

# High-Order Progressive Trajectory Matching for Medical Image Dataset Distillation

Le Dong<sup>1</sup>, Jinghao Bian<sup>1\*</sup>, Jingyang Hou<sup>2</sup>, Jingliang Hu<sup>2</sup>, Yilei Shi<sup>2</sup>,  
Weisheng Dong<sup>1</sup>, Xiao Xiang Zhu<sup>3</sup>, and Lichao Mou<sup>2(✉)</sup>

<sup>1</sup> Xidian University, Xi'an, China

<sup>2</sup> MedAI Technology (Wuxi) Co. Ltd., Wuxi, China

[lichao.mou@medimagingai.com](mailto:lichao.mou@medimagingai.com)

<sup>3</sup> Technical University of Munich, Munich, Germany

**Abstract.** Medical image analysis faces significant challenges in data sharing due to privacy regulations and complex institutional protocols. Dataset distillation offers a solution to address these challenges by synthesizing compact datasets that capture essential information from real, large medical datasets. Trajectory matching has emerged as a promising methodology for dataset distillation; however, existing methods primarily focus on terminal states, overlooking crucial information in intermediate optimization states. We address this limitation by proposing a shape-wise potential that captures the geometric structure of parameter trajectories, and an easy-to-complex matching strategy that progressively addresses parameters based on their complexity. Experiments on medical image classification tasks demonstrate that our method improves distillation performance while preserving privacy and maintaining model accuracy comparable to training on the original datasets. Our code is available at <https://github.com/Bian-jh/HoP-TM>.

**Keywords:** patient privacy · dataset distillation · trajectory matching · image classification.

## 1 Introduction

Deep learning has revolutionized medical image analysis, achieving unprecedented performance across diverse tasks. However, developing deep networks in the medical domain faces a critical challenge: the scarcity of large-scale, diverse training datasets. This limitation arises from multiple domain-specific constraints, including stringent privacy regulations [14,19] and complex data sharing protocols across healthcare institutions [24,7]. While approaches like federated learning [18,25,21] address privacy concerns, they demand substantial computational resources and intricate coordination among participating institutions.

Dataset distillation offers a promising solution to these challenges by synthesizing a small set of training samples that encapsulate the essential information of a large-scale dataset [23]. This approach creates a compact, synthetic

---

\* Work done during an internship at MedAI Technology (Wuxi) Co. Ltd.

dataset that, when used for training, yields models with performance comparable to those trained on the original, full dataset. In the medical imaging context, dataset distillation not only addresses privacy concerns by avoiding direct data sharing but also significantly reduces storage requirements and computational costs associated with model training [10,11,12].

Recent years have witnessed the emergence of various algorithmic frameworks for dataset distillation, including gradient matching methods that align gradients produced by synthetic and real data [8,28], kernel inducing points using neural tangent kernel ridge regression [15,16], and distribution matching approaches that minimize the distance between synthetic and real data distributions [22,27,29]. Among these, trajectory matching [1,2,5,13] has demonstrated superior performance. This approach leverages pre-recorded expert trajectories from models trained on real data and optimizes synthetic examples to reproduce similar parameter trajectories.

Despite its promise, existing trajectory matching methods primarily focus on matching terminal states, essentially comparing only the destinations of parameter trajectories [1,2,5,13]. This simplified approach overlooks crucial information embedded in intermediate optimization states. To address this limitation, we propose matching trajectories with a high-order potential. Specifically, we introduce a shape-wise potential that captures the geometric structure of parameter trajectories, enabling more comprehensive matching.

Moreover, we observe that model parameters exhibit varying degrees of matching complexity during optimization. Building on this insight, we develop an easy-to-complex matching strategy that progressively addresses parameter matching based on complexity. This approach initially matches simpler parameters before advancing to more challenging ones, resulting in robust and effective trajectory matching.

