Dynamic Gradient Sparsification Training for Few-Shot Fine-tuning of CT Lymph Node Segmentation Foundation Model

Zihao Luo[1], Zijun Gao[2], Wenjun Liao[3], Shichuan Zhang[3], Guotai Wang[1,4(✉)], and Xiangde Luo[5(✉)]

[1]School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China.
[3]Department of Radiation Oncology, Sichuan Cancer Hospital & Institute, University of Electronic Science and Technology of China, Chengdu, China
[4]Shanghai Artificial Intelligence Laboratory, Shanghai, China.
[5]Department of Radiation Oncology, Stanford University School of Medicine, Palo Alto, California
`guotai.wang@uestc.edu.cn`, `luoxd96@stanford.edu`

Abstract. Accurate lymph node (LN) segmentation is critical in radiotherapy treatment and prognosis analysis, but is limited by the need for large annotated datasets. While deep learning-based segmentation foundation models show potential in developing high-performing models with fewer samples, their medical adaptation faces domain-specific prior deficiencies in the LN domain and inefficient few-shot fine-tuning for complex clinical practices, highlighting the necessity of an LN segmentation foundation model. In this work, we annotated 36,106 visible LNs from 3,346 publicly available head-and-neck CT scans to establish a robust LN segmentation model (nnUNetv2). Building on this, we propose **D**ynamic **G**radient **S**parsification **T**raining (DGST), a few-shot fine-tuning approach that preserves foundational knowledge while dynamically updating the most critical parameters of the LN segmentation model with few annotations. We validate it on two publicly available LN segmentation datasets: SegRap2023 and LNQ2023. The results show that DGST outperforms existing few-shot fine-tuning methods, achieving satisfactory performance with limited labeled data. We release the dataset, models and all implementations to facilitate relevant research: https://github.com/HiLab-git/LN-Seg-FM.

Keywords: Lymph nodes segmentation · Foundation model · Few-shot fine-tuning

1 Introduction

Accurate lymph node (LN) segmentation in computed tomography (CT) scan is critical for radiotherapy planning, prognosis and follow-up analysis [12], yet

manual delineation of all visible LN remains clinically impractical due to time constraints. While deep learning methods have advanced automated segmentation with various learning paradigms [14,24,21,5], their deployment is hindered by demanding data needs stemming from LNs' anatomical heterogeneity and inter-patient variability. Emerging foundation models address data scarcity challenges [26,19] but face two critical limitations in LN-specific applications: First, generalist [13] or non-LN-specific models [17,1,23] lack domain-specific priors for subtle LN boundary characterization in medical images, leading to suboptimal performance. Second, their excessive computational complexity impedes clinical translation to resource-limited settings. These gaps highlight the unmet need for LN-specialized foundation models in the reality of hospitals that simultaneously achieve: LN variation-robust generalization ensuring exposure to diverse LN variations, computational efficiency to operate within hardware constraints, and stable retraining protocols which is a critical requirement given the constant influx of new patient cohorts with evolving disease patterns in clinical practice. These requirements are interdependent—a prerequisite for practical adoption in radiotherapy workflows.

To address the aforementioned requirements, we conducted an exploratory study focusing on developing and deploying the LN segmentation foundation model. In the first phase, we annotated over 36,106 visible LNs in 3,346 CT scans from a publicly available head and neck (HN) cancer cohort [22], capturing the variability in LN morphology and distribution to establish a solid training base. Considering both performance and efficiency, we selected nnUNet, the most widely used backbone, for its strong capabilities [10,11]. For the retraining for model updating or downstream deployment, parameter-efficient fine-tuning (PEFT) is considered due to its advantage of not requiring additional model structures or adjustments to inference settings [16,20]. However, current methods typically fixed tuning on a subset of model parameters or added new structures to maintain model stability [8,25,4], limiting their ability to capture the complex individual heterogeneity in medical scenarios, particularly in terms of cohort and disease variations, hindering knowledge transfer across large gaps. Therefore, a few-shot fine-tuning method that dynamically balances model stability and flexibility is urgently needed.

