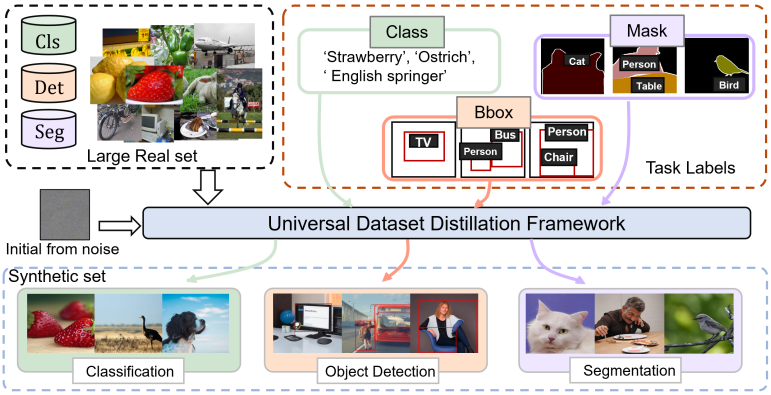
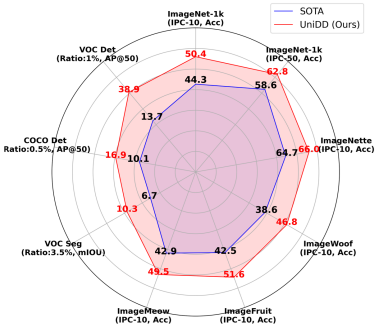
Towards Universal Dataset Distillation via Task-Driven Diffusion

Ding Qi[1*] Jian Li[2†] Junyao Gao[1] Shuguang Dou[1] Ying Tai[5]
Jianlong Hu[2] Bo Zhao[4] Yabiao Wang[3,2†] Chengjie Wang[4,2] Cairong Zhao[1†]
[1]Tongji University [2]Tencent YouTu Lab [3]Zhejiang University
[4]Shanghai Jiao Tong University [5]Nanjing University



(a) One Universal Dataset Distillation Framework for Mutli-Tasks



(b) Improved Performance by UniDD

Figure 1. **(a)** We introduce the Universal Dataset Distillation Framework (UniDD), a novel dataset distillation framework supporting a range of vision tasks, including classification, object detection, and segmentation. UniDD incorporates diverse label types, such as class labels, bounding box coordinates, and pixel-level masks, facilitating the synthesis of task-specific datasets. **(b)** UniDD improves the performance of dataset distillation across diverse tasks on ImageNet [8], Pascal VOC [11, 12], and MS COCO [28].

Abstract

Dataset distillation (DD) condenses key information from large-scale datasets into smaller synthetic datasets, reducing storage and computational costs for training networks. However, most recent research has primarily focused on image classification tasks, with limited exploration in detection and segmentation. Two key challenges remain: (i) Task Optimization Heterogeneity, where existing methods focus on class-level information but fail to address the diverse needs of detection and segmentation, and (ii) Inflexible Image Generation, where current generation methods rely on global updates for single-class targets and lack localized optimization for specific object regions. To address these challenges, we propose UniDD, a universal dataset distillation framework built on a task-driven diffusion model for diverse DD tasks, as shown in Fig. 1. Our approach operates in two stages: Universal Task Knowledge Mining, which captures task-relevant information through task-specific proxy model training, and Universal Task-Driven Diffusion, where these proxies guide the diffusion process to generate task-specific synthetic images. Extensive experiments across ImageNet-1K, Pascal VOC, and MS COCO demonstrate that UniDD consistently outperforms state-of-the-art methods. In particular, on ImageNet-1K with IPC-10, UniDD surpasses previous diffusion-based methods by 6.1%, while also reducing deployment costs.

1. Introduction

Dataset distillation (DD) [44], as an innovative data optimization technique in machine learning [13, 14, 20, 22, 45], condenses essential information from large datasets into smaller proxy sets. Training on these highly ensemble proxy sets enables the model to use much fewer resources while providing comparable performance. Prior DD methods excelled in image classification using various techniques like Gradient Matching [26, 48, 50], Distribution Matching [49, 51], Training Trajectory Matching [3, 7], and Decoupled Methods [42, 46].

*Work done when Ding Qi is an intern in Tencent YouTu Lab, and the advisor is Jian Li.

† Corresponding author

Despite advancements in DD, research on key visual tasks—detection and segmentation—remains limited. We identify the following key challenges: (i) **Task Optimization Heterogeneity.** Existing DD methods are primarily focused on image classification, optimizing distillation by compressing class-level information, but they fall short in addressing the diverse needs of tasks like detection [27] and segmentation, such as object localization and pixel-level information. (ii) **Inflexible Image Generation.** Current image generation methods rely on global updates for single-class targets and lack localized optimization for specific object regions. These limitations become especially evident in images with multiple objects or complex structures, hindering the ability of existing DD to generate distilled data that meets the requirements of diverse tasks. Consequently, the applicability of DD in the visual domain remains constrained.

To address these limitations, we propose a Universal Dataset Distillation (UniDD) framework that optimizes for multi-task learning, including classification, detection, and segmentation, overcoming the shortcomings in task adaptation and image generation. Our approach consists of two stages: **Universal Task Knowledge Mining** and **Universal Task-Driven Diffusion**. In the first stage, we introduce Task-Specific Proxy ($\mathcal{T}_{\mathcal{SP}}$) models, trained on real datasets to extract task-specific information. This allows the $\mathcal{T}_{\mathcal{SP}}$ models to capture and store category information for classification, bounding box positions for detection, and mask data for segmentation within their parameters. Additionally, previous research has often overlooked the importance of contextual information, particularly in high-resolution datasets like ImageNet [8], Pascal VOC [11, 12] and MS COCO [28], where rich object-context relationships exist. To address this, we introduce Task-Aligned Contextual Prompting, which uses a Vision-Language model [33] to generate contextual prompts for real images. In the second stage, we introduce a Task-Driven Diffusion for image synthesis. Specifically, the $\mathcal{T}_{\mathcal{SP}}$ models trained in the first stage supervise the denoising process of the diffusion model by calculating the task loss on the predicted clean images. This guides the diffusion model in generating images that match the given labels. Notably, since the $\mathcal{T}_{\mathcal{SP}}$ models already align with the target domain, there is no need to retrain the diffusion model on the real set, significantly reducing the cost in practical applications.

