Low-Level Dataset Distillation for Medical Image Enhancement

Fengzhi Xu[†], Ziyuan Yang[†,‡,⋆], Mengyu Sun[†], Joey Tianyi Zhou[§,◇], and Yi Zhang[†,⋆]

[†] School of Cyber Science and Engineering, Sichuan University, China
[‡] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[§] Singapore Management University, Singapore
[◇] Institute of Advanced Intelligence and Computing (IAIC), Agency for Science, Technology and Research (A*STAR), Singapore

Abstract. Medical image enhancement is clinically important, but existing methods often rely on large-scale datasets to learn complex image-to-image mappings. Such reliance leads to high training and storage costs and hinders practical deployment.

Dataset distillation (DD) provides a promising solution by synthesizing compact datasets that preserve the training behavior of the original data. However, existing DD methods mainly focus on high-level tasks, where multiple samples share a common semantic label and the distilled data can compress shared representations. In contrast, low-level medical image enhancement involves dense image-to-image mappings, where each sample corresponds to a unique pixel-level target. This fundamental difference makes low-level DD significantly more challenging and inherently underdetermined.

To address this issue, we propose the first DD framework for low-level medical image enhancement. We construct a shared anatomical prior from a representative subject. This prior serves as the initialization and constrains the solution space. We further personalize the prior through a Structure-Preserving Personalized Generation (SPG) module, which incorporates subject-specific structural characteristics while maintaining fidelity. To capture task-specific degradation effects, we use the distilled dataset to form paired high- and low-quality samples for optimization.

More importantly, we align the optimization behavior of networks trained on the distilled dataset with that of networks trained on the full dataset through subject-aware gradient matching. This design ensures that the distilled data preserve essential training dynamics rather than only appearance-level similarity. Finally, only the distilled dataset is shared for downstream use, so raw patient data remain inaccessible and privacy protection is improved. Extensive experiments across multiple modalities and tasks demonstrate the effectiveness of our method. [1]

⋆ Corresponding authors: Ziyuan Yang (cziyuanyang@gmail.com), and Yi Zhang (yzhang@scu.edu.cn)

[1] The code is publicly available at `https://github.com/xufz123/Med_LLDD`.

Keywords: Low-Level Dataset Distillation · Medical Image Enhancement

1 Introduction

Medical image enhancement tasks, such as super-resolution [10, 27] and restoration [8, 28], aim to transform low-quality (LQ) images into high-quality (HQ) ones. These tasks support accurate diagnosis and clinical decision-making. However, existing methods typically rely on large, high-resolution medical datasets to learn complex pixel-level mappings. The associated training and storage costs hinder their practical deployment [19].

To address these challenges, dataset distillation (DD) has emerged as an efficient dataset compression paradigm. It synthesizes a small distilled dataset while maintaining training performance comparable to that of the original large dataset [9, 30]. Although existing DD methods improve data efficiency, they are primarily designed for high-level tasks such as classification, where multiple samples share the same label. This many-to-one mapping allows distilled data to capture shared semantics among samples, so information compression becomes feasible. In contrast, low-level tasks involve a many-to-many mapping and require pixel-level fidelity. This makes low-level DD inherently underdetermined, since a limited number of synthetic samples cannot fully capture or constrain the dense pixel-level relationships between inputs and outputs. Fig. 1 illustrates the fundamental difference between these two settings.

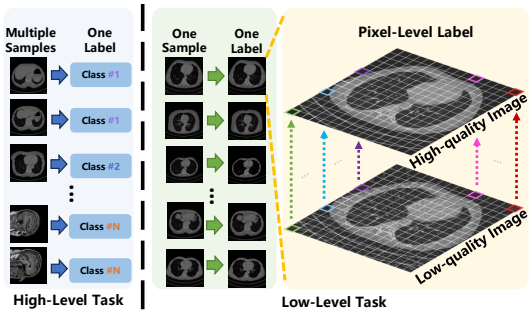

Fig. 1: The comparison with the high-level and low-level tasks.

To address this issue, we propose the first low-level DD method for medical image enhancement. We first construct a shared anatomical prior from a representative subject. This prior serves as a common structural initialization for all subjects. It introduces global constraints that restrict the solution space and help mitigate the underdetermined nature of low-level DD.

Furthermore, substantial anatomical variations exist among different subjects. Therefore, we avoid enforcing cross-subject distillation and instead perform distillation individually for each subject. This design better captures subject-specific characteristics and preserves structural fidelity. It also decomposes the highly underdetermined problem into a series of smaller and more manageable subproblems.

To this end, we design a Structure-Preserving Personalized Generation (SPG) module to personalize the shared anatomical prior for each subject. SPG modulates the prior with a learnable subject-specific code, which allows the distilled

data to adapt to individual structural characteristics. To further preserve pixel-level fidelity, an image is randomly selected from the prior and fused with the subject-specific distilled data. Since different low-level tasks follow distinct degradation models, the distilled data are further mapped to task-specific low-quality representations to construct corresponding HQ-LQ training pairs. Finally, we incorporate subject-specific training knowledge by aligning the gradients computed from the distilled data with those from the corresponding subject's raw data. In this way, the distilled dataset captures both structural and task-relevant information.

Notably, our method does not directly copy subject-level anatomical structures into the distilled data. Instead, it aligns the optimization trajectory of networks trained on the distilled dataset with that of networks trained on the raw dataset. Through this trajectory alignment, subject-specific training dynamics are implicitly encoded in the distilled dataset. As a result, the distilled dataset achieves performance comparable to raw-data training and reduces the need to share raw subject data. Extensive experiments across different settings validate the effectiveness of our method on multiple medical modalities, including Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), as well as across various low-level tasks such as super-resolution and image restoration. The main contributions of this paper are summarized as follows:

- We propose a general DD method for low-level medical image enhancement, applicable across diverse modalities and tasks. To the best of our knowledge, this is the first attempt to explore DD in the low-level domain.
- We leverage anatomical similarities across subjects to build a shared anatomical prior as initialization, while capturing subject-specific variations through subject-wise distillation.
- We introduce a subject-wise personalization module to embed subject- and task-specific knowledge into the distilled data by aligning gradients computed from both distilled and raw data.

