This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Learning Task-Specific Anatomical Priors via Context Distillation for In-Context Medical Image Segmentation

Guoqing Yang[1,2⋆], Shishuai Hu[1,2⋆], and Yong Xia[1,2(✉)]

[1] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, Shaanxi, China
[2] Ningbo Institute of Northwestern Polytechnical University, Ningbo 315048, China
yxia@nwpu.edu.cn

Abstract. In-context learning (ICL) has recently emerged as a promising paradigm for medical image segmentation, enabling frozen pre-trained models to adapt to new tasks using only a few labeled demonstrations (context). However, the effectiveness of ICL critically depends on the quality and representativeness of the provided context set. Redundant, noisy, or suboptimal demonstrations not only degrade segmentation accuracy but also increase computational overhead during inference. In this work, we propose Context Distillation via Masked Image Modeling (CD-MIM), a framework that reformulates context construction as a task of learning compact and task-specific anatomical priors. Instead of relying solely on raw clinical samples, we treat the context as a learnable entity and distill it into a small set of optimized templates. Through a curriculum-driven masking strategy and a meta-learning-based template optimization process, CD-MIM encourages the model to encode essential structural dependencies and cross-patient anatomical variations into condensed context. The resulting distilled context can either replace or complement raw demonstrations, reducing redundancy while improving segmentation robustness. Extensive experiments on seven medical image segmentation tasks and three representative visual in-context learning backbones demonstrate consistent performance gains and improved efficiency. Our findings suggest that explicitly learning distilled context priors provides a scalable and principled solution for stabilizing and enhancing ICL-based medical image segmentation.

Keywords: Medical image segmentation · In-context learning · Dataset distillation.

1 Introduction

Visual In-Context Learning (VICL) has recently emerged as a promising paradigm in computer vision [2, 25, 26, 28]. Unlike conventional deep learning frameworks

⋆ G. Yang and S. Hu contributed equally. Corresponding author: Y. Xia.

that require task-specific fine-tuning, VICL enables a pre-trained and frozen model to adapt to new tasks by conditioning on a small set of labeled demonstrations, referred to as the context. This paradigm is particularly appealing for medical image segmentation, where expert annotations are scarce, anatomical variability is substantial, and maintaining separately fine-tuned models for diverse clinical scenarios is often impractical [8]. Recent studies, including UniverSeg [6], Tyche-TS [22], and related approaches [11, 26], have demonstrated that VICL offers a flexible and scalable alternative to conventional supervised medical image segmentation pipelines.

Despite its promise, the performance of VICL-based segmentation models is highly sensitive to the composition of the context set. Existing studies primarily focus on selecting suitable demonstrations for a given query [10, 29, 30] or improving context–query matching [13, 27]. While these strategies enhance performance, they implicitly assume a relatively small and fixed pool of candidate contexts from which demonstrations are selected. In real-world clinical scenarios, however, the number of available annotated samples continuously grows as more data are accumulated [21]. Consequently, the candidate context pool becomes increasingly large and heterogeneous. Under such circumstances, selecting a limited number of demonstrations, constrained by the model's context length, from a vast and expanding pool becomes computationally expensive and practically challenging. This raises a fundamental question: rather than repeatedly selecting from an ever-growing set of raw samples, can we instead learn compact context representations that distill essential anatomical priors required for the task?

Dataset distillation [16] offers a conceptual foundation for tackling this challenge. It seeks to compress a large training set into a small number of synthetic samples that retain task-relevant knowledge, typically via gradient matching, trajectory matching, or distribution alignment. In medical imaging, distillation has primarily been investigated in privacy-preserving and federated learning scenarios [9,15,17–19], with most efforts centered on classification tasks, while dense prediction problems such as segmentation remain largely underexplored. More importantly, existing distillation frameworks are not tailored to the in-context setting, where the distilled representations are expected to serve as context demonstrations during inference rather than as surrogate training data. This distinction calls for a specialized mechanism capable of encoding spatially structured anatomical priors into compact context templates that can effectively guide in-context segmentation.