In summary, our contributions are threefold:

- We identify key challenges in extending dataset distillation beyond classification and propose the first universal framework for DD across classification, detection, and segmentation tasks.
- We develop a two-stage approach, Universal Task Knowledge Mining and Universal Task-Driven Diffusion for Data Synthesis, which flexibly synthesizes images for multi-tasks. Additionally, we address the previously overlooked need to align potential contextual information within datasets.
- Our method achieves state-of-the-art performance in classification tasks, on the ImageNet-1K dataset, our approach surpasses the previous state-of-the-art method by over 6.1%. Additionally, we establish new benchmarks for detection and segmentation on VOC and COCO, showing significant improvements over coreset methods.

2. Related works

2.1. Dataset Distillation

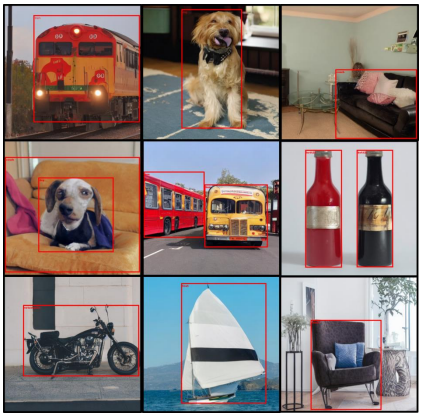
Dataset distillation [44], a dataset reduction method, addresses the challenge of handling substantial data by synthesizing a smaller, representative dataset. Existing dataset distillation methods can be categorized into meta-learning and data matching frameworks based on whether they explicitly mimic the performance of target data. In the meta-learning framework, the distilled data are treated as hyperparameters and optimized in a nested loop fashion according to the trained model's risk. For example, KIP [32] performs kernel ridge regression using the Neural Tangent Kernel (NTK) [25]. The data matching framework includes gradient matching, distribution matching, and trajectory matching. These methods update distilled data by imitating the influence of target data. For example, DC [50] and DM [49] enforce synthetic and real images to have similar gradients or distribution at every training step. MTT [3] proposes to imitate long-range trajectories of optimization steps for real images, thereby better mimicking the training dynamics of real images for synthetic images. However, the aforementioned gradient matching methods require a large number of iterations. To address this, DREAM [31] and IDC [26] focus on more efficient dataset distillation training. Due to the unsatisfactory performance of DM, CAFE [43] and IDM [51] propose improved distribution matching between the real and synthetic data. For the trajectory matching method, DATM [19] scales trajectory matching-based methods to larger synthetic datasets, achieving lossless dataset distillation.

Despite significant progress in previous work, there have been few breakthroughs in dataset distillation for detection and segmentation tasks. In this work, we introduce the first universal framework that simultaneously handles classification, object detection, and semantic segmentation, significantly expanding the generality and scope of dataset distillation research.

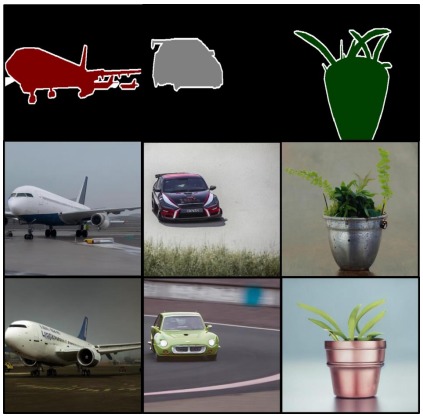
2.2. Diffusion Models

Diffusion models [15, 16, 37, 47] are generative models that show superior fidelity and diversity compared to GANs [17]. It aims to learn the mapping from a Gaussian

(a) Synthetic images for classification.



(b) Synthetic images for detection.



(c) Synthetic images for segmentation.

Figure 2. Visualization of synthetic images generated by UniDD. (a) For classification, images are synthesized on the ImageWoof [8] dataset. (b) For detection, images are synthesized on the Pascal VOC dataset with specified bounding box coordinates. (c) For segmentation, images are synthesized on the Pascal VOC [11, 12] dataset with pixel-level masks. Our UniDD is the first approach to enable high-realism data distillation across classification, detection, and segmentation tasks.

distribution to the real dataset distribution by adding noise in the diffusion process and reversing the noise to the raw image in the denoising process. In the forward process, diffusion models progressively add noise to a clean sample z_0, resulting in a noisy sample z_t:

$$z_t = \sqrt{\alpha_t} z_0 + \sqrt{1-\alpha_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I). \tag{1}$$

In the reverse process, a denoising network ϵ_θ estimates the noise and iteratively recovers the clean sample z_0:

$$\epsilon_\theta(z_t, t) \approx \frac{z_t - \sqrt{\alpha_t} z_0}{\sqrt{1-\alpha_t}}. \tag{2}$$

DDIM [40] provides the following formula for predicting a clean data sample:

$$\hat{z}_0 = \frac{z_t - \sqrt{1-\alpha_t}\epsilon_\theta(z_t, t)}{\sqrt{\alpha_t}}. \tag{3}$$

This formula predicts the clean data point $\hat{z}_0$ from the noisy sample z_t, enabling deterministic sampling for efficient and high-quality image synthesis.

To introduce additional conditions into the diffusion model, classifier guidance [9] has been proposed to train an additional classifier on the noise to guide the model to generate images that are closer to the specified condition. Additionally, [23] introduces classifier-free guidance that trains both an unconditioned and conditioned model simultaneously, achieving more efficient guidance. Following these, recent methods [1, 10, 30] have been proposed to generate high-quality images from diverse conditions.