2 Related Work

2.1 Medical Image Enhancement

Medical image enhancement aims to recover high-quality images from degraded inputs while preserving critical details that support accurate diagnosis. Representative tasks include super-resolution [13, 15], which enhances image details, and restoration [24, 28], which removes noise and artifacts. With the rapid progress of deep learning, medical image enhancement has achieved substantial advances across diverse imaging modalities [1, 22]. However, most existing methods rely on large-scale datasets. This reliance introduces substantial storage and computational overhead, which limits the development and practical application of these approaches.

2.2 Dataset Distillation

Dataset distillation (DD) has emerged as an efficient data compression paradigm. It synthesizes a small distilled dataset while maintaining training performance comparable to that of the raw dataset [9, 30]. Existing DD methods can be broadly classified according to their matching strategies. 1) Performance matching formulates distillation as a bi-level optimization problem [25], with further improvements such as momentum-enhanced inner-loop updates [5] and kernel-based approximations [12,17,18]. 2) Parameter matching aligns model optimization trajectories through gradient matching [33] or multi-step updates [2]. It is often combined with strategies such as symmetric augmentation [31] or learnable soft labels [4]. 3) Distribution matching aligns feature distributions between synthetic and raw data, such as class-wise distribution alignment [32] or layer-wise statistical alignment [23]. However, these methods are primarily tailored for high-level tasks and are not directly applicable to low-level tasks that require dense pixel-level fidelity.

3 Problem Statement

Existing DD methods mainly focus on high-level tasks, such as classification, and are difficult to directly extend to low-level tasks. To clarify the underlying challenge, this section formally defines DD for both high- and low-level tasks, highlights their key distinctions, and discusses the inherent difficulty of low-level DD.

Given a high-level raw dataset $\mathcal{T}^h = \{(\mathbf{x}_n, y_n)\}_{n=1}^N$, where $y_n \in \{1, \ldots, C\}$ and C denotes the number of classes, DD seeks to extract the knowledge of $\mathcal{T}^h$ into a compact synthetic dataset $\mathcal{S}^h = \{\tilde{\mathbf{x}}_m, \tilde{y}_m\}_{m=1}^M$, where $\tilde{y}_m \in \{1, 2, ..., C\}$ and $M \ll N$. The goal is to train a model parameterized by $\theta_{\mathcal{S}^h}$ on $\mathcal{S}^h$ such that it achieves performance comparable to that of a model with parameters $\theta_{\mathcal{T}^h}$ trained on $\mathcal{T}^h$. This objective can be formulated as:

$$\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{P}_{\mathcal{T}^h}} |\mathcal{L}^h(\psi(\mathbf{x}; \theta_{\mathcal{T}^h}), y) - \mathcal{L}^h(\psi(\mathbf{x}; \theta_{\mathcal{S}^h}), y)| \leq \epsilon, \tag{1}$$

where $\mathcal{P}_{\mathcal{T}^h}$ represents the distribution of the raw dataset $\mathcal{T}^h$, and ϵ is a small tolerance value. $\mathcal{L}^h$ and ψ denote the loss function and the network for high-level tasks, respectively.

In high-level tasks, multiple samples often share the same label, which leads to a many-to-one mapping. This property allows synthetic data to focus on common high-level semantics within each class and makes high-level DD feasible. In contrast, low-level tasks involve a many-to-many mapping between samples and labels and require pixel-level fidelity. This difference substantially increases the difficulty of distillation.

For low-level tasks, we define the raw dataset $\mathcal{T}^l = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$, where $\mathbf{y}_n \in \mathbb{R}^{h\times w}$), and the synthetic dataset as $\mathcal{S}^l = \{\tilde{\mathbf{x}}_m, \tilde{\mathbf{y}}_m\}_{m=1}^M$, where $\tilde{\mathbf{y}}_m \in \mathbb{R}^{h\times w}$). Based on these definitions, the objective of low-level DD is formulated as follows:

$$\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{P}_{\mathcal{T}^l}} |\mathcal{L}^l(\phi(\mathbf{x}; \theta_{\mathcal{T}^l}), \mathbf{y}) - \mathcal{L}^l(\phi(\mathbf{x}; \theta_{\mathcal{S}^l}), \mathbf{y})| \leq \epsilon, \tag{2}$$

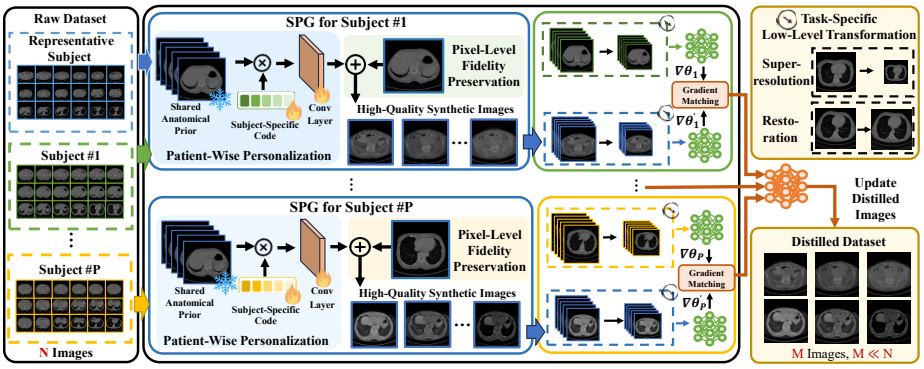

Fig. 2: The overview of our proposed method.

where $\mathcal{P}_{\mathcal{T}^l}$ represents the distribution of the raw dataset $\mathcal{T}^l$, $\phi(\mathbf{x};\theta_{\mathcal{T}^l}) \in \mathbb{R}^{h\times w}$ represents a dense mapping output, such as a restored image, rather than a single class label as in high-level tasks. $\mathcal{L}^l$ and ϕ denote the loss function and the network for low-level tasks, respectively.

