baseline, MGD³, by absolute margins of 1.8%, 1.2%, and 1.6% on ConvNet-6, ResNetAP-10, and ResNet-18, respectively. Under the highly constrained IPC 10 setting, our approach achieves the best performance on both ResNetAP-10 and ResNet-18, while remaining highly competitive on ConvNet-6, closely following IDC-1. Notably, our method is substantially more computationally efficient compared to approaches like IDC-1 and MinimaxDiffusion.

Table 4. Performance comparison on ImageNet-1K under the IPC=50 setting. The best results are marked as bold.

Model	SRe2L	G-VBSM	RDED	EDC	D⁴M	VLCP	MGD³	DD-IGD	Ours
ResNet-18	46.8±0.2	51.8±0.4	56.5±0.1	58.0±0.2	55.2±0.1	60.5±0.2	60.2±0.1	60.3±0.4	**64.0±0.1**
ResNet-50	55.6±0.3	58.7±0.3	-	64.3±0.2	62.4±0.1	-	64.6±0.4	-	**65.5±0.3**
ResNet-101	60.8±0.5	61.0±0.4	61.2±0.4	64.9±0.2	63.4±0.1	-	67.7±0.4	66.8±0.2	**69.9±0.4**

ImageNet-1K. We further evaluate our method on the large-scale ImageNet-1K dataset, as shown in Table 4. Our approach demonstrates strong scalability and achieves the best results on all evaluated architectures. Specifically, we outperform our primary baseline (MGD³) by 3.8% on ResNet-18, 0.9% on ResNet-50, and 2.2% on ResNet-101. To address recent gradient-guidance baselines, we also include the available official Minimax-IGD results reported by DD-IGD [8] in Table 4. CoRe-Diffusion surpasses Minimax-IGD by 3.7% on ResNet-18 and 3.1% on ResNet-101, further demonstrating the effectiveness of our training-free contrastive guidance on large-scale benchmarks.

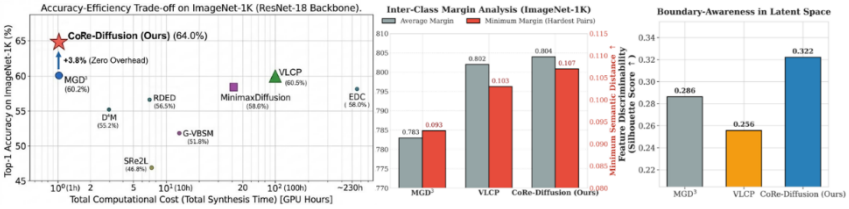

Fig. 4. Comprehensive Performance and Feature Analysis. Left: Accuracy-Efficiency Trade-off on ImageNet-1K (ResNet-18). Middle & Right: Quantitative validation of Boundary-Awareness, demonstrating significantly enlarged inter-class margins (middle) and superior latent feature discriminability via the highest Silhouette Score (right).

Efficiency and Quantitative Analysis. Besides Top-1 accuracy, we evaluate feature discriminability using average inter-class margin, minimum inter-class margin, and Silhouette Score, and analyze practical efficiency in terms of generation cost and network evaluation overhead. As comprehensively illustrated in Fig. 4, CoRe-Diffusion achieves a superior trade-off between synthesis accuracy and computational cost (Left). Unlike tuning-dependent frameworks (e.g., MinimaxDiffusion, VLCP) that incur prohibitive optimization burdens, our strictly training-free method leverages frozen diffusion priors. Furthermore, compared to the MGD³ baseline, our enhancements—repurposing mode clusters for negative guidance and applying an Annealed Sampling Schedule—introduce zero additional network evaluations. Consequently, we match MGD³'s rapid generation speed while delivering substantially higher accuracy. To validate this discriminative gain, we quantitatively analyze the latent space (Middle and Right). By explicitly repelling hard negatives, CoRe-Diffusion substantially enlarges both average and minimum inter-class margins, and attains the highest Silhouette Score compared to existing paradigms. This indicates tighter intra-class clustering and clearer inter-class separation, demonstrating superior generative-discriminative alignment.

4.4 Ablation Experiments

Table 5. Ablation study on ImageNette. Each variant removes one component from the full CoRe-Diffusion model while keeping the remaining settings unchanged.