Our contributions are three-fold:

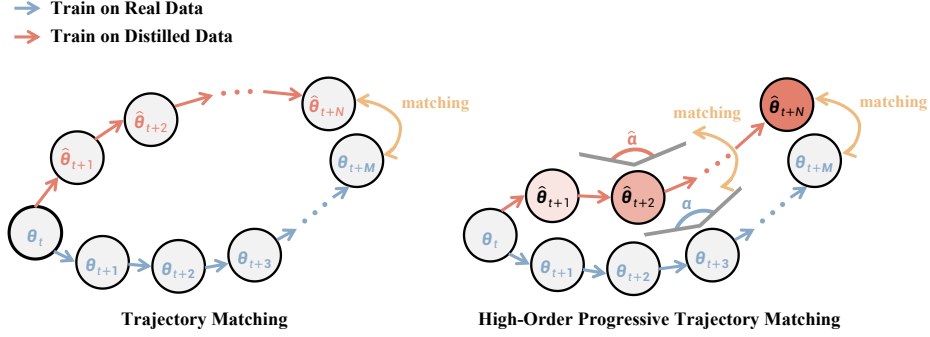
- We propose a shape-wise potential that enables high-order trajectory matching, capturing rich geometric relationships between parameter trajectories.
- We introduce an easy-to-complex matching strategy that enhances trajectory matching by considering parameter-specific characteristics.
- A thorough investigation of dataset distillation in medical imaging, with extensive experiments demonstrating state-of-the-art performance.

## 2 Method

### 2.1 Preliminary

Trajectory matching methods [1,2,5,13] first use multiple teacher networks trained for  $T$  epochs on a real dataset to generate a set of expert trajectories  $\{\tau\}$ , where each trajectory  $\tau$  represents a time sequence of parameters  $\{\theta_t\}_0^T$ .

At each distillation iteration, we randomly sample start parameters  $\theta_t$  and target parameters  $\theta_{t+M}$  ( $t + M \leq T$ ) from an expert trajectory. A student network, initialized with  $\theta_t$ , is then trained on a synthetic dataset  $\mathcal{D}_s$  for  $N$



**Fig. 1.** Comparison of trajectory matching approaches. Left: Conventional trajectory matching method. Right: Our proposed high-order progressive trajectory matching approach.

steps to generate a student trajectory  $\{\hat{\theta}_t\}_t^{t+N}$ . This training process minimizes the cross-entropy loss  $\mathcal{L}_{ce}$ :

$$\hat{\theta}_{t+i+1} = \hat{\theta}_{t+i} - \alpha \nabla \mathcal{L}_{ce}(\mathcal{D}_s; \hat{\theta}_{t+i}), \quad (1)$$

where  $\alpha$  is the learning rate.

The synthetic dataset  $\mathcal{D}_s$  is optimized by minimizing the following matching loss:

$$\mathcal{L}_{tm} = \frac{\|\hat{\theta}_{t+N} - \theta_{t+M}\|_2^2}{\|\theta_{t+M} - \theta_t\|_2^2}. \quad (2)$$

## 2.2 High-Order Trajectory Matching

Current trajectory matching approaches for dataset distillation focus solely on comparing terminal states between expert and student trajectories, potentially missing valuable information about how parameters evolve during training. We propose matching trajectories with a high-order potential, which captures the geometric structure of parameter trajectories to synthesize datasets that better emulate the learning dynamics of training on the original data.

Given sampled start parameters  $\theta_t$  and target parameters  $\theta_{t+M}$ , we first obtain intermediate state parameters at training epoch  $t + \lfloor \frac{M}{2} \rfloor$ , denoted as  $\theta_{t+\lfloor \frac{M}{2} \rfloor}$ , where  $\lfloor \cdot \rfloor$  represents the mathematical floor function. Similarly, we compute intermediate student parameters  $\hat{\theta}_{t+\lfloor \frac{N}{2} \rfloor}$ . To capture trajectory geometry, we introduce a high-order potential function that measures the angle formed by parameters at three time steps:

$$\begin{aligned} \psi(\theta_i, \theta_j, \theta_k) &= \cos \angle \theta_i \theta_j \theta_k = \langle e^{ij}, e^{kj} \rangle, \\ \text{where } e^{ij} &= \frac{\theta_i - \theta_j}{\|\theta_i - \theta_j\|_2}, e^{kj} = \frac{\theta_k - \theta_j}{\|\theta_k - \theta_j\|_2}. \end{aligned} \quad (3)$$

We then match geometric structures of real-data and synthetic-data trajectories using:

$$\mathcal{L}_{\text{hotm}} = \mathcal{L}_1(\psi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+\lfloor \frac{M}{2} \rfloor}, \boldsymbol{\theta}_{t+M}), \psi(\boldsymbol{\theta}_t, \hat{\boldsymbol{\theta}}_{t+\lfloor \frac{N}{2} \rfloor}, \hat{\boldsymbol{\theta}}_{t+N})), \quad (4)$$

where  $\mathcal{L}_1$  is a smooth  $\ell_1$  loss that is less sensitive to outliers [4].