In this work, we propose Context Distillation via Masked Image Modeling (CD-MIM), a framework that learns task-specific context templates for in-context medical image segmentation. We reconceptualize the context not as a static subset of labeled samples, but as a learnable and compact carrier of anatomical priors. Specifically, we introduce a curriculum-driven masking scheduler within a meta-learning-based template optimization process, encouraging the model to reconstruct masked regions progressively and thereby encode structural dependencies into the distilled templates. This design enables the context to capture cross-patient morphological variations while suppressing redundant

or noisy details. Unlike prior approaches that focus solely on context selection or matching, our method directly optimizes the context representations themselves. The resulting distilled templates can replace raw demonstrations, significantly reducing redundancy while stabilizing segmentation performance. We evaluate CD-MIM on seven medical image segmentation tasks across five imaging modalities using three representative VICL backbones, and experimental results demonstrate consistent improvements in query segmentation accuracy. The contributions of this work are three-fold: (1) We propose a context distillation framework tailored to in-context medical image segmentation, which learns compact task-specific anatomical priors instead of relying solely on raw demonstrations. (2) We introduce a curriculum-driven masked modeling strategy integrated with a meta-learning-based optimization procedure to progressively encode structural priors into distilled context templates. (3) We conduct comprehensive evaluations across seven segmentation tasks and three VICL backbones, demonstrating consistent and robust performance gains.

2 Method

2.1 Problem Definition

Let $\mathcal{M}$ denote a pre-trained and frozen VICL segmentation model. Given a context set $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{K}$ contains K demonstrations indicating the same segmentation task, where x_i is a medical image and y_i is its corresponding segmentation mask, the goal of VICL is to predict the segmentation mask $\hat{y^q}$ for a given query medical image x^q conditioned on $\mathcal{S}$:

$$\hat{y^q} = \mathcal{M}(x^q|\mathcal{S}). \tag{1}$$

Despite its potential, the performance of VICL is inherently sensitive to the quality and composition of $\mathcal{S}$, where redundant or noisy demonstrations often lead to sub-optimal accuracy and excessive computational overhead. To address this, the proposed CD is to transform the raw, unrefined context $\mathcal{S}$ into a set of highly condensed and discriminative task-specific templates. Inspired by Masked Image Modeling (MIM) [12], we compel the framework to distill essential spatial correlations and structural dependencies from the context. Unlike generic self-supervised pre-training, our method specifically distills task-relevant structural information into the template by minimizing the loss within the context set $\mathcal{S}$. An overview of the framework is illustrated in Fig. 1. We now delve into its details.

2.2 Context Distillation via Masked Image Modeling

Template Initialization. We initialize a context buffer $\mathcal{B} \subset \mathcal{S}$ via stochastic selection to ensure unbiased coverage of the anatomical distribution. As a potential enhancement to further stabilize the distillation, one may also consider a rule-based diversity selection strategy. Ultimately, the objective of initialization