3. Universal Dataset Distillation Framework

3.1. Problem Definition

Universal Dataset Distillation. Current dataset distillation research primarily targets classification tasks and lacks a generalizable approach for detection and segmentation. To address this, we propose a universal dataset distillation standard that constructs a compact synthetic dataset $\mathcal{S} = (\hat{x_1}, \hat{y_1}), \ldots, (\hat{x}_{|\mathcal{S}|}, \hat{y}_{|\mathcal{S}|})$ from a larger real dataset $\mathcal{T} = (x_1, y_1), \ldots, (x_{|\mathcal{T}|}, y_{|\mathcal{T}|})$, with $|S| \ll |\mathcal{T}|$. The primary difference between tasks lies in label structure: classification uses single-class labels, detection includes bounding boxes with class information, and segmentation requires pixel-level masks. Our goal is for models trained on the distilled dataset $\mathcal{S}$ to perform comparably to those trained on the full dataset $\mathcal{T}$ across classification, detection, and segmentation tasks, as formalized by:

$$\mathbb{E}_{(x,y)\sim P_D} \left| \ell(\theta_{\mathcal{T}}^{task}(x), y) - \ell(\theta_{\mathcal{S}}^{task}(x), y) \right| \leq \epsilon, \tag{4}$$

where P_D denotes the distribution of test data, ℓ represents the performance metric specific to each task, $task \in \{\text{class}, \text{detect}, \text{segment}\}$, and ϵ is a small tolerance value. For each task, the parameter set $\theta_{\mathcal{T}}^{task}$ on the real dataset $\mathcal{T}$ is optimized as follows:

$$\theta_{\mathcal{T}}^{task} = \arg\min_{\theta} \mathbb{E}_{(x,y)\in\mathcal{T}} \left[\ell\left(\phi_\theta^{task}(x), y\right) \right], \tag{5}$$

Similarly, $\theta_{\mathcal{S}}^{task}$ is optimized using the synthetic dataset $\mathcal{S}$.

3.2. Universal Task Knowledge Mining

Task-Specific Proxy Training. Building a universal DD framework is challenging due to the diverse optimization objectives in classification, detection, and segmentation tasks. Recent decoupled DD studies have reached a consensus [34, 42, 46]: a fully trained network can capture and retain essential dataset information. Inspired by this, we extend to general tasks, training a classifier, detector, and segmenter on task-specific datasets to extract critical information—class labels for classification, bounding boxes for

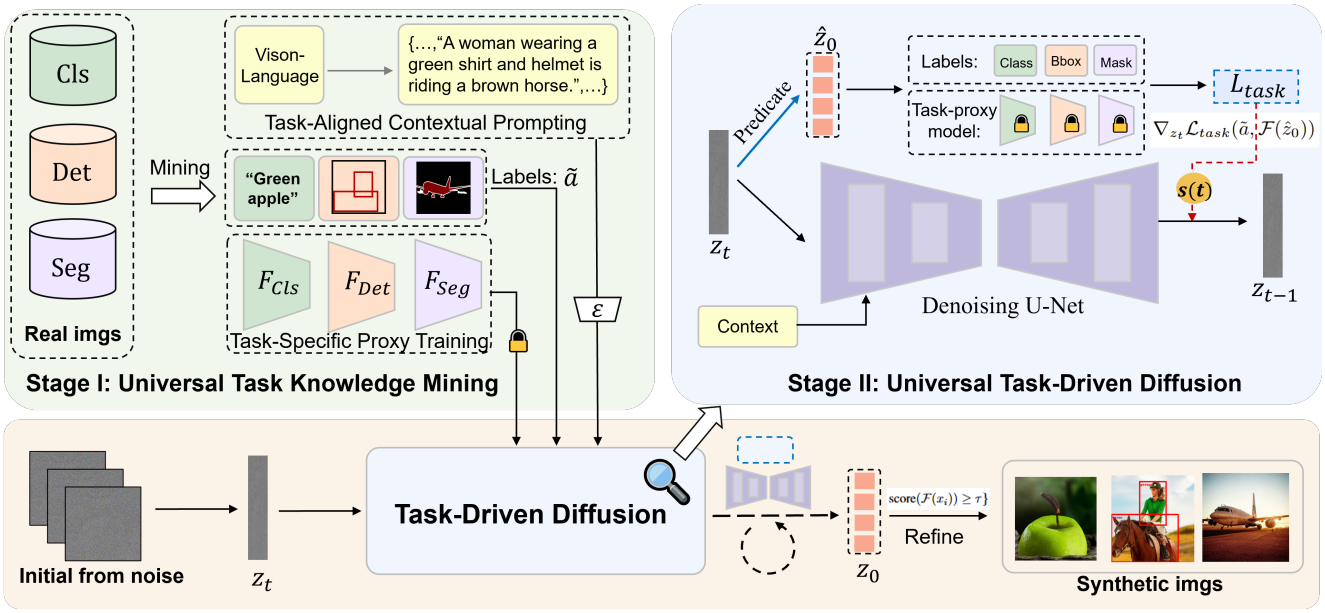


Figure 3. **Universal Dataset Distillation Framework.** This framework consists of two stages: in the Universal Task Knowledge Mining stage, we introduce an easily deployable task-specific proxy ($\mathcal{T}_{\mathcal{SP}}$) model trained on the target dataset and a vision-language model to extract task-aligned context. In the Universal Task-Driven Diffusion stage, we apply the gradient of the supervision loss, derived from the $\mathcal{T}_{\mathcal{SP}}$ model's prediction of the clean image $\hat{z}_0$, at each sampling step.

detection, and pixel-level masks for segmentation. We call this Task-Specific Proxy ($\mathcal{T}_{\mathcal{SP}}$) Model, defined as:

$$\theta = \arg\min_{\theta} \mathcal{L}_{task}(\mathcal{F}(x), y), \tag{6}$$

where $\mathcal{F}$ represents the $\mathcal{T}_{\mathcal{SP}}$ model trained on the target data, θ denotes the model's parameters, $\mathcal{L}_{task}$ represents the loss function specific to the task.

The $\mathcal{T}_{\mathcal{SP}}$ model choice is flexible, allowing adaptation for optimal performance on target datasets. For example, ResNet-18 [21] can serve as a lightweight, fast-deploy option for simple classification, while Faster R-CNN [36] may be used for complex detection tasks to capture key detection information. Unlike previous diffusion-based dataset distillation methods, our approach relies on task-driven learning to align synthetic and real datasets, avoiding costly diffusion model fine-tuning on the target dataset and significantly reducing deployment overhead.

Task-Aligned Contextual Prompting. Previous diffusion-based dataset distillation has largely focused on classification tasks and overlooked the role of natural language in guiding image generation. For diverse tasks, textual prompts must go beyond describing the primary objects in images. As inputs to Stable Diffusion models, prompts can also guide the generation process. To address this, we use a vision-language model to extract natural language descriptions and design task-specific questions to generate task-aligned prompts, a method we call Task-Aligned Contextual Prompting. Following GPT-4o [33], we create tailored prompts for different tasks. For classification, prompts describe object classes, while for detection and segmentation, they also include object relationships and the interaction between objects and the background, providing richer contextual guidance.