From the above formulation, a network trained on the synthetic dataset must capture the pixel-level correspondence in the raw data. However, in low-level tasks, the dense mapping between samples and labels makes it challenging for a small synthetic dataset to fully capture these relationships. Therefore, low-level DD is inherently more underdetermined than high-level DD.

4 Method

4.1 Overview

This paper proposes a low-level DD method for medical image enhancement, which aims to synthesize a compact dataset that maintains training performance comparable to the raw dataset. The overall framework is illustrated in Fig. 2. Based on the anatomical similarity among subjects, we construct a shared anatomical prior from a representative subject as the initialization. We then design an SPG module to personalize this prior with subject-specific structural characteristics and training dynamics. Since degradation models vary across different low-level tasks, we further introduce task-specific low-level transformations to construct paired low- and high-quality images from the distilled data. Finally, the distilled and raw pairs are used to train the image enhancement network, and their gradients are matched to update the distilled data and the SPG module.

4.2 Structure-Preserving Personalized Generation

To generate structurally faithful distilled data, we propose a subject-specific generation pipeline. The proposed SPG pipeline consists of three steps: initialization, subject-wise personalization, and low- and high-quality image pair generation. The details of each stage are described in the following sections.

Initialization. As introduced earlier, high-level tasks usually follow a many-to-one mapping, where different images can correspond to a shared label. In contrast, low-level tasks involve dense image-to-image mappings, where different inputs generally correspond to different pixel-level targets. Consequently, the distilled data need to capture many-to-many mapping knowledge within a limited dataset, which makes low-level DD an underdetermined problem.

To alleviate this issue, we leverage the fact that different subjects share similar anatomical structures and introduce a shared anatomical prior as initialization. This prior provides global constraints that restrict the solution space and help mitigate the underdetermined nature of the problem. Notably, the initialization is robust to the choice of the representative subject, which indicates that anatomical structure priors can be shared across different subjects. Specifically, the shared anatomical prior is constructed by randomly selecting v slices from the 3D volume, such as CT or MRI, of a single representative subject in the raw dataset.It is denoted as $U \in \mathbb{R}^{v\times h\times w}$, where v represents the number of randomly selected images (NRI). These selected slices serve as the shared initialization for subsequent subject-wise distillation.

Subject-Wise Personalization. After initialization, our goal is to incorporate subject-specific information into the distilled dataset. However, subjects differ in their anatomical composition. Forcing a single distilled sample to represent multiple subjects is therefore infeasible and may degrade the quality of the distilled dataset. To avoid this issue, our method performs subject-wise distillation, so that the distilled data can preserve subject-specific structural and training characteristics.

To achieve this, we propose a subject-wise personalization module with two components: a learnable subject-specific code $d_p \in \mathbb{R}^q$ for the p-th subject and a subject-agnostic convolutional layer parameterized by θ^c.

Specifically, we adjust the shared anatomical prior U using the learnable subject-specific code d_p, which can be formulated as:

$$f(U, d_p) = \text{Concat}(d_p^1 \cdot U, ..., d_p^q \cdot U), \tag{3}$$

where $f(\cdot)$ denotes the adjustment operation, and its output lies in $\mathbb{R}^{v\times q\times h\times w}$. d_p^q represents the q-th element of d_p.

The adjusted output is then fed into a subject-agnostic convolutional layer to match its dimension to the number of images per subject (IPS), denoted as i. Finally, the personalized data has a dimension of $\mathbb{R}^{i\times h\times w}$.

Pixel-Level Fidelity Preservation. Since pixel-level fidelity is important in low-level tasks, we introduce a pixel-level fidelity preservation step. Specifically, for the p-th subject, a slice $\mathbf{u}_p \in \mathbb{R}^{h\times w}$ is randomly selected from U and duplicated i times to match the convolutional output dimension. The duplicated slice is then added to the personalized data. The overall SPG process can be formulated as:

$$\tilde{\mathbf{Y}}_\mathbf{p} = \text{Conv}(f(U, d_p); \theta^c) + \mathbf{u}_p \tag{4}$$

where $\tilde{\mathbf{Y}}_\mathbf{p} \in \mathbb{R}^{i\times h\times w}$ denotes the distilled data for the p-th subject.

Low- and High-Quality Image Pair Generation. Different low-level tasks follow distinct degradation models, so we first construct task-specific pairs of low- and high-quality images for subsequent training. Specifically, the personalized data are treated as the high-quality data, and the pair construction process based on the personalized data and the raw high-quality data can be formulated as:

$$\mathcal{S}_p^l = \{D(\tilde{\mathbf{y}}_m), \tilde{\mathbf{y}}_m\}_{m=1}^{i}, \mathcal{T}_p^l = \{D(\mathbf{y}_n^p), \mathbf{y}_n^p\}_{n=1}^{N_p}, \tag{5}$$

where $\tilde{\mathbf{y}}_m \in \mathbb{R}^{h\times w}$ denotes a personalized data sample from $\tilde{\mathbf{Y}}_\mathbf{p}$. $\mathbf{y}_n^p \in \mathbb{R}^{h\times w}$ is the raw high-quality image from the p-th subject, and N_p represents the number of samples. $D(\cdot)$ denotes a degradation operation that simulates low-quality images, such as resizing for super-resolution or adding noise in the measurement data for medical image restoration.

4.3 Subject-Awareness Gradient Matching

Since subjects differ in their anatomical composition, cross-subject distillation may damage individual structural characteristics and prevent the distilled data from converging effectively. Therefore, we split the dense mapping into multiple sub-mappings and perform distillation independently for each subject. This strategy helps prevent training collapse and better preserves subject-specific structural characteristics.

Specifically, our goal is to achieve subject-aware DD, which compresses each subject's raw dataset $\mathcal{T}_p^l$ into a smaller synthetic dataset $\mathcal{S}_p^l$. The collection of all $\mathcal{S}_p^l$ forms the overall distilled dataset $\mathcal{S}^l$. We define the raw dataset $\mathcal{T}^l$ as the collection of medical imaging data from P subjects. Our optimization objective is to ensure that the parameters learned from the distilled dataset are close to those learned from the raw dataset. Accordingly, the objective in Eq. (2) can be reformulated as:

$$\min_{\mathcal{S}^l} \sum_{p=1}^{P} \mathcal{L}_{PM}(\theta_{\mathcal{S}_p^l}, \theta_{\mathcal{T}_p^l}), \tag{6}$$

where $\mathcal{L}_{PM}$ denotes the parameter matching function, and $\theta_{\mathcal{S}_p^l}$ and $\theta_{\mathcal{T}_p^l}$ represent the network parameters learned from the synthetic dataset and the raw dataset of the p-th subject, respectively.