To bridge these gaps, we introduce a novel approach called **D**ynamic **G**radient **S**parsification **T**raining (DGST). Compared with previous work [27,8], it not only maintains stability through sparse parameter updates but also improves efficiency and performance by implementing a dynamic sparsification process via gradients. Specifically, during the fine-tuning phase of the foundation model, DGST dynamically selects and updates the most critical parameters based on the gradient at each iteration. This dynamic selection process ensures a balance between model stability and plasticity. We investigate the LN segmentation foundation models' transferability and DGST's effectiveness on two representative downstream tasks: one within the same anatomical region (SegRap2023) and one across different regions (LNQ2023) to evaluate DGST's ability to handle complex variations in medical scenarios. The contributions of this work can
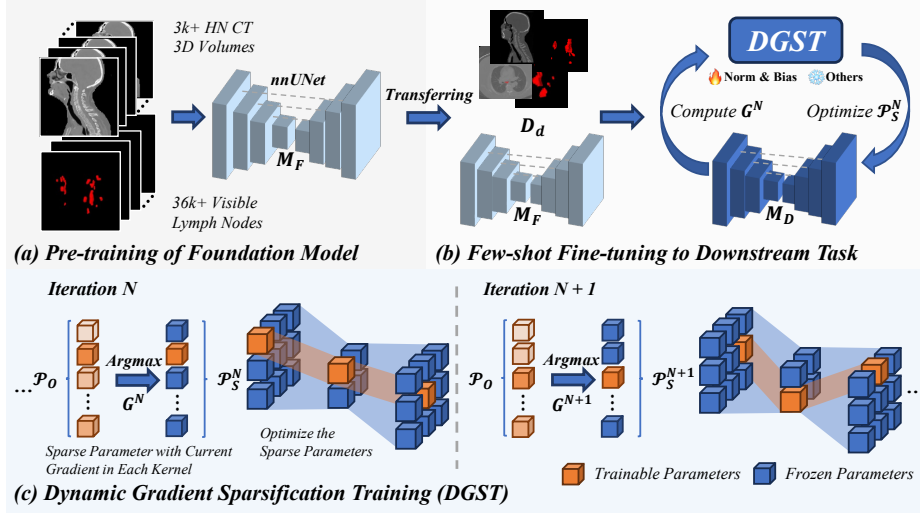
Fig. 1. (a) Pre-training of the foundation model using 3k+ HN CT volumes and 36k+ visible lymph node annotations. (b) Few-shot fine-tuning to downstream tasks by transferring the pre-trained model to new datasets via Dynamic Gradient Sparsification Training (DGST). (c) DGST methodology: At each iteration, the parameters $\mathcal{P}_O$ are sparsified to $\mathcal{P}_S^N$ using the current gradient G^N for each kernel, and then optimized.

be summarized as follows: (1) This pioneering exploratory study on foundation models for LN segmentation involves the annotation and public release of 36,106 visible LNs in 3,346 CT scans from a publicly available HN cancer cohort [22], providing a valuable basis for future research in this area. (2) We propose DGST, a few-shot fine-tuning method tailored to balance the stability and flexibility of foundation models in medical scenarios involving complex disease progression and LN anatomical variability. (3) Comprehensive validation on HN LN segmentation using the SegRap2023 dataset and mediastinal LN segmentation with the LNQ2023 dataset demonstrates the superior performance of DGST over existing methods. Besides, our DGST method achieves comparable performance to sufficient data settings in SegRap2023 with only 10 annotated samples and significantly narrows the performance gap in LNQ2023 with 20 annotated samples.

2   Method

We consider a scenario in which a pre-trained LN segmentation foundation model M_F is adapted to downstream tasks with limited annotated samples $\mathcal{D}_d$. An overview of the framework is shown in Fig. 1, and the subsequent sections formally introduce each of its components.