3.3. Universal Task-Driven Diffusion.

Inspired by recent advancements in guidance diffusion techniques [1, 9], we propose a Task-Driven Diffusion approach for Dataset Distillation. As illustrated in Fig. 3, during each sampling step $S(z_t, t)$ of the diffusion process, the trained $\mathcal{T}_{\mathcal{SP}}$ model $\mathcal{F}$ provides task-specific guidance. Given an label $\tilde{a}$, the original noise prediction $\epsilon_\theta(z_t, t, \tilde{c})$ is adjusted using the output of $\mathcal{F}$. The adjusted noise prediction $\hat{\epsilon}$ is formulated as follows:

$$\hat{\epsilon}_\theta(z_t, t, \tilde{c})) = \epsilon_\theta(z_t, t, \tilde{c}) + \sqrt{1-\alpha_t}\nabla_{z_t}\mathcal{L}_{\text{task}}(\tilde{a}, \mathcal{F}(z_t)), \tag{7}$$

where $\mathcal{L}_{\text{task}}$ represents the task-specific loss function and $\tilde{a} = \{a_{cl}, a_{bbox}, a_{mask}\}$ may denote a class label a_{cl}, a bounding box a_{bbox}, or a pixel-level segmentation mask a_{mask}. This adjustment mechanism effectively guides the diffusion process to produce data closely aligned with the intended distribution for each task. $\tilde{c}$ is a textual prompt.

Because $\mathcal{F}$ is trained on clean images, it may not perform well directly on noisy inputs. Therefore, we first use Equation 3 to estimate the clean image $\hat{z}_0$. Once $\hat{z}_0$ is obtained, we apply $\mathcal{F}$ to compute the task-driven loss as follows:

$$\hat{\epsilon}_\theta(z_t, t, \tilde{c})) = \epsilon_\theta(z_t, t, \tilde{c}) + s(t) \cdot \nabla_{z_t}\mathcal{L}_{task}(\tilde{a}, \mathcal{F}(\hat{z}_0)), \tag{8}$$

where $s(t)$ is the task-driven strength at each sampling step,

defined as:

$$s(t) = \lambda_s \cdot \sqrt{1 - \alpha_t}, \tag{9}$$

with λ_s being a constant coefficient controlling the overall intensity of the task-driven. The term $\nabla_{z_t}\mathcal{L}_{task}(\tilde{a}, \mathcal{F}(\hat{z}_0))$ represents the gradient of the loss function computed based on the $\mathcal{T}_{\mathcal{SP}}$ model's prediction of the clean image $\hat{z}_0$. Specifically, this gradient is computed as:

$$\nabla_{z_t}\mathcal{L}_{task}(\tilde{a}, \mathcal{F}(\hat{z}_0)) = \nabla_{z_t}\mathcal{L}_{task}\left(\tilde{a}, \mathcal{F}\left(\frac{z_t - \sqrt{1-\alpha_t}\epsilon_\theta}{\sqrt{\alpha_t}}\right)\right). \tag{10}$$

Unlike previous approaches [18, 41] that required finetuning the diffusion model, such as those involving SD [37], to align with new data domains, our method leverages a $\mathcal{T}_{\mathcal{SP}}$ model trained specifically on the target dataset. This model encapsulates critical information about the target task, effectively driving the diffusion process to generate data that is well-aligned with the target task without the need for additional finetuning. This not only conserves time and resources but also mitigates the risk of overfitting the diffusion model to the specific task.

Proxy-Driven High-Realism Refinement. To ensure the high realism and accuracy of the generated images, we employ a trained $\mathcal{T}_{\mathcal{SP}}$ model, to score each image. By setting a well-defined threshold, we filter out low-quality images based on their scores. The remaining images are then relabeled using the same $\mathcal{T}_{\mathcal{SP}}$ model. This combined process of filtering and relabeling can be expressed as:

$$\{x, y\} = \{x_i, \mathcal{F}_{\text{label}}(x_i) \mid \text{score}(\mathcal{F}(x_i)) \geq \tau\} \tag{11}$$

where $\mathcal{F}(x_i)$ represents the model's output score for image x_i, τ is the quality threshold, and $\mathcal{F}_{\text{label}}$ generates the accurate label y_i for the filtered image. This ensures that both the generated images and their labels are of high quality and closely match the desired data distribution.

Algorithm 1 Universal Dataset Distillation Approach

Input: Real data $\mathcal{T}$, Diffusion model ϵ_θ, $\mathcal{T}_{\mathcal{SP}}$ model $\mathcal{F}$, Sampling steps T, Task-Driven strength $s(t)$
Output: Synthetic dataset $\mathcal{S}$.

Step I: Universal Task Knowledge Mining
Train $\mathcal{T}_{\mathcal{SP}}$ model $\mathcal{F}$ on $\mathcal{T}$ using Equ. 6
Use GPT-4o to generate task-aligned context for classification, detection, and segmentation datasets
Step II: Universal Task-Driven Diffusion for Data Synthetic
for each sampling step $t = 1$ to T **do**
    Generate noisy sample z_t.
    Predict clean image $\hat{z}_0$ using Equ. 3
    Calculate guidance loss $\nabla_{z_t}\mathcal{L}_{task}(c, \mathcal{F}(\hat{z}_0))$.
    Adjust noise prediction using Equ. 8
    Update z_{t-1} using adjusted noise prediction.
end for
Score each generated image x_i using $\mathcal{F}(x_i)$.
Filter and relabel synthetic set based on threshold τ using Equ. 11.

4. Experiments

4.1. Experimental Setups

Datasets and Evaluation Metric. To enhance practicality, we prioritize high-resolution and large-scale datasets. (1) **Classification task**: our experiments are primarily conducted on the full ImageNet 1K dataset [8], with IPC settings commonly set to 10 and 50, following the configurations in [41]. Additionally, based on the dataset partitioning settings from previous work [3], we also use the ImageNet 10-class subsets, such as ImageWoof, ImageFruit, ImageNette, and ImageMeow. These 10-class subsets contain very similar categories, making them challenging for dataset distillation. Top-1 accuracy (%) is reported in the experimental tables. (2) **Object detection task**: we conduct experiments on the widely-used Pascal VOC [11, 12] and MS COCO [28] datasets, using mAP (Mean Average Precision) and AP_{50} (Average Precision at IoU threshold of 0.50) as evaluation metric. (3) **Semantic segmentation task**: we utilize the Pascal VOC dataset, with mIoU (Mean Intersection over Union) as evaluation metrics.