### 2.3 Easy-to-Complex Matching

While high-order trajectory matching improves geometric alignment, we observe that parameter matching difficulty varies significantly across different network components. Some student parameters align easily with their teacher counterparts, while others require more complex optimization. To address this challenge, we propose an adaptive optimization strategy that progressively matches parameters from easy to complex, thereby improving the quality of the synthesized dataset.

Formally, we define per-parameter losses  $\ell_i$  derived from  $\mathcal{L}_{\text{tm}}$  (see Eq. (2)) and introduce binary variables  $\mathbf{v} = [v_1, v_2, \dots, v_K]$ , where  $K$  represents the number of parameters to be matched.  $\ell_i$  preserves the functional form of  $\mathcal{L}_{\text{tm}}$  but applies it at the individual parameter level rather than across the entire parameter set. The synthetic dataset  $\mathcal{D}_s$  and binary variables are jointly optimized by minimizing:

$$\min_{\mathcal{D}_s, \mathbf{v}} \mathbb{E}(\mathcal{D}_s, \mathbf{v}, \kappa) = \sum_{i=1}^K v_i \ell_i - \kappa \sum_{i=1}^K v_i, \text{ s.t. } v_i \in [0, 1], \quad (5)$$

where  $\kappa$  is a threshold parameter for controlling learning pace. Eq. (5) indicates that each parameter matching loss  $\ell_i$  is discounted by a weight  $v_i$ . We use alternative convex search to solve Eq. (5). Specifically, at each distillation iteration,  $\mathcal{D}_s$  and  $\mathbf{v}$  are optimized alternately. When  $\mathcal{D}_s$  is fixed, the optimal  $\mathbf{v}^*$  can be determined by:

$$v_i^* = \begin{cases} 1, & \text{if } \ell_i < \kappa, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

This means that parameters yielding loss values below  $\kappa$  are classified as easy parameters, while those exceeding  $\kappa$  exhibit high matching complexity.

During the optimization of  $\mathcal{D}_s$  with fixed  $\mathbf{v}$ , only easy parameters are selected, and their corresponding loss values contribute to the update of  $\mathcal{D}_s$ .

As the distillation progresses, we incrementally increase  $\kappa$  every  $P$  iterations by a growth factor  $\mu$  ( $\mu > 1$ ):

$$\kappa = \kappa_{\text{base}} \mu^{\lfloor \frac{i}{P} \rfloor}. \quad (7)$$

This gradual increase in  $\kappa$  enables the incorporation of increasingly complex parameters into the optimization of the synthetic dataset.

**Algorithm 1:** Pseudocode of Our Method

---

**Input:**  $\{\tau\}$ : Set of expert parameter trajectories  
 $M$ : Number of updates between initial and target expert parameters  
 $N$ : Number of student network updates per distillation iteration  
 $T^-, T^+$ : Sampling epoch range  
 $\mu$ : Growing factor  
 $P$ : Number of iterations for threshold  $\kappa$  increment.  
Initialize synthetic dataset  $\mathcal{D}_s$  with random samples  
Initialize threshold  $\kappa = \kappa_{\text{base}}$   
Initialize latent weights  $\mathbf{v} = \mathbf{0}$   
**for** each distillation iteration **do**  
    Sample trajectory  $\tau \in \{\tau\}$  with  $\tau = \{\theta_t\}_0^T$   
    Sample starting epoch where  $T^- \leq t \leq T^+$   
    Initialize student network  $\hat{\theta}_t = \theta_t$   
    **for**  $i = 0$  to  $N - 1$  **do**  
        Update student network:  
         $\hat{\theta}_{t+i+1} = \hat{\theta}_{t+i} - \alpha \nabla \mathcal{L}_{\text{ce}}(\mathcal{D}_s; \hat{\theta}_{t+i})$   
    **end**  
    Update  $\kappa$  using Eq. (7) and  $\mathbf{v}$  using Eq. (6)  
    Obtain intermediate parameters  $\theta_{t+\lfloor \frac{M}{2} \rfloor}$  and  $\hat{\theta}_{t+\lfloor \frac{N}{2} \rfloor}$   
    Compute loss with high-order matching using Eq. (8)  
    Update  $\mathcal{D}_s$  and learning rate  $\alpha$   
**end**  
**Output:** Distilled dataset  $\mathcal{D}_s$