2.1   Pre-training of Foundation Model

To comprehensively capture the variability of LNs, we utilized the large-scale HN CT dataset RADCURE [22], denoted as $\mathcal{D}_F$. We re-delineated this dataset with all visible LNs to encounter the advanced radiotherapy treatment trends [15], and to employ a fully supervised manner to construct the LN segmentation foundation model. The pre-training of the foundation model is represented as:

$$M_F = \arg \min_{\theta_F \in M_F} \mathcal{L}_{CE+Dice}(\mathcal{D}_F : \theta_F) \tag{1}$$

Where M_F and θ_F represent the foundation model and its parameters, respectively; $\mathcal{L}_{CE+Dice}$ denotes the combined cross-entropy and Dice loss.

2.2   Few-shot Fine-tuning to Downstream Task

Few-shot fine-tuning is considered an efficient approach for knowledge transferring from foundation models, as it does not require modifications to the pretraining process and facilitates easier deployment than other methods [16,20]. In the few-shot scenario, we assume that adapting the LN segmentation foundation model M_F to downstream tasks with a limited number of annotated volumes $\mathcal{D}_d$ can alleviate the target institution's resource constraints. The goal is to fine-tune M_F on a new dataset $\mathcal{D}_d$ with limited labeled data, this can be formulated as:

$$M_d = \arg \min_{\theta_d \in M_d} \mathcal{L}_{CE+Dice}(\mathcal{D}_d : \theta_d) + R_{penalty}(\theta_F, \theta_d) \tag{2}$$

where M_d and θ_d represent the fine-tuned downstream model and its parameters, respectively; $\mathcal{L}_{CE+Dice}$ is the combined loss; $\mathcal{D}_d$ is the few-shot dataset for the downstream task; The regularization term $R_{penalty}$ helps prevent overfitting and maintain model plasticity by constraining model parameters. This can involve freezing most of the pre-trained parameters θ_F and updating a small subset (e.g., only fine-tuning the bias [25]) or adding new trainable parameters (e.g., using LoRA [8] or Adapter [7] in key modules).

2.3   Dynamic Gradient Sparsification Training

Due to limited annotations and specific medical image data, advanced data augmentation is essential, but excessive fine-tuning risks overfitting. Previous work utilized a sparse parameter update strategy as a regularization mechanism, controlling the upper bound of model stability to mitigate overfitting [6,27], but statically sparsing limits the model's plasticity when critical parameters vary across cases. To address this, we introduce **D**ynamic **G**radient **S**parsification **T**raining (DGST), a method designed to achieve a balance between preventing overfitting and maintaining model plasticity during the fine-tuning of UNet-like foundation models under few-shot conditions. In our DGST approach, the gradient is computed at the start of each iteration, specifically at iteration N, The gradient set is formulated as:

$$G^N = \left\{ g_{\theta_i}^N \middle| g_{\theta_i}^N = \nabla_{\theta_i} \mathcal{L}(\mathcal{D}_d), \ \theta_i \in \mathcal{P}_O \right\}, \tag{3}$$

where $g_{\theta_i}^N$ is the gradient of parameter θ_i at iteration N; $\mathcal{P}_O$ denotes the set of original model parameters. $\mathcal{L}(\mathcal{D}_d)$ is the loss of the downstream annotated dataset. Further, we assume that the model comprises K convolutional and transposed convolutional kernels, denoted as $\{C_k\}_{k=1}^K$, and select the Top-γ gradient parameters with the highest absolute values in each kernel, forming the set of key sparse parameters to be updated, denoted as $\mathcal{P}_S^N$, is formulated as:

$$\mathcal{P}_S^N = \bigcup_{k=1}^K \underset{\theta_i \in \mathcal{P}_O \cap C_k}{\arg \overset{(\gamma)}{\max}} \left| g_{\theta_i}^N \right| \tag{4}$$