Implementation Details. Unless otherwise specified, the Task-Specific Proxy $\mathcal{T}_{\mathcal{SP}}$ model used for the classification task is ResNet-18 [21], the detector is Faster R-CNN [36], and the segmentation model is LR-ASPP [24], all utilizing the official pre-trained weights from PyTorch. For the diffusion model, we use SD $v1.5$ [37] as the baseline. For quantitative comparisons, we generate 256×256 images for classification tasks and 512×512 images for detection and segmentation tasks. Synthetic images are created using 250 denoising steps based on a predefined number of random noise samples, aligned with the IPC settings. The threshold τ for Proxy-Driven High-Realism Refinement is set to 0.9 for classification, 0.85 for detection, and 0.75 for segmentation. Generally, this operation results in a data discard rate of approximately 5-10%. All experiments are conducted on NVIDIA V100 GPUs, with further details provided in the special experiments section. We generate the distilled dataset 3 times, train the network from scratch 5 times per dataset (with pre-trained backbone weights for the detector and segmenter), and evaluate on the original test set, reporting results as $\bar{x} \pm std$.

4.2. Results on Classification

We perform comprehensive comparisons across datasets of various scales. For large-scale datasets ImageNet-1k, we compare TESLA [7], SRe^{2}L [46], D^{4}M [41], RDED [42], and MiniMax Diffusion [18] (abbreviated as MiMxDiff in

Dataset	IPC	Method	Accuracy(%)
ImageNet-1K	1000	Full dataset	69.8
	10	TESLA	7.7
		SRe^{2}L	21.3
		D^{4}M	27.9
		RDED	42.0
		MiMxDiff	<u>44.3</u>
		UniDD (Ours)	**50.4 (↑6.1)**
	50	SRe^{2}L	46.8
		D^{4}M	55.2
		RDED	56.5
		MiMxDiff	<u>58.6</u>
		UniDD (Ours)	**62.8 (↑4.2)**

Table 1. Performance comparison with state-of-the-art methods on large-scale dataset ImageNet-1K. All methods employ ResNet18 as test model. All standard deviations in this table are less than one. The results of Full dataset come from the official PyTorch. The best results are marked in bold, while the second-best results are underlined.

the table). For small-scale dataset ImageNet-Subset, we compare Random selection, DM [49], MTT [3], HaBa [29], DataDAM [38], and MiniMax Diffusion.

Quantitative comparison on ImageNet-1K and ImageNet-Subset. As shown in the Tab 1, we utilize ResNet-18 as the $\mathcal{T}_{\mathcal{SP}}$ model, trained on ImageNet-1K. We report results for IPC-10 and IPC-50, evaluating all outcomes with the ResNet-18 model. To ensure a fair comparison, our training and testing pipeline is aligned with that of RDED and D^{4}M. Our method, UniDD, achieves state-of-the-art performance on this large-scale dataset. Specifically, at IPC-10, UniDD surpasses optimization-based methods, delivering substantial gains over TESLA and SRe^{2}L. Additionally, it demonstrates significant advantages over diffusion-based methods, outperforming D^{4}M by 22.5% and MiniMax Diffusion by 6.1%. These results underscore UniDD's effectiveness by using a lightweight $\mathcal{T}_{\mathcal{SP}}$ model to guide SD denoising, aligning generative and target domains without requiring diffusion model finetuning, thus advancing performance in dataset distillation.

As shown in the Tab 2, we conduct experiments on the ImageNet-subset under the IPC-10 setting. For consistency, the test model is standardized to a 5-layer ConvNet across all methods, except for MiniMax Diffusion, which uses a 6-layer ConvNet. Our method stands out among the comparison methods, outperforming the second-best by 8.2% on ImageWoof, 9.1% on ImageFruit, and 6.6 % on ImageMeow. Additionally, our method surpasses the diffusion-based MiniMax Diffusion by 9.8 % on ImageWoof, further demonstrating the effectiveness of UniDD approach in aggregating information from the original dataset.

Qualitative Result. In Fig. 2 (a), we present the synthetic images generated by UniDD on ImageWoof. Benefiting from the powerful generative capability of SD and the use of a $\mathcal{T}_{\mathcal{SP}}$ model to align the real and synthetic data domains, our synthetic images for the classification task exhibit significantly higher realism compared to optimization-based methods [3, 46, 50]. However, unlike other diffusion-based dataset distillation methods [18, 41], we do not finetune SD, greatly reducing the deployment time of UniDD.

4.3. Results on Detection and Segmentation

We deploy our UniDD framework for detection and segmentation tasks and conduct experiments on the Pascal VOC and MS COCO datasets. Since dataset distillation is rarely applied to these tasks, we reference earlier dataset distillation works (DD [44], DC [50]) and compare UniDD with core-set selection methods. Specifically, we construct the following baseline methods: Random [35], K-center [39], and Herding [2, 5]. Additionally, we develop a method that selects a subset with balanced class representation, which we refer to as Uniform.

Quantitative Result. In Table 3, we compare coreset selection methods with UniDD across different compression ratios on the object detection task. According to the official MMDetection implementation [4], the detector Faster R-CNN [36] achieves 51.4% mAP and 80.3% AP_{50} on Pascal VOC with the full dataset, and 32.6% mAP and 51.4% AP_{50} on MS COCO, providing an approximate performance upper bound. On Pascal VOC, UniDD demonstrates significant performance gains, achieving 7.7% mAP and 19.2% AP_{50} higher than the Random method at an extremely low compression ratio of 0.5%. Additionally, UniDD surpasses Uniform by 11.1% mAP and 21.2% AP_{50} at a 1% compression ratio. On the more challenging MS COCO dataset, UniDD also shows a clear advantage, surpassing the best coreset selection method by 3.7%, 3.4%, and 3.4% mAP at compression ratios of 0.25%, 0.5%, and 1%, respectively. This indicates UniDD's robustness, achieving strong results even on more complex datasets, which can be attributed to the Task Knowledge Mining in the first stage that extracts sufficient task-relevant information.