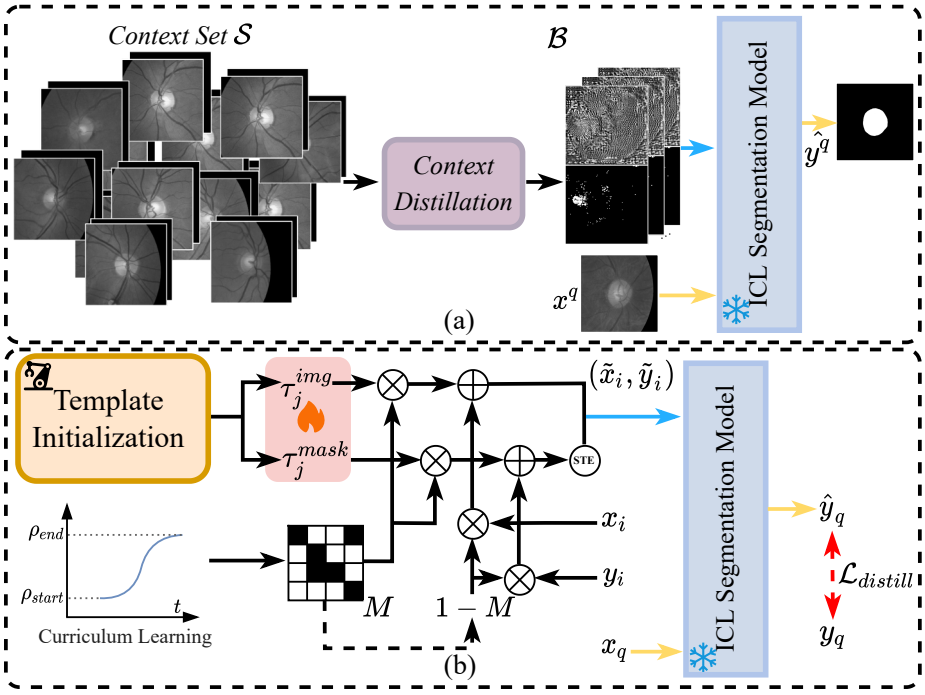

Fig. 1. (a) Overall pipeline. The context set $\mathcal{S}$ is distilled into compact, task-specific templates that guide query segmentation. (b) Context distillation module. The templates are optimized through a meta-learning framework. A curriculum-driven scheduler progressively increases the masking ratio ρ_t to enable gradual distillation. The STE preserves gradient flow for discrete labels. The loss $\mathcal{L}_{distill}$ optimizes only the templates to minimize the discrepancy between the query prediction and ground truth.

is to anchor the templates within the critical regions of the anatomical distribution. These anchors provide a reliable basis for the CD-MIM process, allowing the framework to successfully distill redundant clinical information into compact, stabilized structural representations.

Curriculum-driven Masking Scheduler. A key objective of context distillation is to ensure that the learnable templates capture essential anatomical priors rather than mere low-level textures. To achieve this, we introduce a curriculum-driven masking scheduler that regulates the information density of the support context as distillation progresses. Unlike standard Masked Image Modeling (MIM) which trains a model for self-supervised reconstruction, our scheduler serves as a progressive distillation constraint that compels the learnable templates to provide highly informative structural guidance under increasingly sparse conditions. Let ρ_t denote the masking ratio at training epoch t. To facilitate an easy-to-hard learning trajectory, ρ_t is formulated as a dynamic

pacing function:

$$\rho_t = \rho_{start} + (\rho_{end} - \rho_{start}) \cdot \eta(t), \tag{2}$$

where ρ_{start} and ρ_{end} represent the initial and terminal masking ratios, and $\eta(t) \in [0, 1]$ is a time-dependent scaling factor. In the early stages, a lower ρ_t allows the templates to align with the global geometric layout of the support samples. As training evolves, the increasing ρ_t reduces the visible portions of the raw support data, forcing the optimization process to encode dense structural dependencies and cross-patient morphological priors into the learnable templates $(\tau_j^{img}, \tau_j^{mask})$. This mechanism ensures that the final distilled context remains highly discriminative and effective even when the available support information is significantly constrained.

Meta-learning-based Template Optimization. During the distillation phase, the context buffer $\mathcal{B}$ is treated as a set of learnable templates. To optimize these templates, we adopt a meta-learning-based training procedure executed within $\mathcal{S}$. Specifically, in each iteration, a single sample $(x_q, y_q) \in \mathcal{S}$ is designated as the pseudo-query, while the remaining samples $\mathcal{S} \setminus \{(x_q, y_q)\}$ serve as the support context. To maintain computational efficiency, we implement a mini-batch training strategy coupled with gradient accumulation.