where $\arg \overset{(\gamma)}{\max}$ refers to get the γ parameters with the largest values. Moreover, bias parameters capture task-specific output shifts, while normalization parameters control feature scaling and stability [4,25], we include both in $\mathcal{P}_S^N$. Thus, at iteration N, the parameter update rule is as follows:

$$\theta_i \leftarrow \begin{cases} \theta_i - \eta g_{\theta_i}^N, & \text{for } \theta_i \in \mathcal{P}_s^N \\ \theta_i, & \text{otherwise} \end{cases} \tag{5}$$

where η is the learning rate at iteration N, $g_{\theta_i}^N$ is calculated in Eq.3. Finally, through iterative execution, the model is enabled to progressively prioritize the most critical parameters. Unlike traditional static sparse parameter constraints [27,25], our method implements dynamic sparsity optimization by adjusting parameters via gradient, enabling adaptive updates at each iteration. This approach significantly enhances gradient descent step efficiency, particularly highlighting the effectiveness of our DGST method in achieving a well-balanced trade-off between model flexibility and stability during few-shot fine-tuning.

3 Experiment and Results

3.1 Experimental Details

Dataset. In this study, we employed the RADCURE dataset [22] to train a foundational model for LN segmentation. An exploratory analysis was performed to delineate 36,106 visible LNs across 3,346 HN CT volumes to assess the foundation model's transferability. Additionally, we also used two publicly available LN segmentation datasets for the few-shot fine-tuning experiment: 120 CT volumes from SegRap2023 [18] for HN LN segmentation and 120 CT volumes from LNQ2023 [3,9] for mediastinal LN segmentation. Due to the scarcity of labeled data in the few-shot scenario, establishing a stable validation set is challenging. Therefore, we used an 8:2 training-test split. For SegRap2023, few-shot fine-tuning experiments were conducted with 3, 5, and 10 shots, while for LNQ2023, we tested 5, 10, and 20 shots.
Implementation Details. All training and inference were performed using the nnUNetv2 framework in PyTorch [10], running on a cluster with 8 NVIDIA V100

Table 1. Quantitative comparison of different few-shot fine-tuning methods on two datasets is presented, with results reported as mean ± standard deviation. The best and second-best results are highlighted in bold and underlined, respectively.

SegRap2023						
Method	*3-shot*		*5-shot*		*10-shot*	
	DSC(%)↑	NSD(%)↑	DSC(%)↑	NSD(%)↑	DSC(%)↑	NSD(%)↑
From scratch	46.68±14.46	39.10±13.86	56.52±9.48	48.15±7.74	58.56±10.44	50.98±8.30
Full	62.66±8.45	55.16±9.86	64.86±9.36	57.68±9.41	65.99±9.12	59.32±8.21
LinearProb [2]	59.31±11.13	50.48±10.70	58.95±11.43	50.05±10.59	59.15±11.37	50.31±10.53
Bias [25]	63.48±9.07	55.04±9.27	66.42±8.68	58.83±7.88	67.35±8.72	59.96±7.90
Adapter [7]	62.97±8.92	54.41±9.74	66.77±8.51	59.35±7.69	67.04±8.80	59.69±8.35
Lora [8]	**65.73±8.30**	**58.08±9.24**	66.34±8.59	58.66±8.06	67.02±8.87	59.72±7.87
Affine-IN [4]	64.89±8.44	56.23±8.47	66.33±8.82	58.75±7.77	67.82±8.47	60.64±7.50
DGST(Ours)	65.05±8.07	57.09±7.18	**67.36±8.45**	**60.25±7.62**	**68.44±8.59**	**61.97±7.21**
All-shot	69.20±9.06, 62.33±7.47 (96-shot, from scratch)					