---

## 2.4 Overall Loss

Combining our high-order trajectory matching with the easy-to-complex strategy, we optimize the synthetic dataset using the following loss:

$$\mathcal{L} = \sum_{i=1}^K v_i \ell_i - \kappa \sum_{i=1}^K v_i + \lambda \mathcal{L}_{\text{hotm}}, \quad (8)$$

where  $\lambda$  is a balancing hyper-parameter. This unified loss function enables our method to simultaneously leverage geometric trajectory information while adaptively focusing on manageable parameters during optimization. The pseudocode of our method is shown in Algorithm 1.

## 3 Experiments

### 3.1 Datasets

We evaluate our method on two public medical datasets: PathMNIST [26] and COVID19-CXR [17]. PathMNIST contains 107,180 histopathology images of colon biopsies across 9 categories, split into 89,996/10,004/7,180 images for training, validation, and testing, respectively. COVID19-CXR comprises chest X-ray

**Table 1.** Performance comparison of different methods on the PathMNIST dataset.

	IPC				
	1	5	10	100	1000
Real Dataset	89.89 $\pm$ 0.49				
DM [27]	38.39 $\pm$ 4.39	62.85 $\pm$ 0.73	66.99 $\pm$ 1.04	82.04 $\pm$ 0.88	87.29 $\pm$ 0.59
IDM [29]	<b>50.39<math>\pm</math>0.53</b>	69.32 $\pm$ 1.62	72.74 $\pm$ 1.13	82.05 $\pm$ 0.89	87.51 $\pm$ 0.21
MTT [1]	29.84 $\pm$ 1.06	47.30 $\pm$ 0.37	60.74 $\pm$ 1.03	82.90 $\pm$ 0.47	87.73 $\pm$ 0.27
FTD [2]	29.36 $\pm$ 0.77	55.99 $\pm$ 1.02	62.06 $\pm$ 0.93	82.81 $\pm$ 0.88	87.65 $\pm$ 0.39
DATM [5]	45.74 $\pm$ 1.66	64.94 $\pm$ 1.01	73.18 $\pm$ 0.90	84.07 $\pm$ 1.03	89.15 $\pm$ 0.11
ATT [13]	48.42 $\pm$ 2.29	56.95 $\pm$ 1.75	68.92 $\pm$ 1.09	83.86 $\pm$ 0.67	88.41 $\pm$ 0.32
Ours	47.71 $\pm$ 1.22	<b>73.92<math>\pm</math>0.66</b>	<b>77.23<math>\pm</math>0.65</b>	<b>84.82<math>\pm</math>0.40</b>	<b>89.86<math>\pm</math>0.36</b>

**Table 2.** Performance comparison of different methods on the COVID19-CXR dataset.

	IPC			
	1	5	10	50
Real Dataset	91.49 $\pm$ 0.42			
DM [27]	50.85 $\pm$ 1.86	65.13 $\pm$ 0.92	72.68 $\pm$ 3.54	79.10 $\pm$ 1.52
IDM [29]	67.78 $\pm$ 0.92	68.51 $\pm$ 0.92	72.77 $\pm$ 0.82	79.28 $\pm$ 0.94
MTT [1]	56.98 $\pm$ 1.89	74.94 $\pm$ 1.74	83.30 $\pm$ 0.17	86.11 $\pm$ 0.52
FTD [2]	60.55 $\pm$ 2.08	74.14 $\pm$ 0.90	84.09 $\pm$ 0.46	86.96 $\pm$ 0.44
DATM [5]	68.46 $\pm$ 0.95	79.59 $\pm$ 0.34	82.11 $\pm$ 0.53	87.38 $\pm$ 0.31
ATT [13]	67.83 $\pm$ 2.02	73.80 $\pm$ 0.92	75.12 $\pm$ 0.72	87.62 $\pm$ 0.44
Ours	<b>69.71<math>\pm</math>0.87</b>	<b>80.81<math>\pm</math>0.38</b>	<b>84.42<math>\pm</math>0.26</b>	<b>87.71<math>\pm</math>0.24</b>

images with 3,616 COVID-19 cases, 6,012 lung opacity cases, 10,192 normal cases, and 1,345 viral pneumonia cases. We use an 8:2 train-test split for this dataset.