The model performs ICL-based segmentation using the masked support context $(\tilde{x}_i, \tilde{y}_i)$. The masking operation is formulated using the masking ratio ρ_t and learnable template:

$$\begin{aligned} \tilde{x}_i &= (1 - M) \odot x_i + M \odot \tau_j^{img}, \\ \tilde{y}_i &= \mathrm{STE}((1 - M) \odot y_i + M \odot \tau_j^{mask}), \end{aligned} \tag{3}$$

where M is a binary mask sampled with ratio ρ_t, $\odot$ denotes the element-wise product, and τ represents the learnable template. Since the segmentation masks y_i are discrete, the direct optimization of masked labels is non-differentiable. To circumvent this, we apply the Straight-Through Estimator (STE) [3] during the discretization of $\tilde{y}_i$, allowing gradients to flow back to the templates through the identity mapping in the backward pass.

The resulting prediction $\hat{y}_q$ is then compared against its ground truth y_q. The optimization objective $\mathcal{L}_{distill}$ is defined as the sum of Binary Cross-Entropy (BCE) and Dice loss:

$$\mathcal{L}_{distill} = \mathcal{L}_{BCE}(y_q, \hat{y}_q) + \mathcal{L}_{Dice}(y_q, \hat{y}_q). \tag{4}$$

The gradients derived from $\mathcal{L}_{distill}$ are back-propagated to update the learnable templates, while the parameters of $\mathcal{M}$ remain frozen. This refinement process ensures that the templates are transformed from raw, potentially redundant demonstrations into highly condensed and discriminative task-specific priors. By distilling representative anatomical features and structural dependencies, this mechanism effectively mitigates the sensitivity inherent in context selection. Furthermore, it inherently reduces the search complexity by diminishing the search space.

3 Experiments and Analysis

3.1 Experimental Setup

VICL Backbones. We evaluate the proposed CD-MIM framework using UniverSeg [6], Tyche-TS [22], and SegGPT [26] as VICL backbones. UniverSeg and Tyche-TS are VICL models built upon a cross-block design and trained on large-scale medical image segmentation datasets, enabling in-context segmentation across diverse medical tasks. In contrast, SegGPT adopts a Vision Transformer backbone and is pre-trained via masked image modeling on large-scale natural image segmentation data. Evaluating across these backbones allows us to assess the effectiveness of CD-MIM under different architectural designs and pre-training regimes.

Datasets and Evaluation Metrics. We evaluate CD-MIM on seven tasks across five modalities. Specifically, the evaluated datasets include GlaS (GLS, 165 images) [24], REFUGE (ROC and ROD, 1,200 images) [20], PanDental (PDM, 99 images) [1], ACDC (ALV, 1,377 images) [4], LGG (1,271 images) [5], and FH-PS-AOP (FPA, 4,000 images) [14]. For each dataset, samples are randomly split into context and query sets with a ratio of 3:7. Notably, none of the evaluated datasets are included in the training data of the above VICL backbone models. Segmentation performance is assessed using the Dice Similarity Coefficient (DSC). A higher DSC indicates better segmentation performance.

Implementation Details. All experiments are conducted using the officially released pre-trained VICL models. For UniverSeg and Tyche-TS, the context size is fixed to 1. Input images are resized to 128×128, converted to grayscale, and normalized to the intensity range $[0, 1]$. For SegGPT, the context size is also set to 1, while input images are resized to 448×448, converted to three-channel color images, and normalized using the standard mean and standard deviation. For the distillation process, we employ the Adam optimizer with a learning rate of 0.003 across 200 epochs. The curriculum pacing function $\eta(t)$ is implemented as a linear scheduler, while the masking bounds ρ_{start} and ρ_{end} are empirically configured to 0.25 and 0.75, respectively.

3.2 Experimental Results and Analysis

We evaluate the proposed CD-MIM against four representative baselines for establishing the in-context demonstration set from the candidate pool: RS [23], which randomly select subsets and is averaged over five runs for stability; Top-K [30], which selects the most effective pairs based on one-shot VICL performance; VPR [29], which retrieves nearest neighbors using a pre-trained CLIP encoder and cosine similarity; and Herding [7], a classic coreset selection method that chooses samples to best approximate the mean embedding of the entire distribution. Within these methods, RS and Herding serve as coreset selection baselines, while Top-K and VPR operate as task-level and sample-level selection strategies, respectively.