In Table 4, we present experimental results on the VOC dataset for the semantic segmentation task. Based on the MMSegmentation [6] implementation, LR-ASPP with a MobileNetV3 backbone achieves 32.9% mIoU on the full dataset. UniDD outperforms coreset selection methods at all compression ratios, demonstrating its effectiveness in dataset distillation for segmentation tasks. This result further underscores UniDD's versatility across different tasks. By leveraging task-driven diffusion synthesis alongside task knowledge mining, UniDD generates synthetic data well-suited to complex, pixel-level predictions, illustrating its

Dataset	Random	DM	MTT	HaBa	DataDAM	MiMxDiff	UniDD (Ours)	Full
ImageNette	47.7±2.4	58.1±0.3	63.0 ±1.3	64.7 ±1.6	59.4±0.4	62.0±0.2	**66.0±0.7**	87.4 ±1.0
ImageWoof	27.0±1.9	31.4±0.5	35.8 ±1.8	38.6 ±1.2	34.4±0.4	37.0±1.0	**46.8±0.9**	67.0 ±1.3
ImageFruit	21.4±1.2	-	40.3 ±1.3	42.5 ±1.6	-	-	**51.6±0.5**	63.9 ±2.0
ImageMeow	29.0±1.1	-	40.4 ±2.2	42.9 ±0.9	-	-	**49.5±0.7**	66.7 ±1.1

Table 2. Performance comparison with other state-of-the-art methods on ImageNet-Subset. IPC is set to 10. The best results are marked in bold, while the second-best results are underlined. "-" were not officially implemented in the original papers.

Methods	Object Detection			
	Pascal VOC		MS COCO	
	mAP	AP50	mAP	AP50
Ratio	0.5%		0.25%	
Random	0.8±0.2	3.1±0.4	0.5±0.1	1.7±0.3
Uniform	0.9±0.1	3.4±0.3	0.8±0.2	2.4±0.5
K-Center	0.5±0.1	2.1±0.3	0.4±0.1	1.5±0.2
Herding	0.6±0.2	2.4±0.2	0.5±0.1	1.8±0.4
UniDD (Ours)	**8.5±0.4**	**22.3±0.6**	**4.5±0.3**	**10.3±0.4**
Ratio	1%		0.5%	
Random	4.2±0.5	13.7±0.6	3.7±0.2	10.1±0.3
Uniform	5.7±0.2	17.7±0.4	3.4±0.4	9.5±0.6
K-Center	3.6±0.6	12.3±0.3	3.2±0.5	9.3±0.5
Herding	3.5±0.5	11.9±0.5	3.5±0.3	9.7±0.3
UniDD (Ours)	**16.8±0.5**	**38.9±0.7**	**7.1±0.4**	**16.9±0.3**
Ratio	2%		1%	
Random	12.4±0.4	34.3±0.5	7.2±0.8	17.3±0.9
Uniform	13.8±0.3	36.2±0.4	7.4±0.5	17.6±0.5
K-Center	10.9±0.6	29.3±0.6	6.1±0.3	15.4±0.6
Herding	10.4±0.4	28.7±0.7	6.7±0.4	16.3±0.7
UniDD (Ours)	**23.9±0.5**	**48.5±0.6**	**10.8±0.4**	**22.5±0.5**
Full	51.4±0.8	80.3±0.4	32.6±0.7	51.4±0.8

Table 3. Performance comparison with coreset selection methods on Pascal VOC and MS COCO for the object detection task, using Faster R-CNN as the detector. "Ratio" denotes the compression ratio.

Methods	Semantic Segmentation (mIoU)			
	Ratio: 3.5%	Ratio: 7%	Ratio: 14%	Full
Random	6.7±0.4	9.6±0.5	11.3±0.4	
Uniform	6.5±0.6	9.3±0.4	12.1±0.5	
K-Center	6.1±0.4	8.5±0.6	10.4±0.7	32.9±0.6
Herding	6.2±0.7	8.7±0.4	10.6±0.3	
UniDD (Ours)	**10.3±0.5**	**12.3±0.3**	**14.9±0.7**	

Table 4. Performance comparison with coreset selection methods on Pascal VOC for the semantic segmentation task, using LR-ASPP as the segmenter. The metric is mIoU, and "Ratio" denotes the compression ratio.

versatility across a range of vision tasks.

Qualitative Result. As shown in Fig. 2 (b) and 2 (c), we present visualizations of detection and segmentation data synthesized using the UniDD framework. These images are generated with $\mathcal{T}_{\mathcal{SP}}$ models, Faster R-CNN for detection and LR-ASPP for segmentation, which effectively steer the diffusion process toward the specified labels, bounding boxes, and masks. Previous research in dataset distillation has predominantly focused on single-category label synthesis, where each image aligns with a single class. In contrast, our UniDD framework overcomes this limitation, enabling dataset distillation for more complex, multi-class tasks. This advancement not only broadens the applicability of dataset distillation but also offers valuable insights for future research in this domain.

4.4. Ablation Study

Component Analysis. We conduct a comprehensive ablation study on the proposed UniDD, analyzing its components: (1) Baseline (a standard SD $v1.5$ checkpoint), (2) Task-driven Diffusion with the $\mathcal{T}_{\mathcal{SP}}$ model, (3) Task-Aligned Contextual Prompting (abbreviated as Context), and (4) Proxy-Driven High-Realism Refinement (abbreviated as Refine). As shown in Table 5, the proposed task-driven approach significantly improves the performance of the baseline, demonstrating that this guidance effectively bridges the gap between real and synthetic datasets. The contextual information is more effective for more complex detection and segmentation datasets, as it reveals underlying target relationships. The refinement module also proves to be effective, as the randomness of the SD model can lead to the generation of images that deviate from the annotations or result in failed generations.

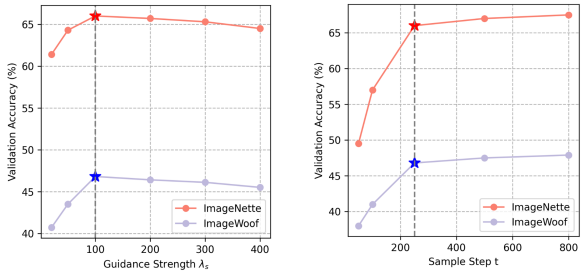
Methods	Dataset			
	Cls. / Accuracy (%)		Det. / mAP (%)	Seg. / mIoU (%)
	ImageNette	ImageWoof	VOC	VOC
Baseline	60.4±0.9	39.1±1.1	11.5±0.2	7.2±0.3
Baseline+$\mathcal{T}_{\mathcal{SP}}$	64.2±0.8	45.3±0.6	14.7±0.4	8.9±0.5
Baseline+$\mathcal{T}_{\mathcal{SP}}$+Context	64.7±0.4	46.1±0.7	15.9±0.3	9.6±0.4
Baseline+$\mathcal{T}_{\mathcal{SP}}$+Context+Refine	66.0±0.7	46.8±0.9	16.8±0.5	10.3±0.5

Table 5. Ablation study on UniDD. For classification datasets, IPC is set to 10 with ResNet-18 as the test model. For detection on VOC, the compression ratio is set to 1% with Faster R-CNN as the test model. For segmentation on VOC, the ratio is set to 3.5% with LR-ASPP as the test model.