We aim to match the gradients computed from both the raw and distilled data. However, incorporating training information from multiple subjects into a single distilled sample is challenging due to substantial anatomical differences across subjects. To stand this, we visualize the gradient data from different subjects using t-SNE [14], as shown in Fig. 3. Gradients from different subjects exhibit

Fig. 3: t-SNE visualization of gradient data from different subjects. Different colors denote different subjects.

clear differences, while gradients from
the same subject are more similar.

Therefore, we propose a subject-wise distillation process to avoid cross-subject information conflicts, which could otherwise lead to training collapse. The subject-aware gradient matching process can be formulated as:

$$\mathcal{L}_{PGM}(\nabla\theta_p, \nabla\theta'_p) = 1 - \frac{\nabla\theta'_p \cdot \nabla\theta_p}{||\nabla\theta'_p||||\nabla\theta_p||}, \tag{7}$$

where $\nabla\theta_p$ and $\nabla\theta'_p$ denote the gradients computed on the raw dataset $\mathcal{T}_p^l$ and the synthetic dataset $\mathcal{S}_p^l$ for the p-th subject, respectively.

To compute the gradients used in the distillation, we employ the commonly used mean squared error (MSE) as the task loss to measure pixel-level fidelity between the network output and the corresponding high-quality image. The loss function is defined as:

$$\mathcal{L}^l_{MSE}(\mathbf{x}, \mathbf{y}) = ||\phi(D(\mathbf{y}); \theta) - \mathbf{y}||^2, \tag{8}$$

where $\phi_\theta(D(\mathbf{y}))$ denotes the enhanced image and $\mathbf{y}$ is the ground truth from the raw dataset. For the synthetic dataset, the loss is defined analogously by replacing the paired data with the synthetic dataset $\mathcal{S}_p^l$.

In this way, we obtain a compact low-level distilled dataset that is significantly smaller than the raw dataset, while still preserving essential training information.

4.4 Downstream Task Usage

Medical data are typically large-scale and sensitive, and cannot be directly shared due to privacy regulations and ethical concerns. To enable efficient training while preserving privacy, we provide the distilled dataset $\mathcal{S}^l = \{\tilde{\mathbf{x}}_m, \tilde{\mathbf{y}}_m\}_{m=1}^M$ to downstream users.

Based on this synthetic dataset, downstream users utilize paired high- and low-quality data and optimize image enhancement networks using the following formulation:

$$\phi_{\theta_{\mathcal{S}^l}} = \arg\min_\theta ||\phi(\tilde{\mathbf{x}}; \theta) - \tilde{\mathbf{y}}||^2, \tag{9}$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ denote the synthetic low- and high-quality image pairs generated from $\mathcal{S}^l$, respectively.

In our pipeline, subjects' raw data are never shared with downstream users. Instead, only the distilled dataset is accessed, which preserves subject-level and task-relevant information while avoiding exposure of raw data. This enables efficient training with significantly reduced data size compared to the original dataset. The corresponding visualization is provided in Sec. 5.4. Moreover, the proposed method is task-agnostic and can be readily adapted to different low-level medical image enhancement tasks by changing the degradation model $D(\cdot)$.

5 Experiment

5.1 Experimental Setup

Implementation Details. All experiments are conducted using PyTorch on an NVIDIA RTX 3090 GPU. We perform iterative distillation for 2,000 optimization steps. SRCNN [6] and REDCNN [3], two widely used networks, are employed for the super-resolution and restoration tasks, respectively. For our method, three independent distilled datasets are generated in each experiment, and five training–testing runs are conducted for each distilled dataset, resulting in 15 test scores. We report their mean and variance. For CT super-resolution and CT restoration, all networks are trained for 300 epochs, while MRI super-resolution networks are trained for 600 epochs. For coreset-based baselines, after selecting the subset, we perform multiple independent training–testing runs on the selected data and report the mean and variance of the resulting test scores.
Datasets. We evaluate the proposed method on multiple imaging modalities, including CT and MRI. For CT, we employ the public "NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge"[2] dataset [16]. In the super-resolution task, 50 images per subject from 10 subjects are used for training, while in the restoration task, all images from 10 subjects are used for training. For both tasks, data from two additional subjects are used for testing. For MRI, Calgary-Campinas-359 dataset[3] [21] is employed in our experiments. In the super-resolution task, 100 high-resolution MRI images per subject from 10 subjects are used for training, and data from two additional subjects are used for testing.
Data Simulation. We validate our method on two common medical image enhancement tasks: super-resolution and image restoration. For the super-resolution task, the degradation operator downsamples the 512×512 CT images by a factor of 4. For MRI, images are first resized to 256×256 and then downsampled by a factor of 2 to generate low-quality inputs. For the low-dose CT (LDCT) restoration task, degradation is simulated by adding noise to the projections followed by back-projection, with the photon count set to 10^4.
Comparison Methods. We compare our method to five baseline methods: Random, Random*, Uniform, Herding, and K-Center. Random selects a subset of images from all subjects, while Random* samples images from a single subject. Uniform selects samples at equal intervals across all subjects. Herding [26] minimizes the distance between the coreset and dataset centers in feature space, and K-Center Greedy (K-Center) [7, 20] solves a minimax facility location problem.
Evaluation Metrics. We adopt Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as evaluation metrics. For both measures, higher scores reflect better perceptual fidelity and structural integrity of the reconstructed images. Since low-level tasks lack explicit class concepts, we use the number of raw images used for initialization or selection (NRI) as a measure of dataset size.