LNQ2023						
Method	*5-shot*		*10-shot*		*20-shot*	
	DSC(%)↑	NSD(%)↑	DSC(%)↑	NSD(%)↑	DSC(%)↑	NSD(%)↑
From scratch	37.11±22.90	34.03±19.79	41.43±24.18	41.87±21.12	55.76±25.57	55.15±23.60
Full	49.09±21.88	48.72±19.06	53.68±24.25	54.09±22.06	62.47±23.44	63.00±21.94
LinearProb [2]	6.56±6.95	6.13±5.07	7.35±9.08	7.05±6.84	6.88±7.71	6.41±5.67
Bias [25]	44.51±22.24	40.99±18.94	51.62±23.68	48.56±20.82	53.43±21.98	50.09±20.15
Adapter [7]	46.99±23.51	43.43±20.71	51.79±23.78	48.46±22.30	57.13±22.48	54.48±20.58
Lora [8]	47.93±20.72	46.21±19.06	52.13±22.63	49.78±20.60	59.25±22.07	56.03±21.62
Affine-IN [4]	45.61±21.83	41.75±19.20	51.10±23.29	47.96±21.35	57.89±23.08	55.16±21.17
DGST(Ours)	**50.94±22.41**	**48.76±19.86**	**54.94±24.36**	**55.24±22.11**	**63.82±21.69**	**63.84±20.31**
All-shot	67.12±18.65, 67.30±18.56 (96-shot, from scratch)					

GPUs. The foundation model was trained and fine-tuned with the default full-resolution 3D U-Net backbone in nnUNet, using a batch size of 2, a patch size of $80\times112\times224$, SGD optimization, an initial learning rate of 0.01 and polynomial decay with power of 0.9 for 2000 epochs. The fine-tuning stage employed the same setup with an initial learning rate of 0.001 for 50 epochs, the hyperparameter γ is set to 1. Data augmentation was performed following the default settings of nnUNet. Segmentation performance was assessed in 3D volumes using the Dice similarity coefficient (DSC) for voxel overlap and Normalized Surface Dice (NSD) with a 1 mm tolerance for boundary accuracy.

Baselines. We first employed two reference methods: **From Scratch**, which trains the model from random initialization, and **Full**, which fine-tunes the entire foundation model. We also employed several PEFT methods: **LinearProb** [2], which updates only the classifier head; **Bias** [25], which tunes only the bias parameters; and **Affine-IN** [4], which fine-tunes the affine parameters of the instance normalization layers. Additionally, we applied **LoRA** [8] and **Adapter** [7] methods, both of which target only the 3D convolution layers, ensuring efficient fine-tuning through low-rank updates or small auxiliary modules. For comparison, we also used all 96 samples with full-parameter training from scratch, referred to **All-shot**. All methods were evaluated using the final checkpoint.
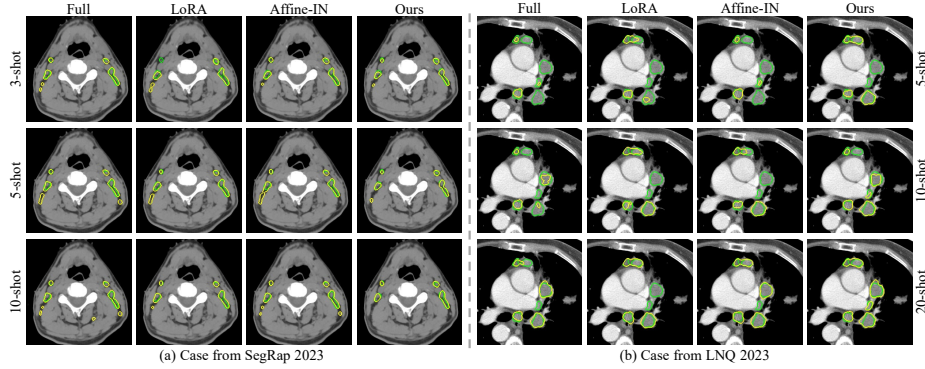
Fig. 2. Qualitative comparison of different fine-tuning methods. The ground truth and predictions are shown in green and yellow contours, respectively.