Cross-Architecture Analysis. Cross-Architecture experi-

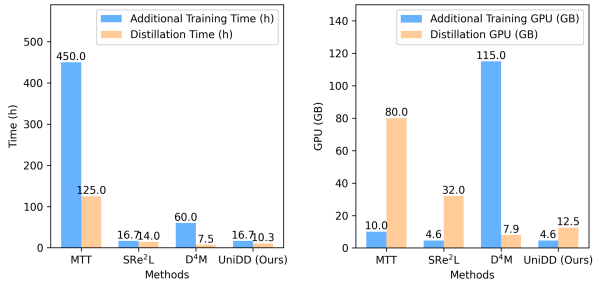
Dataset	Test Model	$\mathcal{T}_{\mathcal{SP}}$ Model		
		ResNet-18	ResNet-50	ResNet-101
ImageNette	ResNet-18	66.0±0.7	68.8±0.7	69.5±0.7
	ResNet-50	67.2±0.6	69.5±1.1	70.7±0.9
	ResNet-101	69.3±0.9	70.9±1.2	**72.1±0.6**
ImageWoof	ResNet-18	46.8±0.9	49.2±0.5	50.1±0.8
	ResNet-50	47.3±0.7	50.0±0.8	50.9±0.6
	ResNet-101	48.9±1.1	51.5±0.6	**52.4±0.7**

Table 6. $\mathcal{T}_{\mathcal{SP}}$ Model Ablation and Cross-Architecture Testing: We use ResNet-18, ResNet-50, and ResNet-101 as $\mathcal{T}_{\mathcal{SP}}$ models to generate synthetic images. During testing, the same three models are used. IPC is set to 10.



(a) Accuracy vs. Guidance strength (b) Accuracy vs. Sample Step

Figure 4. Hyper-parameter analysis on (a) guidance strength λ_s and (b) sampling step t. The star symbol indicates the parameters chosen for trade-off.



(a) Training Time overhead (b) Training GPU overhead

Figure 5. Comparison of distillation cost on ImageNet-1k using our UniDD with comparison methods (MTT [3], SRe^{2}L [46], and D^{4}M [41]), tested on V100 GPUs.

ments are crucial for validating the generalizability of synthetic data. We compare different combinations of $\mathcal{T}_{\mathcal{SP}}$ models and testing models. In this study, we set up three $\mathcal{T}_{\mathcal{SP}}$ models: ResNet-18, ResNet-50, and ResNet-101. These models are also used during testing to evaluate the cross-architecture performance of the synthetic datasets. As shown in Table. 6, using more complex $\mathcal{T}_{\mathcal{SP}}$ models improves the guidance effectiveness. Moreover, when testing across different architectures, the performance does not degrade as seen in previous optimization-based methods [3, 49, 50], demonstrating the generalizability of our synthetic data.

Guidance Stregth Analysis. As shown in Fig. 4 (a), we analyze the performance variations of ImageWoof and ImageNette at IPC-10 under different guidance strength settings λ_s, with Conv5 as the test model. The best results are achieved at $\lambda_s = 100$, but further increasing the guidance strength interferes with the quality of diffusion-generated images, leading to a decline in test performance.

Sample Step Analysis. As shown in Fig. 4 (b), we analyze the impact of different SD sample steps on the performance of synthetic datasets under the IPC-10 setting, with ImageWoof and ImageNette resolutions set to 256x256. Performance improves significantly with sample steps below 250, but gains are minimal beyond this point while time consumption increases. Thus, we select 250 steps as the optimal trade-off.

Distillation Cost Analysis. As shown in Fig. 5, we evaluate the time and GPU cost of various DD methods. Previous methods primarily focus on the time required to synthesize images, often neglecting the time needed to train or deploy the DD model. For MTT [3], the deployment time corresponds to the expert trajectory training time, while for D^{4}M [41], it is the time required to finetune the SD model. The slight increase in UniDD's distillation time and GPU cost is mainly due to the added supervision process of the $\mathcal{T}_{\mathcal{SP}}$ model and the refinement stage, where approximately 5%-10% of low-quality generated data is discarded. Combining the deployment and distillation costs for comparison, UniDD demonstrates the best advantage, indicating its superior suitability for practical deployment.

5. Conclusion

In this study, we introduce the first Universal Dataset Distillation (UniDD) framework, capable of handling classification, detection, and segmentation tasks within a unified approach. UniDD follows a two-stage process: in the first stage, Task-Specific Proxy Training and Task-Aligned Contextual Prompting extract critical information from real datasets. In the second stage, Universal Task-driven Diffusion synthesizes images with specific categories, bounding boxes, and segmentation masks tailored to each task. We conduct extensive experiments to validate UniDD's effectiveness across all three tasks. For classification, UniDD achieves state-of-the-art results on ImageNet-1k and ImageNet-Subset. In detection and segmentation, we establish the first dataset distillation benchmark on Pascal VOC and MS COCO, comparing UniDD to coreset selection methods. Encouraging results demonstrate that our UniDD consistently outperforms these methods, especially under extremely low compression ratios.

Acknowledgments. This work was supported by National Natural Science Fund of China (62473286).