Model	CoRe-Diffusion	w/o Annealed Schedule	w/o Negative Guidance	w/o Real Anchor	Large λ(neg) (0.1)	Single Neg. Mode
ConvNet	**60.2±1.8**	58.9±1.3	58.4±1.6	59.5±2.0	56.3±1.2	31.4±0.6
ResNetAP	**68.9±0.9**	67.6±1.8	65.9±2.0	67.1±2.4	63.7±0.6	36.2±0.7
ResNet	**65.6±2.2**	64.8±1.4	63.8±1.8	64.6±1.7	60.8±1.2	34.4±0.6

Table 5 shows that the complete CoRe-Diffusion framework consistently outperforms all ablated variants. Notably, the exclusion of Contrastive Negative Guidance yields the most significant performance decline, confirming that repulsive forces from hard negatives are indispensable for preventing boundary collapse and enforcing inter-class separability. Furthermore, both replacing Real-Anchors with arithmetic centroids and removing the Annealed Sampling Schedule lead to marked degradation, validating that strict adherence to the data manifold and smooth trajectory evolution are prerequisites for high-fidelity synthesis.

We further analyze the sensitivity of the negative guidance mechanism. In our empirical trials, we observe that $\lambda(\text{neg})$ works best when it is kept approximately one order of magnitude smaller than the positive guidance scale λ_{pos}. This is because negative guidance is intended to serve as a weak boundary regularizer rather than a dominant sampling objective. When the negative scale becomes excessively large ($\lambda(\text{neg}) = 0.1$), accuracy degrades across all architectures (e.g., dropping to 63.7% on ResNetAP), as over-repulsion overwhelms the positive attraction, forces trajectories off the natural image manifold, and distorts target semantics. More drastically, replacing our Dynamic Mode Alignment (top-k aggregation) with a single negative mode triggers a severe performance collapse (to 36.2% on ResNetAP). A single negative mode introduces erratic, instance-specific noise, deflecting the trajectory via arbitrary local features rather than shared confounding semantics. This firmly validates our top-k averaging strategy, which ensures a stable, noise-filtered repulsive field against ambiguous boundaries.

4.5 Visualization

Fig. 5. Qualitative comparison among VLCP (Left), MGD³ (Middle), and CoRe-Diffusion (Right).

Fig. 5 compares synthesized samples across three challenging scenarios. First, for **Inter-class Discriminability** (Col 1), MGD³ suffers from boundary collapse (e.g., generating short-legged Foxhounds), whereas our Contrastive Negative Guidance ensures unambiguous, class-specific proportions. Second, for **Structural Integrity** (Col 2), MGD³ yields distorted tubing on French horns due to off-manifold drift. In contrast, our Real-Anchor Discovery strictly preserves complex geometric connectivity. Finally, for **Detail Refinement** (Col 3), MGD³'s rigid truncation produces blurry textures and denoising artifacts (e.g., red pigment separating from the nose), while our Annealed Schedule smoothly releases ODE control to render delicate, artifact-free details. Furthermore, VLCP (left block) adopts a fundamentally different, resource-heavy paradigm by packing multiple instances into a single image. While this Mixup-like strategy injects dense semantics to boost classification, it frequently yields severe visual anomalies and conflated features.

5 Discussion and Limitations

CoRe-Diffusion is primarily designed for high-resolution natural image dataset distillation, especially ImageNet-scale GDD scenarios. Following common large-scale protocols, we evaluate IPC settings such as IPC=10, 20, and 50, where IPC=10 already represents a highly compact low-data regime for high-resolution ImageNet-style distillation and enables direct comparison with existing GDD baselines. More extreme IPC=1/5 settings are valuable but may become highly sensitive to initialization, soft-label supervision, and downstream training schedules, and are therefore left as future work. Similarly, cross-domain generalization, such as transferring from natural images to medical images, is an important direction but also strongly depends on whether the base diffusion prior covers the target-domain distribution. Without a medical-domain diffusion prior or explicit domain adaptation, such evaluation may entangle the effect of dataset distillation with the domain gap of the generative prior itself. Regarding backbone compatibility, CoRe-Diffusion does not modify or fine-tune DiT-specific parameters; it only requires a diffusion sampler, a clean-data estimate, and empirical real-image anchors in the corresponding latent space. Therefore, the proposed guidance principle is potentially compatible with other diffusion backbones, while this paper adopts DiT-XL/2 to ensure a controlled and fair comparison with recent ImageNet-scale GDD methods. In practical deployment, CoRe-Diffusion remains lightweight because it is training-free, requires no VLM-based prototype generation, no surrogate-gradient backpropagation, and no additional diffusion network evaluations beyond the original sampling process. Its extra memory cost mainly comes from storing a small number of real anchors, which scales linearly with the number of classes and modes and is negligible compared with the frozen diffusion inference cost.