[2] The Mayo dataset link is https://www.aapm.org/grandchallenge/lowdosect/
[3] The CC-359 dataset link is https://www.ccdataset.com/

Table 1: Quantitative super-resolution results of different methods on CT and MRI trained with SRCNN as the backbone. The Compression Rate (CR) measures the ratio of non-selected raw images to the total number of images in the raw dataset.

		CT (x4)				MRI (x2)			
		NRI=5/CR=99%		NRI=10/CR=98%		NRI=5/CR=99.5%		NRI=10/CR=99%	
Methods		PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑
Full Data		PSNR=36.10±0.02		SSIM=94.86±0.03		PSNR=29.07±0.00		SSIM=90.46±0.02	
Base	Random	31.05±0.35	87.38±1.32	32.66±0.41	91.06±0.51	25.54±0.08	79.27±0.42	26.80±0.11	83.63±0.55
	Random*	31.37±0.36	87.98±0.67	32.76±0.19	90.78±0.69	25.70±0.13	79.87±0.57	26.74±0.13	83.54±0.44
	Uniform	31.39±0.24	87.84±0.98	32.46±0.28	90.67±0.26	25.94±0.06	80.80±0.25	26.77±0.06	83.79±0.23
	Herding	31.41±0.23	88.80±0.40	32.86±0.26	91.18±0.48	25.63±0.13	79.57±0.58	26.72±0.09	83.62±0.36
	K-Center	31.59±0.24	89.22±0.40	32.92±0.23	91.34±0.28	25.70±0.09	79.97±0.50	26.76±0.10	83.70±0.36
Ours	IPS=1	32.73±0.22	90.50±0.78	32.93±0.28	90.15±1.90	26.95±0.21	71.78±15.14	27.02±0.28	77.04±8.22
	IPS=5	**34.45±0.20**	**92.07±0.93**	**34.57±0.14**	**92.83±0.28**	**28.16±0.18**	**87.73±0.44**	**28.35±0.10**	**88.27±0.33**

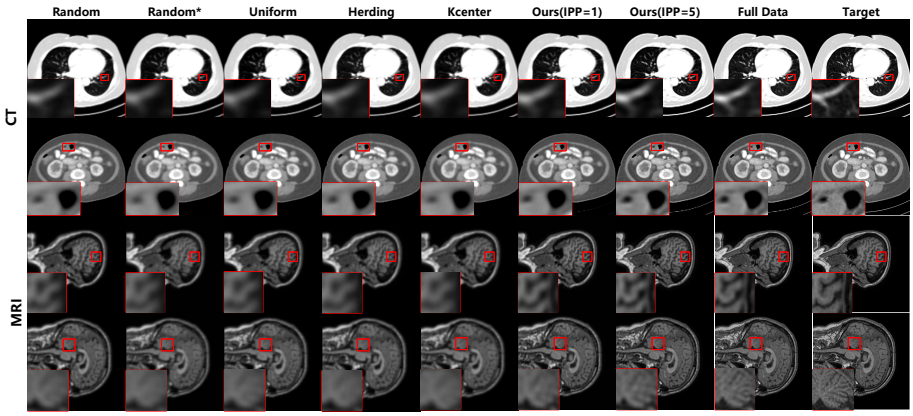

Fig. 4: Qualitative super-resolution results of different methods on CT and MRI. The display window is [-950, 50] HU for the first row and [-290, 310] HU for the second row.

5.2 Comparison with Other Methods

Super-Resolution in the CT Modality. For the CT modality, quantitative results are shown in Tab. 1. Overall, coreset selection methods achieve similar performance under the same NRI setting, while Herding and K-Center demonstrate slightly better stability. Compared with these methods, our method consistently outperforms them across all settings and reaches the best performance at IPS=5. Increasing IPS further improves performance and leads to more stable optimization. As NRI increases, all methods benefit from additional information and show performance gains. Our method achieves the best results, with a PSNR of 34.57 dB and an SSIM of 92.83. The qualitative results in Fig. 4 further demonstrate that our method better recovers anatomical details and produces results closer to the ground truth. We also include additional 3D CT results in the ***Appendix***.

Super-Resolution in the MRI Modality. For the MRI modality, the results in Tab. 1 show trends consistent with those observed in CT. Coreset selection methods achieve comparable performance, while our method consistently outper-

Table 2: Quantitative super-resolution results using SRCNN distillation across different network architectures for the CT modality.

		SRCNN		EDSR		SCTSRN	
Algorithm		PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑
Full data		36.10±0.02	94.86±0.03	38.29±0.03	96.19±0.01	35.01±0.09	94.68±0.07
Base	Random	32.66±0.41	91.06±0.51	34.14±0.04	92.66±0.07	34.48±0.02	94.18±0.06
	Uniform	32.46±0.28	90.67±0.26	34.03±0.11	92.40±0.21	34.48±0.01	94.22±0.04
	Herding	32.86±0.26	91.18±0.48	33.99±0.20	92.52±0.37	34.48±0.02	94.20±0.12
	K-Center	32.92±0.23	91.34±0.28	34.13±0.17	92.68±0.23	34.47±0.01	94.20±0.07
Ours	IPS=1	33.24±0.29	91.74±0.32	34.45±0.14	92.52±0.23	34.50±0.01	94.20±0.05
	IPS=5	**34.79±0.04**	**93.16±0.19**	**35.48±0.05**	**94.57±0.03**	**34.52±0.01**	**94.30±0.01**

forms them and obtains the best results across all settings. As NRI increases, all methods show improved performance. Remarkably, our approach reaches up to 97% of the performance obtained using the full dataset in terms of both PSNR and SSIM. This improvement can be attributed to the fact that our method preserves subject-specific information through subject-aware optimization and gradient alignment, as described in the previous section. The qualitative results in Fig. 4 further show that our method produces images with improved visual fidelity and better anatomical detail preservation.

5.3 Generalization Evaluation

Cross-Architecture Experiment. We conduct cross-architecture experiments to evaluate whether the distilled dataset can generalize across different downstream models. Specifically, the dataset is distilled using SRCNN [6] and then used to train and evaluate different architectures, including SRCNN, EDSR [11], and SCTSRN [29], under the NRI=10 setting. Quantitative results are reported in Tab. 2. Our method achieves the best performance when using EDSR, reaching a PSNR of 35.48dB and an SSIM of 94.57, even outperforming the results obtained with SRCNN. These results indicate that the distilled dataset captures transferable information from the raw dataset and does not depend on a specific network architecture.Overall, the results demonstrate strong cross-architecture generalization capability.