References

[1] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 843–852, 2023. 3, 4

[2] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Proceedings of the European conference on computer vision (ECCV)*, pages 233–248, 2018. 6

[3] George Cazenavette, Tongzhou Wang, Antonio Torralba, Alexei A Efros, and Jun-Yan Zhu. Dataset distillation by matching training trajectories. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4750–4759, 2022. 1, 2, 5, 6, 8

[4] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 6

[5] Yutian Chen, Max Welling, and Alex Smola. Super-samples from kernel herding. *arXiv preprint arXiv:1203.3472*, 2012. 6

[6] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. 6

[7] Justin Cui, Ruochen Wang, Si Si, and Cho-Jui Hsieh. Scaling up dataset distillation to imagenet-1k with constant memory. In *International Conference on Machine Learning*, pages 6565–6590. PMLR, 2023. 1, 5

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1, 2, 3, 5

[9] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 3, 4

[10] Dave Epstein, Allan Jabri, Ben Poole, Alexei Efros, and Aleksander Holynski. Diffusion self-guidance for controllable image generation. *Advances in Neural Information Processing Systems*, 36:16222–16239, 2023. 3

[11] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. 1, 2, 3, 5

[12] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111:98–136, 2015. 1, 2, 3, 5

[13] Junyao Gao, Xinyang Jiang, Huishuai Zhang, Yifan Yang, Shuguang Dou, Dongsheng Li, Duoqian Miao, Cheng Deng, and Cairong Zhao. Similarity distribution based membership inference attack on person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, pages 14820–14828, 2023. 1

[14] Junyao Gao, Xinyang Jiang, Shuguang Dou, Dongsheng Li, Duoqian Miao, and Cairong Zhao. Re-id-leak: Membership inference attacks against person re-identification. *International Journal of Computer Vision*, 132(10):4673–4687, 2024. 1

[15] Junyao Gao, Yanchen Liu, Yanan Sun, Yinhao Tang, Yanhong Zeng, Kai Chen, and Cairong Zhao. Styleshot: A snapshot on any style. *arXiv preprint arXiv:2407.01414*, 2024. 2

[16] Junyao Gao, Yanan Sun, Fei Shen, Xin Jiang, Zhening Xing, Kai Chen, and Cairong Zhao. Faceshot: Bring any character into life. *arXiv preprint arXiv:2503.00740*, 2025. 2

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[18] Jianyang Gu, Saeed Vahidian, Vyacheslav Kungurtsev, Haonan Wang, Wei Jiang, Yang You, and Yiran Chen. Efficient dataset distillation via minimax diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15793–15803, 2024. 5, 6

[19] Ziyao Guo, Kai Wang, George Cazenavette, Hui Li, Kaipeng Zhang, and Yang You. Towards lossless dataset distillation via difficulty-aligned trajectory matching. *arXiv preprint arXiv:2310.05773*, 2023. 2

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4, 5

[22] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022. 1

[23] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 3

[24] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019. 5

[25] Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018. 2

[26] Jang-Hyun Kim, Jinuk Kim, Seong Joon Oh, Sangdoo Yun, Hwanjun Song, Joonhyun Jeong, Jung-Woo Ha, and Hyun Oh Song. Dataset condensation via efficient synthetic-data parameterization. In *International Conference on Machine Learning*, pages 11102–11118. PMLR, 2022. 1, 2

[27] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. Dsfd: dual shot face detector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5060–5069, 2019. 2

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 1, 2, 5

[29] Songhua Liu, Kai Wang, Xingyi Yang, Jingwen Ye, and Xinchao Wang. Dataset distillation via factorization. *Advances in Neural Information Processing Systems*, 35:1100–1113, 2022. 6

[30] Xihui Liu, Dong Huk Park, Samaneh Azadi, Gong Zhang, Arman Chopikyan, Yuxiao Hu, Humphrey Shi, Anna Rohrbach, and Trevor Darrell. More control for free! image synthesis with semantic diffusion guidance. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 289–299, 2023. 3

[31] Yanqing Liu, Jianyang Gu, Kai Wang, Zheng Zhu, Wei Jiang, and Yang You. Dream: Efficient dataset distillation by representative matching. *arXiv preprint arXiv:2302.14416*, 2023. 2

[32] Timothy Nguyen, Zhourong Chen, and Jaehoon Lee. Dataset meta-learning from kernel ridge-regression. In *International Conference on Learning Representations*, 2020. 2

[33] OpenAI. Gpt-4 technical report. https://openai.com/research/gpt-4, 2023. Accessed: 2024-11-13. 2, 4

[34] Ding Qi, Jian Li, Jinlong Peng, Bo Zhao, Shuguang Dou, Jialin Li, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Fetch and forge: Efficient dataset condensation for object detection. *Advances in Neural Information Processing Systems*, 37:119283–119300, 2024. 3

[35] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017. 6

[36] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 4, 5, 6

[37] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 5

[38] Ahmad Sajedi, Samir Khaki, Ehsan Amjadian, Lucy Z Liu, Yuri A Lawryshyn, and Konstantinos N Plataniotis. Datadam: Efficient dataset distillation with attention matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17097–17107, 2023. 6

[39] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017. 6

[40] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 3

[41] Duo Su, Junjie Hou, Weizhi Gao, Yingjie Tian, and Bowen Tang. Dˆ 4: Dataset distillation via disentangled diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5809–5818, 2024. 5, 6, 8

[42] Peng Sun, Bei Shi, Daiwei Yu, and Tao Lin. On the diversity and realism of distilled dataset: An efficient dataset distillation paradigm. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9390–9399, 2024. 1, 3, 5

[43] Kai Wang, Bo Zhao, Xiangyu Peng, Zheng Zhu, Shuo Yang, Shuo Wang, Guan Huang, Hakan Bilen, Xinchao Wang, and Yang You. Cafe: Learning to condense dataset by aligning features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12196–12205, 2022. 2

[44] Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A Efros. Dataset distillation. *arXiv preprint arXiv:1811.10959*, 2018. 1, 2, 6

[45] Yichao Yan, Zanwei Zhou, Zi Wang, Jingnan Gao, and Xiaokang Yang. Dialoguenerf: Towards realistic avatar face-to-face conversation video generation. *Visual Intelligence*, 2 (1):24, 2024. 1

[46] Zeyuan Yin, Eric Xing, and Zhiqiang Shen. Squeeze, recover and relabel: Dataset condensation at imagenet scale from a new perspective. *arXiv preprint arXiv:2306.13092*, 2023. 1, 3, 5, 6, 8

[47] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2

[48] Bo Zhao and Hakan Bilen. Dataset condensation with differentiable siamese augmentation. In *International Conference on Machine Learning*, pages 12674–12685. PMLR, 2021. 1

[49] Bo Zhao and Hakan Bilen. Dataset condensation with distribution matching. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6514–6523, 2023. 1, 2, 6, 8

[50] Bo Zhao, Konda Reddy Mopuri, and Hakan Bilen. Dataset condensation with gradient matching. In *International Conference on Learning Representations*, 2020. 1, 2, 6, 8

[51] Ganlong Zhao, Guanbin Li, Yipeng Qin, and Yizhou Yu. Improved distribution matching for dataset condensation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7856–7865, 2023. 1, 2