Table 1. Performance comparison (DSC%) across seven segmentation tasks. The best performance for each task is highlighted in **bold**, while the second best is indicated with underline.

$\mathcal{M}$	Methods	GLS	ROC	ROD	PDM	ALV	LGG	FPA	Average
UniverSeg	RS	42.71	27.31	51.87	65.89	40.03	15.29	38.20	40.19
	Top-K	**53.00**	37.82	61.48	72.03	51.91	**28.00**	53.84	51.15
	VPR	46.43	44.92	63.10	68.01	61.31	20.30	44.19	49.75
	Herding	41.64	31.31	53.72	68.06	44.34	5.64	42.92	41.09
	Ours	47.78	**54.85**	**81.76**	**88.15**	**69.76**	20.95	**82.22**	**63.64**
Tyche-TS	RS	52.47	37.65	60.34	85.22	58.08	23.99	60.07	53.97
	Top-K	**62.25**	53.37	71.86	88.39	75.11	37.35	75.07	66.20
	VPR	57.96	55.99	74.76	85.21	**82.48**	32.09	**93.78**	68.90
	Herding	52.64	45.51	65.62	88.41	61.21	8.63	61.96	54.85
	Ours	58.84	**57.96**	**84.80**	**90.73**	80.10	**45.88**	83.64	**71.71**
SegGPT	RS	71.56	61.67	85.15	89.88	77.69	46.05	68.87	71.55
	Top-K	**85.58**	72.22	89.47	**92.26**	**88.82**	**74.26**	80.12	**83.25**
	VPR	82.75	65.21	87.10	90.07	87.76	57.72	**91.40**	80.29
	Herding	80.08	67.26	86.17	91.75	80.05	16.98	72.75	70.72
	Ours	78.37	**82.78**	**92.33**	85.12	88.70	71.01	68.66	81.00

Table 2. Ablation study on template initialization. "Raw" denotes the performance using raw context without distillation, while "Ours" represents the performance using the distilled templates.

	UniverSeg	Tyche-TS	SegGPT
Raw	40.61	52.08	71.68
Ours	$63.64_{\uparrow 23.03}$	$71.71_{\uparrow 19.63}$	$81.00_{\uparrow 9.32}$

Table 3. Ablation study of masking strategies and curriculum bounds on UniverSeg.

Component	Configuration	DSC (%)
Baseline	w/o Masking	31.29
Fixed Ratio	0.25	48.15
	0.50	59.40
Curriculum Bounds	$[0.10, 0.50]$	56.75
	$[0.40, 0.60]$	62.73
	$[0.25, 0.75]$	63.64

As shown in Table 1, CD-MIM delivers strong performance across multiple settings, achieving peak average DSC scores of 63.64% for UniverSeg and 71.71% for Tyche-TS. For these medical-specific backbones, our method outperforms both selection-based and coreset-sampling alternatives on average. This indicates that distilled anatomical templates effectively capture the task manifold's core features more reliably than individual raw clinical samples. While the performance gain is substantial for UniverSeg and Tyche-TS, CD-MIM exhibits a moderate response on SegGPT, trailing slightly behind the Top-K. Nevertheless, CD-MIM still improves the raw SegGPT baseline by +9.32%, demonstrating its plug-and-play effectiveness. Ultimately, by encoding cross-patient structural dependencies into condensed learnable templates, our method bridges the performance gap inherent in small-context regimes, facilitating accurate segmentation even when the support information is limited to a single distilled demonstration.