6 Conclusion

In this paper, we introduce CoRe-Diffusion, a strictly training-free framework that resolves the generative-discriminative mismatch and off-manifold distortion bottlenecks in GDD. Via Contrastive Negative Guidance, our method widens decision margins by repelling trajectories from confounding hard negatives. Concurrently, Manifold-Aligned Real-Anchor Discovery and an Annealed Sampling Schedule ground the synthesis to empirical data distributions, eliminating structural artifacts and preserving fine high-frequency details. Extensive evaluations demonstrate that CoRe-Diffusion establishes a new state-of-the-art across diverse benchmarks—notably achieving a +3.8% improvement on ImageNet-

1K—while incurring negligible additional computational overhead. Injecting boundary-aware contrastive priors and an annealed sampling schedule into diffusion sampling provides a highly scalable and efficient paradigm for future dataset condensation, with promising extensions to ultra-low IPC regimes, cross-domain distillation, and broader diffusion backbones.

7 References

[1] Tongzhou Wang et al.: Dataset Distillation. CoRR (2018)

[2] Bo Zhao et al.: Dataset Condensation with Gradient Matching. ICLR (2021)

[3] Bo Zhao and Hakan Bilen: Dataset Condensation with Distribution Matching. WACV (2023)

[4] George Cazenavette et al.: Dataset Distillation by Matching Training Trajectories. CVPR Workshops (2022)

[5] William Peebles and Saining Xie: Scalable Diffusion Models with Transformers. ICCV (2023)

[6] Robin Rombach et al.: High-Resolution Image Synthesis with Latent Diffusion Models. CVPR (2022)

[7] Jianyang Gu et al.: Efficient Dataset Distillation via Minimax Diffusion. CVPR (2024)

[8] Mingyang Chen et al.: Influence-Guided Diffusion for Dataset Distillation. ICLR (2025)

[9] Yawen Zou et al.: Dataset Distillation via Vision-Language Category Prototype. CoRR (2025)

[10] Jeffrey A. Chan-Santiago et al.: MGD3: Mode-Guided Dataset Distillation using Diffusion Models. ICML (2025)

[11] Zeyuan Yin et al.: Squeeze, Recover and Relabel: Dataset Condensation at ImageNet Scale From A New Perspective. NeurIPS (2023)

[12] Shitong Shao et al.: Generalized Large-Scale Data Condensation via Various Backbone and Statistical Matching. CVPR (2024)

[13] Shitong Shao et al.: Elucidating the Design Space of Dataset Condensation. NeurIPS (2024)

[14] Duo Su et al.: Generative Dataset Distillation Based on Diffusion Model. ECCV Workshops (2024)

[15] Duo Su et al.: D⁴M: Dataset Distillation via Disentangled Diffusion Model. CVPR (2024)

[16] Brian Bernhard Moser et al.: Unlocking Dataset Distillation with Diffusion Models. NeurIPS (2025)

[17] Minguk Kang et al.: Distilling Diffusion Models Into Conditional GANs. ECCV (2024)

[18] Tim Salimans and Jonathan Ho: Progressive Distillation for Fast Sampling of Diffusion Models. ICLR (2022)

[19] Yuqi Jia et al.: Unlocking the Potential of Federated Learning: The Symphony of Dataset Distillation via Deep Generative Latents. ECCV (2024)

[20] Xiao Cui et al.: Optimizing Distributional Geometry Alignment with Optimal Transport for Generative Dataset Distillation. NeurIPS (2025)

[21] Duo Su et al.: Diffusion Models as Dataset Distillation Priors. CoRR (2025)

[22] Lin Zhao et al.: Taming Diffusion for Dataset Distillation with High Representativeness. ICML (2025)

[23] Haoxuan Wang et al.: CaO_2: Rectifying Inconsistencies in Diffusion-Based Dataset Distillation. ICCV (2025)

[24] Xin Zhang et al.: Breaking Class Barriers: Efficient Dataset Distillation via Inter-Class Feature Compensator. ICLR (2025)

[25] Jonathan Ho and Tim Salimans: Classifier-Free Diffusion Guidance. NeurIPS Workshops (2021)

[26] Ting Chen et al.: A Simple Framework for Contrastive Learning of Visual Representations. ICML (2020)

[27] Kaiming He et al.: Momentum Contrast for Unsupervised Visual Representation Learning. CVPR (2020)

[28] Olga Russakovsky et al.: ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. (2015)

[29] Jeremy Howard: Imagenette and Imagewoof. fast.ai (2019)
[30] Jang-Hyun Kim et al.: Dataset Condensation via Efficient Synthetic-Data Parameterization. ICML (2022)
[31] Peng Sun et al.: On the Diversity and Realism of Distilled Dataset: An Efficient Dataset Distillation Paradigm. CVPR (2024)
[32] Max Welling: Herding dynamical weights to learn. ICML (2009)