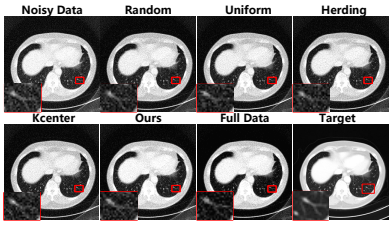

Fig. 5: Qualitative results of different algorithms.

Table 3: Quantitative restoration results of different methods trained on REDCNN for the CT modality.

		NRI=5/CR=99.67%		NRI=10/CR=99.34%	
Methods		PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑
Full data		PSNR=28.90±0.01		SSIM=58.08±0.05	
Base	Random	28.06±0.01	56.54±0.06	28.10±0.02	56.75±0.12
	Uniform	28.04±0.01	57.06±0.06	28.07±0.00	57.06±0.06
	Herding	28.05±0.01	56.74±0.07	28.10±0.02	56.75±0.17
	K-Center	28.05±0.00	56.56±0.04	28.11±0.05	56.78±0.32
Ours	IPS=1	28.00±0.07	57.20±0.62	28.09±0.03	56.36±0.42
	IPS=5	**28.38±0.08**	**57.45±0.63**	**28.38±0.14**	**57.43±0.94**

Image Restoration Tasks. To further verify the generality of our method, we conduct experiments on the CT restoration task using REDCNN [3]. As shown in Tab. 3, our method achieves the best restoration performance and is comparable to training with the full dataset. The quantitative results in Fig. 5 further show that models trained on the distilled dataset produce higher-quality reconstructions and preserve more fine-grained anatomical details.

Large-Scale Dataset Experiment. To evaluate scalability, we further conduct experiments on a large-scale dataset containing 48 subjects from the CT dataset [16], while two additional subjects are used for testing. Since computing gradients over all subjects simultaneously is computationally expensive, we perform gradient accumulation over every 5 subjects before updating the synthetic samples. This process is repeated until all subjects are processed. As shown in Tab. 4, under the same NRI setting, our method achieves a PSNR of 36.08 dB and an SSIM of 94.94, which is closest to full-dataset training performance. This result demonstrates that our method can scale to larger datasets while maintaining strong performance.

Table 4: Quantitative super-resolution results of different methods based on the large-scale dataset.

Methods		NRI=5/CR=99.92%		NRI=10/CR=99.85%	
		PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑
Full Data		PSNR=36.95±0.02		SSIM=95.58±0.02	
Base	Random	31.27±0.23	87.79±0.57	32.65±0.34	91.64±0.29
	Random*	31.81±0.18	89.26±0.67	31.47±0.12	88.45±0.74
	Uniform	31.21±0.15	87.22±0.98	32.95±0.12	91.54±0.75
	Herding	31.55±0.20	88.03±1.04	32.74±0.22	91.66±0.29
	K-Center	31.75±0.45	89.13±0.84	32.76±0.28	91.74±0.36
Ours	IPS=1	34.80±0.39	92.37±2.17	34.91±0.18	92.91±1.54
	IPS=5	**36.08±0.06**	**94.94±0.05**	**36.07±0.07**	**95.00±0.08**

5.4 Ablation Study and Analysis

Hyper-parameter Verification. To evaluate the impact of the subject-specific adjustment code dimensions, we conduct hyperparameter studies as shown in Fig. 6. The left and right subfigures correspond to NRI=5 and NRI=10, respectively. The horizontal axis denotes the dimension of the subject-specific code, and the vertical axis reports PSNR and SSIM. The results show that a 2-dimensional code consistently achieves the best and most stable performance. Therefore, we set the default dimension to 2 in all experiments.

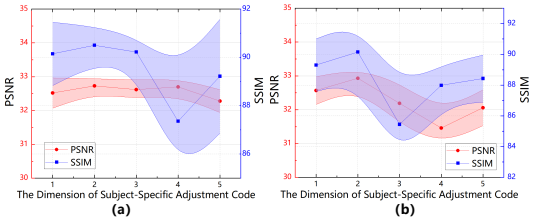

Fig. 6: Parameter validation experiments for the dimension of the subject-specific adjustment code. (a) and (b) denote the results under NRI=5 and NRI=10, respectively.

Ablation Study. We evaluate the effectiveness of each component in SPG. "Ours†" refers to our method initialized with random noise, while "Ours‡" removes the pixel-level fidelity preservation step. As shown in Tab. 5, both components contribute to performance improvement. In particular, the full model

achieves strong compression ratios (up to nearly 99%) while maintaining performance close to full-data training. Additional ablation results are provided in the ***Appendix***.

Table 5: Ablation study of our method. The reduction rate (the ratio of compressed to raw storage size) is provided below the storage size.

Methods	IPS=1			IPS=5		
	PSNR (dB)↑	SSIM↑	Storage↓	PSNR (dB)↑	SSIM↑	Storage↓
Full Data	PSNR=36.10±0.02		SSIM=94.86±0.03		Storage=39.9MB	
Ours†	11.54	45.79	817KB	11.30	45.55	4085KB
	±0.28	±0.36	−98.00%	±0.19	±0.23	−90.00%
Ours‡	16.95	51.57	410KB	30.55	88.11	412KB
	±1.40	±1.83	−98.99%	±0.06	±0.16	−98.99%
Ours NRI=5	32.73	**90.50**	410KB	34.45	92.07	412KB
	±0.22	±0.78	−98.99%	±0.20	±0.93	−98.99%
Ours NRI=10	**32.93**	90.15	819KB	**34.57**	**92.83**	822KB
	±0.28	±1.90	−97.99%	±0.14	±0.28	−97.98%

Visualization of Distilled Data. We present examples of distilled data in Fig. 7. Under NRI=10 and IPS=1, the synthesized samples for CT and MRI show that the distilled data capture shared anatomical structures while preserving inter-subject variability. This helps reduce the need to store raw patient data.