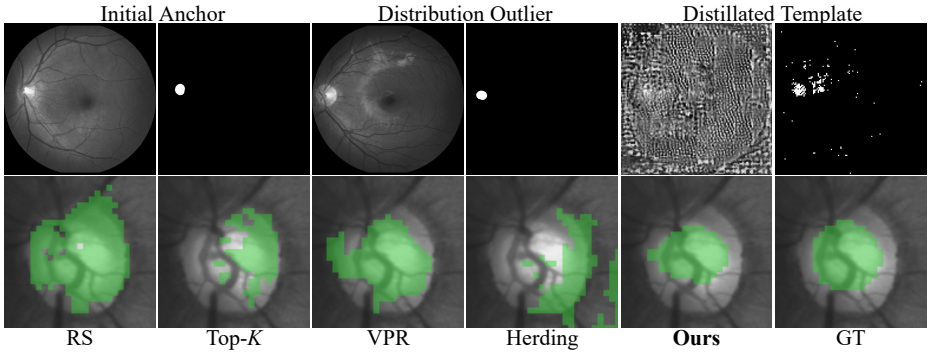

Fig. 2. Qualitative visualization. Top: Examples of two raw demonstrations and the distilled template. Bottom: Comparison on a challenging query image. CD-MIM achieves better boundary precision and structural consistency with the ground truth (GT) compared to other baselines.

3.3 Ablation Analysis

Effectiveness of Template Initialization. We evaluate the effectiveness of learned templates against their static "Raw" initialization, as summarized in Table 2. Transitioning to distilled templates yields drastic performance improvements across all backbones. Specifically, we observe a substantial DSC increase of 23.03% for UniverSeg and 19.63% for Tyche-TS, alongside a consistent gain of 9.32% for SegGPT. These results confirm that simply selecting raw context is suboptimal for medical ICL due to inherent information redundancy and cross-patient noise. In contrast, the proposed method acts as a systematic knowledge compression mechanism, successfully distilling expansive clinical data into compact, highly discriminative anatomical templates.

Masking Strategy and Curriculum Bounds. To isolate the proposed masking strategy, we conduct ablation experiments using UniverSeg. As shown in Table 3, removing the masking schedule drops DSC to 31.29%, confirming its contribution. Additionally, the dynamic curriculum outperforms fixed masking ratios of 0.25 and 0.50. We also test various curriculum bounds, including $[0.10, 0.50]$, $[0.40, 0.60]$, and $[0.25, 0.75]$. They achieved DSC scores of 56.75%, 62.73%, and 63.64%, respectively. This variance highlights the necessity of curriculum bounds regulation.

Qualitative Analysis. The quantitative benefits of CD-MIM are further supported by the qualitative results shown in Fig. 2. The first row demonstrates the transformation of disparate raw samples, including a initial anchor and a distribution outlier, into a single distilled template. The learned representation effectively distills task-relevant information into the template. Consequently, as shown in the second row, while other baselines like RS and VPR suffer from structural inaccuracies or over-segmentation on challenging queries, CD-MIM provides relatively precise anatomical guidance. The resulting masks exhibit better coherence and boundary alignment with the ground truth.

4 Conclusion

We present CD-MIM, a framework that distills raw clinical demonstrations into compact, task-specific anatomical templates for in-context medical image segmentation. By employing a curriculum-driven masking strategy, our method compresses cross-patient variations into stable priors, mitigating the inherent redundancy and sensitivity of VICL. Extensive benchmarking across seven tasks and three backbones demonstrates consistent improvements in segmentation fidelity and robustness, providing an efficient pathway for deploying in-context segmentation models within clinical workflows.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 92470101, in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang, China, under Grant 2025C01201(SD2), in part by the Project of National Key Clinical Specialty (Department of Medical Imaging, No. 2024017), in part by the Ningbo Key Laboratory of Digital Imaging and Medical-Engineering Interdisciplinarity (No. 2024031).

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

References

1. Abdi, A.H., Kasaei, S., Mehdizadeh, M.: Automatic segmentation of mandible in panoramic x-ray. Journal of Medical Imaging **2**(4), 044003–044003 (2015)
2. Bai, Y., Geng, X., Mangalam, K., Bar, A., Yuille, A.L., Darrell, T., Malik, J., Efros, A.A.: Sequential modeling enables scalable learning for large vision models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22861–22872 (2024)
3. Bengio, Y., Léonard, N., Courville, A.: Estimating or propagating gradients through stochastic neurons for conditional computation. arXiv preprint arXiv:1308.3432 (2013)
4. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)
5. Buda, M., Saha, A., Mazurowski, M.A.: Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. Computers in Biology and Medicine **109**, 218–225 (2019)
6. Butoi, V.I., Ortiz, J.J.G., Ma, T., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Universeg: Universal medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21438–21451 (2023)
7. Castro, F.M., Marín-Jiménez, M.J., Guil, N., Schmid, C., Alahari, K.: End-to-end incremental learning. In: European Conference on Computer Vision. pp. 233–248. Springer (2018)
8. Dissanayake, T., George, Y., Mahapatra, D., Sridharan, S., Fookes, C., Ge, Z.: Few-shot learning for medical image segmentation: A review and comparative study. ACM Computing Surveys **58**(1), 1–36 (2025)