3.2 Results

Quantitative and Qualitative Results. The quantitative performance of different few-shot fine-tuning methods on the SegRap2023 and LNQ2023 datasets is presented in Table 1. For SegRap2023, the "Full" approach achieves better results than "From scratch" with DSC (15.98%, 8.34%, 7.43%) and NSD (16.06%, 9.53%, 8.34%) at 3, 5, and 10-shot, respectively. In LNQ2023, it outperforms DSC (11.98%, 12.25%, 6.71%) and NSD (14.69%, 12.22%, 7.85%) in 5, 10 and 20-shot, respectively. These results demonstrate our LN segmentation foundation model's strong transferability in both HN and mediastinal areas. In the SegRap2023 dataset, which has a lower domain discrepancy, all few-shot fine-tuning methods, except for "LinearProb", outperform the "Full" method, highlighting the advantage of the sparse parameter strategy in mitigating overfitting. Our DGST method achieved the second-best result in the 3-shot setting and the best performance in the 5-shot and 10-shot settings. These results suggest that, with slightly more training samples, our method effectively balances anti-overfitting and plasticity, demonstrating superior flexibility. In contrast, for the higher-domain gap dataset LNQ2023, other methods fail to outperform the "Full" approach, whereas our method consistently delivers the best results across all settings. Fig.2 shows qualitative results between different methods in both two downstream tasks. While all methods scale with training data, only our DGST consistently excels in reducing both false positives and missed detections for LN segmentation. Both quantitative and qualitative results demonstrate DGST's capability to maintain model plasticity, enabling continuous learning of intricate LN patterns across diverse clinical scenarios.

Sensitivity Analysis of γ. We examined the influence of the hyperparameter γ in the DGST method, which controls the selection of top gradient parameters for sparsification. Experiments were conducted on the SegRap2023 (10-shot) and LNQ2023 (20-shot) datasets with γ values of 1, 2, 3, 5, and 10. Additionally, we present the results of the "Full" method for comparison. As shown in Fig.3, the

Table 2. Ablation study on different parameter sparsification strategies for the few-shot fine-tuning of our LN segmentation foundation model. Full: Full parameters fine-tuning; Bias+Norm: Tuning on bias and normalization parameters; DRST: Dynamic Random Sparsification Training; SGST: Static Gradient Sparsification Training.

| Method	SegRap2023				LNQ2023				Iteration Duration
	3-shot		10-shot		5-shot		20-shot		
	DSC(%)	NSD(%)	DSC(%)	NSD(%)	DSC(%)	NSD(%)	DSC(%)	NSD(%)	
Full	62.66±8.45	55.16±9.86	65.99±9.12	59.32±8.21	49.09±21.88	48.72±19.06	62.47±23.44	63.00±21.94	0.2085s
Encoder Only	59.33±11.13	50.48±10.69	59.23±11.34	50.37±10.53	6.49±6.86	6.03±5.02	6.92±7.72	6.43±5.66	0.0880s
Decoder Only	59.08±10.81	51.12±11.28	64.24±9.96	57.37±8.45	46.97±24.15	45.21±21.52	59.68±22.24	58.88±20.17	0.1605s
Bias+Norm	64.21±9.08	55.79±9.17	67.26±8.24	60.21±7.63	45.91±21.27	41.82±19.34	56.12±21.32	54.19±20.13	0.1534s
DRST	**65.61±8.53**	**57.64±8.30**	67.59±8.93	60.20±7.83	43.56±21.12	40.44±18.05	52.19±22.13	48.93±20.91	0.2193s
SGST	64.58±7.99	56.84±8.40	67.78±8.65	61.01±6.89	48.25±21.40	46.44±19.01	60.28±25.18	60.02±23.70	0.2197s
DGST	65.05±8.07	57.09±7.18	**68.44±8.59**	**61.97±7.21**	**50.94±22.41**	**48.76±19.86**	**63.82±21.69**	**63.84±20.31**	0.2813s
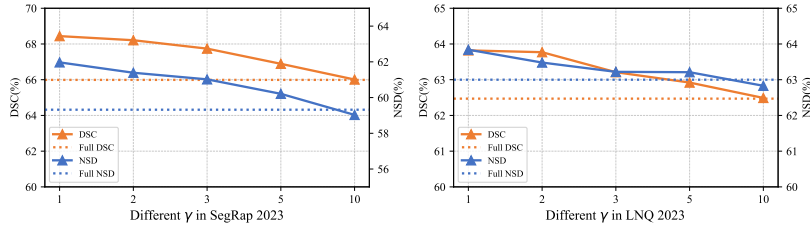
Fig. 3. Sensitivity analysis of hyperparameter γ