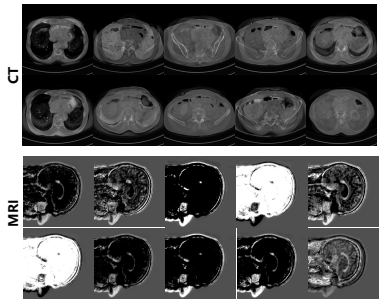

Fig. 7: Distillation results of our methods. Each image represents the synthetic data of a subject.

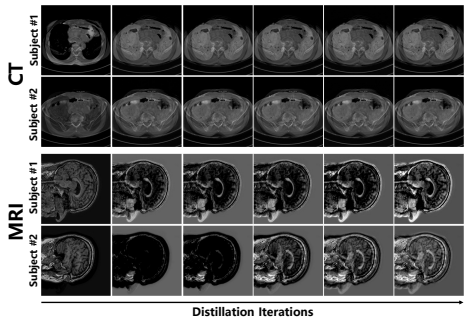

Fig. 8: Visualization of the distillation process.

Visualization of Distillation Process. We further visualize the evolution of distilled samples across optimization iterations in Fig. 8. The results show that the synthetic data gradually converge to stable structures during training, indicating that subject-specific information is effectively encoded into the distilled dataset. Notably, throughout the entire process, no subject data is explicitly transmitted to the user, thereby preserving privacy. Additional privacy-related experiments can be found in our ***Appendix***.

Initialization Strategies. Our method employs a shared anatomical prior derived from a representative subject as initialization. To validate the effectiveness

of this strategy, we compare several initialization schemes. Specifically, we separately use subject #1 and subject #3 as representative subjects to evaluate the robustness of our method to the choice of anatomical prior. In addition, we include two baseline initialization strategies:

- *Random:* the distilled data are initialized with randomly sampled noise, without any anatomical or structural prior;
- *Average:* the mean image, aligned and computed across all subjects, is used as a shared structural prior.

For the Average baseline, we align imaging volumes across subjects according to their slice indices and compute a slice-wise mean to obtain the average prior.

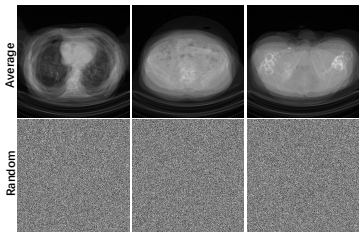

Fig. 9: The initialization examples with different strategies.

Table 6: Quantitative super-resolution results with different initialization strategies.

	Methods	NRI=5/CR=99% PSNR (dB)↑	NRI=5/CR=99% SSIM↑	NRI=10/CR=98% PSNR (dB)↑	NRI=10/CR=98% SSIM↑
IPS=1	Random	18.41±0.84	54.50±1.13	18.84±1.32	55.38±2.00
	Average	27.20±8.43	73.60±20.49	25.52±8.77	72.01±18.13
	Ours w/ subject #1	32.73±0.22	90.50±0.78	32.93±0.28	90.15±1.90
	subject #3	32.26±0.32	88.55±2.02	32.02±0.28	88.91±0.93
IPS=5	Random	31.84±2.37	81.23±8.99	28.35±4.69	75.28±12.16
	Average	30.29±0.27	86.10±0.53	30.66±0.30	87.36±0.91
	Ours w/ subject #1	34.45±0.20	92.07±0.93	34.57±0.14	92.83±0.28
	subject #3	34.27±0.17	92.00±0.41	34.18±0.19	91.40±0.47

For the CT super-resolution task, the results under different initialization strategies are shown in Tab. 6. Using a single representative subject as the shared anatomical prior yields the best performance. Besides, the Random initialization fails to converge under limited data conditions, as it does not provide any structural constraint for optimization. Furthermore, the Average baseline suffers from significant anatomical variability across subjects. Directly averaging different volumes blurs critical structural boundaries and limits its effectiveness as a structural prior. The initialization examples in Fig. 9 further show that both the Random and Average baselines fail to preserve meaningful structural details.

Sampling Strategies Analysis. We investigate the impact of different sampling strategies from a single subject when constructing the shared anatomical prior, as shown in Tab. 7. Under high compression ratios, consecutive sampling tends to focus on local anatomical regions and provides limited coverage of diverse anatomical structures. In contrast, random sampling offers broader anatomical coverage and achieves better distillation performance. Therefore, we adopt random sampling as the default strategy.

Images Per Subject (IPS) Analysis. To further explore the performance upper bound under larger IPS settings, we conduct additional experiments, with results reported in Tab. 7. Larger IPS preserves more information and leads to improved performance, but it also introduces higher computational and storage costs. This indicates a trade-off between efficiency and performance.

Table 7: Experiments under different sampling strategies analysis and IPS settings.

Results	Adjacent	Spaced(Step=3)	Spaced(Step=5)	Spaced(Step=10)	Random(ours)	IPS=1	IPS=5	IPS=10	IPS=15
PSNR	26.24±0.13	31.57±0.09	31.84±0.34	31.89±0.16	**32.73±0.22**	32.73±0.22	34.45±0.20	34.56±0.15	34.77±0.11
SSIM	64.85±0.33	88.20±0.21	88.90±0.45	89.81±0.53	**90.50±0.78**	90.50±0.78	92.07±0.93	92.52±0.30	92.68±0.42

6 Conclusion

We propose the first low-level DD framework for medical image enhancement. Our approach leverages anatomical similarity to construct a shared prior, which is further personalized through a learnable modulation module to generate subject-specific distilled data. Task-specific high–low quality pair construction and subject-aware gradient alignment ensure that both task-relevant information and subject-level structural characteristics are preserved without requiring access to raw data. Extensive experiments across multiple modalities, tasks, and network architectures demonstrate that our method achieves strong enhancement performance using a compact synthetic dataset, providing an efficient and privacy-preserving solution for medical image enhancement. For future work, we will extend the proposed framework to additional imaging modalities and more diverse low-level tasks, and further investigate its applicability in broader data-efficient learning scenarios.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under 62271335, in part by the Sichuan Science and Technology Program under Grant 2025ZNSFSC0470, in part by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-RP-2024-033), and in part by the Japan Science and Technology Agency (JST) and the Agency for Science, Technology and Research (A*STAR) under the Japan-Singapore Joint Call (Project No. R24I6IR133)