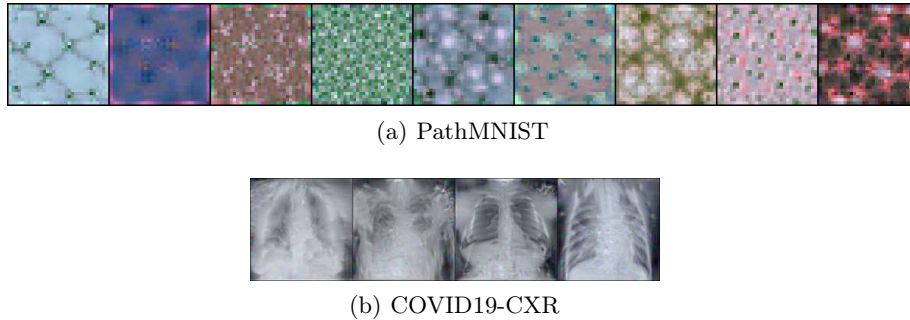
### 3.2 Evaluation Metrics

We evaluate distillation performance using classification accuracy. For each IPC (images per class) setting, we synthesize distilled datasets and train five networks with different random initializations, reporting the mean classification accuracy and standard deviation.

### 3.3 Implementation Details

Following prior work [27,1,2,5,13], we employ 3-layer and 5-layer ConvNets [3] as our distillation networks for PathMNIST and COVID19-CXR, respectively. We train 100 teacher networks to obtain pre-recorded expert trajectories. Moreover, we set  $P = 500$ .

For PathMNIST, we resize images to 32 $\times$ 32 pixels and set  $\kappa_{\text{base}} = 0.5$  and  $\mu = 1.5$ . We use  $\kappa_{\text{base}} = 5 \times 10^{-9}$  for IPC = 1, 5, 10 and  $\kappa_{\text{base}} = 5 \times 10^{-8}$  for IPC = 100, 1000. For COVID19-CXR, images are resized to 112 $\times$ 112 with  $\lambda = 1.0$  and  $\kappa_{\text{base}} = 1 \times 10^{-10}$ . We set  $\mu = 1.3$  for IPC = 1, 5, 10 and  $\mu = 1.6$  for IPC = 50. We follow [5] for the sampling parameters  $t$ ,  $M$ , and  $N$ . All experiments are conducted on NVIDIA RTX 4090 GPUs.



**Fig. 2.** Visualization of synthetic images generated through our method on two datasets with  $IPC = 1$ .

**Table 3.** Cross-architecture generalization results on the PathMNIST dataset with  $IPC = 10$ .

Method	Evaluation Model			
	ConvNet [3]	ResNet18 [6]	VGG11 [20]	AlexNet [9]
DM [27]	66.99 $\pm$ 1.04	61.66 $\pm$ 1.75	54.68 $\pm$ 1.38	44.30 $\pm$ 4.37
IDM [29]	72.74 $\pm$ 1.13	67.21 $\pm$ 0.75	57.72 $\pm$ 0.59	37.26 $\pm$ 2.75
MTT [1]	60.74 $\pm$ 1.03	58.67 $\pm$ 0.70	46.87 $\pm$ 3.04	42.44 $\pm$ 4.66
FTD [2]	62.06 $\pm$ 0.93	58.65 $\pm$ 1.46	47.33 $\pm$ 3.01	46.13 $\pm$ 5.10
DATM [5]	73.18 $\pm$ 0.90	60.36 $\pm$ 1.78	49.61 $\pm$ 3.21	47.67 $\pm$ 1.84
ATT [13]	68.92 $\pm$ 1.09	59.52 $\pm$ 1.39	48.97 $\pm$ 1.37	34.14 $\pm$ 1.77
Ours	<b>77.23<math>\pm</math>0.65</b>	<b>68.36<math>\pm</math>1.52</b>	<b>59.61<math>\pm</math>2.06</b>	<b>48.85<math>\pm</math>4.44</b>

### 3.4 Comparison with State-of-the-Art

As shown in Table 1, our method achieves superior performance across all settings on PathMNIST except  $IPC = 1$ . Notably, with  $IPC = 1000$  (approximately 10% of the original dataset size), our method approaches the accuracy achieved when training on the full dataset. On COVID19-CXR (Table 2), we consistently outperform all baselines across different  $IPC$  settings. The improvements over other trajectory matching-based methods demonstrate the effectiveness of our high-order potential and easy-to-complex matching strategy.