results indicate that smaller γ values produce similar performance, while larger values lead to a performance decline, falling close to the "Full".

Ablation study. We performed an ablation study to evaluate different parameter sparsification strategies, with the baseline "Full" method as a reference. We first fine-tuned only the encoder ("Encoder Only") or decoder ("Decoder Only"), finding that sparsification on general structures underperformed. We then explored tuning bias and normalization parameters ("Bias+Norm") and tested a modified approach with random parameter selection ("Dynamic Random Sparsification Training, DRST"). While sparsification improved SegRap2023 performance compared to the "Full" method, it failed to enhance results on LNQ2023, suggesting insufficient consideration of model plasticity. We further implemented a static parameter selection approach ("Static Gradient Sparsification Training, SGST") via initial accumulated gradients. Comparisons of performance and time cost for each strategy are shown in Table 2, demonstrating that gradient-based parameter selection is crucial for model plasticity. However, effective sparsification necessitates dynamic updates based on current gradients to balance stability and flexibility, albeit at the cost of micro-longer times.

4   Conclusion

In conclusion, this study introduces DGST, a novel approach designed to enhance the fine-tuning of foundation models for LN segmentation with few-shot

annotations. By dynamically selecting and updating the most critical parameters based on the gradient at each iteration, DGST effectively balances model stability and flexibility, mitigating the risks of overfitting while preserving the model's ability to adapt to new medical scenarios. The results across SegRap2023 and LNQ2023 demonstrate the superior performance of DGST compared to existing fine-tuning methods. Besides, we will release the dataset of 36,106 annotated LNs and the validated framework to advance the deployment of robust, resource-efficient segmentation tools in evolving clinical workflows.

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

References

1. Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., et al.: Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. MedIA **98**, 103310 (2024)
2. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: ICLR (2019)
3. Dorent, R., Khajavi, R., Idris, T., Ziegler, E., Somarouthu, B., Jacene, H., LaCasce, A., Deissler, J., Ehrhardt, J., Engelson, S., et al.: Lnq 2023 challenge: Benchmark of weakly-supervised techniques for mediastinal lymph node quantification. Machine Learning for Biomedical Imaging **3**, 1–15 (2025)
4. Frankle, J., Schwab, D.J., Morcos, A.S.: Training batchnorm and only batchnorm: On the expressive power of random features in cnns. In: ICLR (2021)
5. Fu, J., Wang, G., Lu, T., Yue, Q., Vercauteren, T., Ourselin, S., Zhang, S.: Um-cam: Uncertainty-weighted multi-resolution class activation maps for weakly-supervised segmentation. PR **160**, 111204 (2025)
6. Fu, Z., Yang, H., So, A.M.C., Lam, W., Bing, L., Collier, N.: On the effectiveness of parameter-efficient fine-tuning. In: AAAI. vol. 37, pp. 12799–12807 (2023)
7. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: ICML. pp. 2790–2799. PMLR (2019)
8. Hu, E.J., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. In: ICLR (2022)
9. Idris, T., Somarouthu, S., Jacene, H., LaCasce, A., Ziegler, E., Pieper, S., Khajavi, R., Dorent, R., Pujol, S., Kikinis, R., Harris, G.: Mediastinal lymph node quantification (lnq): Segmentation of heterogeneous ct data (2024), https://doi.org/10.7937/QVAZ-JA09