References

1. Ahmad, W., Ali, H., Shah, Z., Azmat, S.: A new generative adversarial network for medical images super resolution. Scientific Reports **12**(1), 9533 (2022)
2. Cazenavette, G., Wang, T., Torralba, A., Efros, A.A., Zhu, J.Y.: Dataset distillation by matching training trajectories. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4750–4759 (2022)
3. Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging **36**(12), 2524–2535 (2017)
4. Cui, J., Wang, R., Si, S., Hsieh, C.J.: Scaling up dataset distillation to imagenet-1k with constant memory. In: International Conference on Machine Learning. pp. 6565–6590. PMLR (2023)
5. Deng, Z., Russakovsky, O.: Remember the past: Distilling datasets into addressable memories for neural networks. Advances in Neural Information Processing Systems **35**, 34391–34404 (2022)

6. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38**(2), 295–307 (2015)
7. Farahani, R.Z., Hekmatfar, M.: Facility location: concepts, models, algorithms and case studies. Springer Science & Business Media (2009)
8. Kumar, R.R., Priyadarshi, R.: Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches. Multimedia Tools and Applications **84**(12), 10817–10875 (2025)
9. Lei, S., Tao, D.: A comprehensive survey of dataset distillation. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(1), 17–32 (2023)
10. Li, Y., Sixou, B., Peyrin, F.: A review of the deep learning methods for medical images super resolution problems. Irbm **42**(2), 120–133 (2021)
11. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
12. Loo, N., Hasani, R., Amini, A., Rus, D.: Efficient dataset distillation using random feature approximation. Advances in Neural Information Processing Systems **35**, 13877–13891 (2022)
13. Lu, H., Mei, J., Qiu, Y., Li, Y., Hao, F., Xu, J., Tang, L.: Information sparsity guided transformer for multi-modal medical image super-resolution. Expert Systems with Applications **261**, 125428 (2025)
14. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
15. Mahapatra, D., Bozorgtabar, B., Garnavi, R.: Image super-resolution using progressive generative adversarial networks for medical image analysis. Computerized Medical Imaging and Graphics **71**, 30–39 (2019)
16. McCollough, C.: Tu-fg-207a-04: overview of the low dose ct grand challenge. Medical physics **43**(6Part35), 3759–3760 (2016)
17. Nguyen, T., Chen, Z., Lee, J.: Dataset meta-learning from kernel ridge-regression. arXiv preprint arXiv:2011.00050 (2020)
18. Nguyen, T., Novak, R., Xiao, L., Lee, J.: Dataset distillation with infinitely wide convolutional networks. Advances in Neural Information Processing Systems **34**, 5186–5198 (2021)
19. Qiu, T., Wen, C., Xie, K., Wen, F.Q., Sheng, G.Q., Tang, X.G.: Efficient medical image enhancement based on cnn-fbb model. IET Image Processing **13**(10), 1736–1744 (2019)
20. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018), `https://openreview.net/forum?id=H1aIuk-RW`
21. Souza, R., Lucena, O., Garrafa, J., Gobbi, D., Saluzzi, M., Appenzeller, S., Rittner, L., Frayne, R., Lotufo, R.: An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. NeuroImage **170**, 482–494 (2018)
22. Umirzakova, S., Ahmad, S., Khan, L.U., Whangbo, T.: Medical image super-resolution for smart healthcare applications: A comprehensive survey. Information Fusion **103**, 102075 (2024)
23. Wang, K., Zhao, B., Peng, X., Zhu, Z., Yang, S., Wang, S., Huang, G., Bilen, H., Wang, X., You, Y.: Cafe: Learning to condense dataset by aligning features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12196–12205 (2022)

24. Wang, T., Cao, Y., Lu, Z., Huang, Y., Lu, J., Fan, F., Shan, H., Zhang, Y.: Smart: Self-supervised learning for metal artifact reduction in computed tomography using range null space decomposition. IEEE Transactions on Medical Imaging pp. 1–1 (2025). https://doi.org/10.1109/TMI.2025.3616003
25. Wang, T., Zhu, J.Y., Torralba, A., Efros, A.A.: Dataset distillation. arXiv preprint arXiv:1811.10959 (2018)
26. Welling, M.: Herding dynamical weights to learn. In: Proceedings of the 26th Annual International Conference on Machine Learning. p. 1121–1128. ICML '09, Association for Computing Machinery, New York, NY, USA (2009). https://doi.org/10.1145/1553374.1553517, https://doi.org/10.1145/1553374.1553517
27. Yang, H., Wang, Z., Liu, X., Li, C., Xin, J., Wang, Z.: Deep learning in medical image super resolution: a review. Applied Intelligence **53**(18), 20891–20916 (2023)
28. Yang, Z., Chen, Y., Wang, Z., Shan, H., Chen, Y., Zhang, Y.: Patient-level anatomy meets scanning-level physics: Personalized federated low-dose ct denoising empowered by large language model. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5154–5163 (2025)
29. Yu, H., Liu, D., Shi, H., Yu, H., Wang, Z., Wang, X., Cross, B., Bramler, M., Huang, T.S.: Computed tomography super-resolution using convolutional neural networks. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 3944–3948. IEEE (2017)
30. Yu, R., Liu, S., Wang, X.: Dataset distillation: A comprehensive review. IEEE transactions on pattern analysis and machine intelligence **46**(1), 150–170 (2023)
31. Zhao, B., Bilen, H.: Dataset condensation with differentiable siamese augmentation. In: International Conference on Machine Learning. pp. 12674–12685. PMLR (2021)
32. Zhao, B., Bilen, H.: Dataset condensation with distribution matching. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6514–6523 (2023)
33. Zhao, B., Mopuri, K.R., Bilen, H.: Dataset condensation with gradient matching. In: Ninth International Conference on Learning Representations 2021 (2021)