Fig. 2 visualizes distilled images from both datasets. The synthesized images exhibit highly abstract characteristics while maintaining discriminative features, effectively achieving data anonymization through information reduction.

### 3.5 Discussion

In this section, we investigate the generalization capability of our distilled images, analyze the effects of the proposed strategies, and examine the impact of hyperparameter choices.

**Cross-Architecture Generalization** We evaluate the generalization capability of our distilled images ( $IPC = 10$ ) on PathMNIST across three different

**Table 4.** Ablation study on the proposed components using the PathMNIST dataset. HO: high-order trajectory matching; E2C: easy-to-complex matching.

HO	E2C	IPC				
		1	5	10	100	1000
✓		29.84±1.06	47.30±0.37	60.74±1.03	82.90±0.47	87.73±0.27
		47.03±2.61	73.14±0.39	76.77±0.62	84.28±0.53	89.58±0.43
	✓	42.48±4.21	49.42±1.64	65.66±1.30	84.19±0.19	89.19±0.26
✓	✓	<b>47.71±1.22</b>	<b>73.92±0.66</b>	<b>77.23±0.65</b>	<b>84.82±0.40</b>	<b>89.86±0.36</b>

**Table 5.** Effect of hyperparameter  $\lambda$  on the PathMNIST dataset.

$\lambda$	IPC				
	1	5	10	100	1000
0	29.84±1.06	47.30±0.37	60.74±1.03	82.90±0.47	87.73±0.27
0.1	45.12±2.22	72.76±0.82	<b>77.04±0.42</b>	83.91±1.21	89.14±0.43
0.5	<b>47.03±2.61</b>	<b>73.14±0.39</b>	76.77±0.62	<b>84.28±0.53</b>	<b>89.58±0.43</b>
1	46.59±3.36	72.92±0.85	76.67±0.71	84.10±0.91	89.32±0.37
10	43.61±4.46	70.39±0.89	75.12±0.22	84.18±1.07	88.86±0.51

architectures. Results in Table 3 show that despite some performance degradation across architectures, our method consistently outperforms competing approaches, demonstrating strong generalization ability.

**Analysis of Matching Strategies** Table 4 presents ablation studies on our two key matching strategies. The high-order trajectory matching yields significant improvements over the baseline by capturing richer geometric patterns in parameter trajectories. The easy-to-complex matching strategy further enhances performance across all IPC settings. When both strategies are combined, we achieve the best results.

**Hyper-Parameter Study** We examine the impact of the high-order loss weight  $\lambda$ , excluding the easy-to-complex matching strategy to isolate its effect. As shown in Table 5, while different non-zero weights all improve performance compared to using only the original matching loss (Eq. (2)), an excessive weight (e.g.,  $\lambda = 10$ ) can dominate the distillation process and limit gains. Based on empirical results, we set  $\lambda = 0.5$  for experiments.

## 4 Conclusion

In this paper, we introduce a high-order progressive trajectory matching approach for medical image dataset distillation. Our method’s shape-wise potential captures geometric relationships in parameter trajectories, while the easy-to-complex matching strategy adaptively addresses parameters based on their optimization complexity. Experiments on various medical image classification tasks demonstrate that our approach achieves state-of-the-art performance in dataset distillation while maintaining privacy guarantees and model accuracy.



comparable to training on full datasets. This advancement enables efficient sharing of medical imaging data while respecting privacy constraints and reducing computational requirements. Future work could explore extending this approach to more complex medical imaging tasks such as segmentation and detection, as well as incorporating domain-specific medical knowledge into the distillation process. In addition, theoretical analysis of privacy guarantees could enhance the practical applicability of our approach in clinical settings.