10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: MICCAI. pp. 488–498. Springer (2024)
12. Ji, H., Hu, C., Yang, X., Liu, Y., Ji, G., Ge, S., Wang, X., Wang, M.: Lymph node metastasis in cancer progression: molecular mechanisms, clinical significance and therapeutic interventions. STTT **8**(1), 367 (2023)
13. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: ICCV. pp. 4015–4026 (2023)
14. Li, Y., Dan, T., Li, H., Chen, J., Peng, H., Liu, L., Cai, H.: Npcnet: jointly segment primary nasopharyngeal carcinoma tumors and metastatic lymph nodes in mr images. TMI **41**(7), 1639–1650 (2022)
15. Liao, W., Zhao, Y., Zhang, J., Peng, G., Deng, Y., Zhang, X., Ren, J., Zhang, S.: Visible lymph node irradiation (vnrt) of the low-risk neck levels in patients with nasopharyngeal carcinoma: An exploratory trial protocol (2023), pREPRINT available at Research Square
16. Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., Raffel, C.A.: Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. NeurIPS **35**, 1950–1965 (2022)
17. Liu, J., Zhang, Y., Chen, J.N., Xiao, J., Lu, Y., A Landman, B., Yuan, Y., Yuille, A., Tang, Y., Zhou, Z.: Clip-driven universal model for organ segmentation and tumor detection. In: ICCV. pp. 21152–21164 (2023)
18. Luo, X., Fu, J., Zhong, Y., Liu, S., Han, B., Astaraki, M., Bendazzoli, S., Toma-Dasu, I., Ye, Y., Chen, Z., et al.: Segrap2023: A benchmark of organs-at-risk and gross tumor volume segmentation for radiotherapy planning of nasopharyngeal carcinoma. MedIA **101**, 103447 (2025)
19. Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., Rajpurkar, P.: Foundation models for generalist medical artificial intelligence. Nature **616**(7956), 259–265 (2023)
20. Mosbach, M., Pimentel, T., Ravfogel, S., Klakow, D., Elazar, Y.: Few-shot fine-tuning vs. in-context learning: A fair comparison and evaluation. In: Findings of ACL. pp. 12284–12314 (2023)
21. Wang, H., Chen, J., Zhang, S., He, Y., Xu, J., Wu, M., He, J., Liao, W., Luo, X.: Dual-reference source-free active domain adaptation for nasopharyngeal carcinoma tumor segmentation across multiple hospitals. TMI **43**(12), 4078–4090 (2024)
22. Welch, M.L., Kim, S., Hope, A.J., Huang, S.H., Lu, Z., Marsilla, J., Kazmierski, M., Rey-McIntyre, K., Patel, T., O'Sullivan, B., et al.: Radcure: An open-source head and neck cancer ct dataset for clinical radiation therapy insights. Medical Physics **51**(4), 3101–3109 (2024)
23. Ye, Y., Xie, Y., Zhang, J., Chen, Z., Xia, Y.: Uniseg: A prompt-driven universal segmentation model as well as a strong representation learner. In: MICCAI. pp. 508–518. Springer (2023)
24. Yu, Q., Wang, Y., Yan, K., Li, H., Guo, D., Zhang, L., Shen, N., Wang, Q., Ding, X., Lu, L., et al.: Effective lymph nodes detection in ct scans using location debiased query selection and contrastive query representation in transformer. In: ECCV. pp. 180–198. Springer (2024)
25. Zaken, E.B., Goldberg, Y., Ravfogel, S.: Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. In: ACL. pp. 1–9 (2022)

26. Zhang, S., Metaxas, D.: On the challenges and perspectives of foundation models for medical image analysis. MedIA **91**, 102996 (2024)
27. Zhang, Z., Zhang, Q., Gao, Z., Zhang, R., Shutova, E., Zhou, S., Zhang, S.: Gradient-based parameter selection for efficient fine-tuning. In: CVPR. pp. 28566–28577 (2024)