9. Dong, L., Bian, J., Hou, J., Hu, J., Shi, Y., Dong, W., Zhu, X.X., Mou, L.: High-order progressive trajectory matching for medical image dataset distillation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 273–283. Springer (2025)
10. Gao, J., Lao, Q., Kang, Q., Liu, P., Du, C., Li, K., Zhang, L.: Boosting your context by dual similarity checkup for in-context learning medical image segmentation. IEEE Transactions on Medical Imaging **44**(1), 310–319 (2024)
11. Gao, Y., Liu, D., Li, Z., Li, Y., Chen, D., Zhou, M., Metaxas, D.N.: Show and segment: Universal medical image segmentation via in-context learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20830–20840 (2025)
12. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16000–16009 (2022)
13. Hu, S., Liao, Z., Zhen, L., Fu, H., Xia, Y.: Cycle context verification for in-context medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 141–151. Springer (2025)
14. Jieyun, B., ZhanHong, O.: Pubic symphysis-fetal head segmentation and angle of progression (Apr 2023). https://doi.org/10.5281/zenodo.7851339, https://doi.org/10.5281/zenodo.7851339
15. Jin, H., Liu, S., Cong, C., Feng, Q., Liu, Y., Huang, L., Hu, Y.: Fedwsidd: Federated whole slide image classification via dataset distillation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 178–188. Springer (2025)
16. Lei, S., Tao, D.: A comprehensive survey of dataset distillation. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(1), 17–32 (2023)
17. 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. IEEE (2020)
18. 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)
19. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Dataset distillation for medical dataset sharing. arXiv preprint arXiv:2209.14603 (2022)
20. Orlando, J.I., Fu, H., Breda, J.B., Van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Analysis **59**, 101570 (2020)
21. Qazi, M.A., Hashmi, A.U.R., Sanjeev, S., Almakky, I., Saeed, N., Gonzalez, C., Yaqub, M.: Continual learning in medical imaging: a survey and practical analysis. ACM Computing Surveys (2024)
22. Rakic, M., Wong, H.E., Ortiz, J.J.G., Cimini, B.A., Guttag, J.V., Dalca, A.V.: Tyche: Stochastic in-context learning for medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11159–11173 (2024)
23. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: icarl: Incremental classifier and representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2001–2010 (2017)
24. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon

histology images: The glas challenge contest. Medical Image Analysis **35**, 489–502 (2017)
25. Wang, X., Wang, W., Cao, Y., Shen, C., Huang, T.: Images speak in images: A generalist painter for in-context visual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6830–6839 (2023)
26. Wang, X., Zhang, X., Cao, Y., Wang, W., Shen, C., Huang, T.: Seggpt: Towards segmenting everything in context. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1130–1140 (2023)
27. Xie, J., Tonioni, A., Rauschmayr, N., Tombari, F., Schiele, B.: Test-time visual in-context tuning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19996–20005 (2025)
28. Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Dong, H., Qiao, Y., Gao, P., Li, H.: Personalize segment anything model with one shot. In: The Twelfth International Conference on Learning Representations (2024)
29. Zhang, Y., Zhou, K., Liu, Z.: What makes good examples for visual in-context learning? Advances in Neural Information Processing Systems **36**, 17773–17794 (2023)
30. Zhu, Y., Ma, H., Zhang, C.: Exploring task-level optimal prompts for visual in-context learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 11031–11039 (2025)