**Acknowledgments.** The paper was supported by the Open Fund of Intelligent Control Laboratory 2024-ZKSYS-KF02-03.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this paper.

## References

1. 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)
2. Du, J., Jiang, Y., Tan, V.Y., Zhou, J.T., Li, H.: Minimizing the accumulated trajectory error to improve dataset distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3749–3758. (2023)
3. Gidaris, S., Komodakis, N.: Dynamic few-shot visual learning without forgetting. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4367–4375. (2018)
4. Girshick, R.: Fast R-CNN. In: IEEE International Conference on Computer Vision, pp. 1440–1448. (2015)
5. Guo, Z., Wang, K., Cazenavette, G., Li, H., Zhang, K., You, Y.: Towards loss-less dataset distillation via difficulty-aligned trajectory matching. arXiv preprint arXiv:2310.05773 (2023)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. (2016)
7. Jin, H., Luo, Y., Li, P., Mathew, J.: A review of secure and privacy-preserving medical data sharing. *IEEE Access* **7**, 61656–61669 (2019)
8. Kim, J.H., Kim, J., Oh, S.J., Yun, S., Song, H., Jeong, J., Ha, J.W., Song, H.O.: Dataset condensation via efficient synthetic-data parameterization. In: International Conference on Machine Learning, pp. 11102–11118. (2022)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1106–1114. (2012)
10. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Soft-label anonymous gastric x-ray image distillation. In: 2020 IEEE International Conference on Image Processing (ICIP), pp. 305–309. (2020)
11. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Compressed gastric image generation based on soft-label dataset distillation for medical data sharing. *Computer Methods and Programs in Biomedicine* **227**, 107189 (2022)
12. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Dataset distillation for medical dataset sharing. arXiv preprint arXiv:2209.14603 (2022)

13. Liu, D., Gu, J., Cao, H., Trinitis, C., Schulz, M.: Dataset distillation by automatic training trajectories. In: European Conference on Computer Vision, pp. 334–351. (2025)
14. McGraw, D., Dempsey, J.X., Harris, L., Goldman, J.: Privacy as an enabler, not an impediment: Building trust into health information exchange. *Health Affairs* **28**(2), 416–427 (2009)
15. Nguyen, T., Chen, Z., Lee, J.: Dataset meta-learning from kernel ridge-regression. arXiv preprint arXiv:2011.00050 (2020)
16. Nguyen, T., Novak, R., Xiao, L., Lee, J.: Dataset distillation with infinitely wide convolutional networks. In: Advances in Neural Information Processing Systems, pp. 5186–5198. (2021)
17. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Al Maadeed, S., Zughair, S.M., Khan, M.S., Chowdhury, M.E.: Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine* **132**, 104319 (2021)
18. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R.M., Trask, A., Xu, D., Baust, M., Cardoso, M.J.: The future of digital health with federated learning. *NPJ Digital Medicine* **3**(1), 1–7 (2020)
19. Shen, N., Bernier, T., Sequeira, L., Strauss, J., Silver, M.P., Carter-Langford, A., Wiljer, D.: Understanding the patient privacy perspective on health information exchange: A systematic review. *International Journal of Medical Informatics* **125**, 1–12 (2019)
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
21. Tian, Y., Wang, J., Jin, Y., Wang, L.: Communication-efficient federated skin lesion classification with generalizable dataset distillation. In: Medical Image Computing and Computer Assisted Intervention, pp. 14–24. (2023)
22. 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)
23. Wang, T., Zhu, J.Y., Torralba, A., Efros, A.A.: Dataset distillation. arXiv preprint arXiv:1811.10959 (2018)
24. Weitzman, E.R., Kaci, L., Mandl, K.D.: Sharing medical data for health research: The early personal health record experience. *Journal of Medical Internet Research* **12**(2), e1356 (2010)
25. Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J., Wang, F.: Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research* **5**, 1–19 (2021)
26. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2-a large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data* **10**(1), 41 (2023)
27. 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)
28. Zhao, B., Mopuri, K.R., Bilen, H.: Dataset Condensation with Gradient Matching. In: International Conference on Learning Representations, (2021)

29. Zhao, G., Li, G., Qin, Y., Yu, Y.: Improved distribution matching for dataset condensation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7856–7865. (2023)