Protocol-Manifold Coverage Learning (PMCL): Source-Free Active Adaptation for Multi-Center Nasopharyngeal Carcinoma MRI Segmentation

Jiawen Lu[1,2], Yahan Zheng[2], Qiaojuan Guo[1,2*], and Xiangjun Wu[1*]

[1] Fuzhou University, Fuzhou, China
[2] Fujian Cancer Hospital, Fuzhou, China
guoqiaojuan@163.com, xiangjunwu@fzu.edu.cn
*Corresponding authors.

Abstract. The cross-center deployment of primary gross tumor volume (GTVp) segmentation models for nasopharyngeal carcinoma (NPC) on CE-T1 MRI is frequently compromised by *protocol mixture*. This induces systemic domain shifts, primarily manifesting as boundary drift—severe inflation of the 95% Hausdorff Distance (HD95) despite stable Dice scores. To address this, we introduce **Protocol-Manifold Coverage Learning (PMCL)**, a source-free active adaptation framework. Operating within a minimal slice budget, PMCL selects target queries independently of source data or pseudo-labels. It quantifies protocol variation via a compact signature, isolates boundary-band Protocol Disagreement (PD) using protocol-aware perturbations, and enforces Group Prior Coverage to prevent sampling collapse. Evaluated on 1,289 volumes across six centers under a 10% annotation budget, PMCL outperforms state-of-the-art methods, improving average Dice by ∼0.10 and reducing mean HD95 by ∼3.9 mm.

Keywords: Nasopharyngeal carcinoma · MRI · Segmentation · Active learning · Source-free active adaptation · Protocol mixture

1 Introduction

Precise delineation of the primary gross tumor volume (GTVp) for nasopharyngeal carcinoma (NPC) on contrast-enhanced T1-weighted (CE-T1) MRI is fundamental to radiotherapy planning. Accurate GTVp boundaries underpin clinical target volume construction and dose prescription, making boundary reliability clinically relevant beyond volumetric overlap [6]. Automated segmentation can support this workflow, but cross-institutional generalization remains challenging in NPC cohorts [19]. MRI data vary across scanner vendors, acquisition settings, sampling patterns, and reconstruction scenarios, creating mismatches between training and target data [16]. In our cohort, such *protocol mixture* is associated with *protocol-induced boundary drift*, where geometric errors increase despite comparatively stable Dice scores. Because Dice and HD95 capture complementary overlap- and distance-based properties, this motivates boundary-focused evaluation and adaptation [9].

Table 1. Multi-center CE-T1 cohort and compact protocol signatures (median [5–95%], natural log). A is in-plane pixel area (mm^2), and $R = \mathrm{FFT}_H/\mathrm{FFT}_L$.

Center	N	$\log(A)$	$\log(R)$
Center1	577	$-1.207\,[-1.689, -0.940]$	$-3.356\,[-4.257, -2.359]$
Center2	275	$-2.858\,[-2.858, -1.434]$	$-5.448\,[-6.160, -4.157]$
CenterA	50	$-1.355\,[-1.562, -0.842]$	$-4.656\,[-5.293, -3.062]$
CenterB	50	$-1.355\,[-1.355, -1.355]$	$-3.644\,[-4.215, -3.039]$
CenterC	60	$-1.515\,[-1.689, -1.352]$	$-3.925\,[-4.620, -3.217]$
CenterFS	277	$-1.601\,[-1.695, -1.470]$	$-4.395\,[-5.081, -3.839]$

Center1 and Center2 are private; CenterA/B/C are from M-NPC [14]; CenterFS is public [4].

Unsupervised Domain Adaptation often uses adversarial output-space or entropy alignment [10, 12], while Source-Free Domain Adaptation adapts a pre-trained source model without source data, commonly via target-side information maximization or pseudo-labeling [5]. Source-Free Active Domain Adaptation (SFADA) further queries a small target subset for expert annotation. Existing SFADA methods emphasize representativeness, reference matching, and calibration [14, 15], but do not explicitly allocate annotations to under-covered acquisition modes or quantify boundary instability under protocol-relevant perturbations. Biased or redundant sample selection remains a limitation [18], and uncertainty-based querying does not guarantee diversity or distributional coverage [13]. Under protocol mixture, this may over-sample dense protocol regions while overlooking sparse but important acquisition modes.

We study 1,289 CE-T1 MRI volumes from six centers: private Center1 and Center2 cohorts, CenterA/B/C from M-NPC [14], and the public CenterFS dataset [4]. Table 1 summarizes cohort composition and two compact protocol-signature variables: in-plane sampling geometry, $\log(dx\,dy)$, and image-frequency sharpness, $\log(\mathrm{FFT}_H/\mathrm{FFT}_L)$. Figure 1 shows partially overlapping but distinct cross-center distributions, motivating explicit coverage of the target protocol space.

We propose **Protocol-Manifold Coverage Learning (PMCL)**, which formulates pre-deployment adaptation as coverage-constrained querying in this protocol space. PMCL combines: (i) a compact volume-level protocol signature; (ii) boundary-band **Protocol Disagreement (PD)**, which measures prediction instability under protocol-aware perturbations; and (iii) **Group Prior Coverage**, which distributes a fixed annotation budget across protocol modes. Query selection requires no source images, stored source features, source-derived reference vectors, target latent-space clustering, or reference-star selection.

Across ten source–target adaptation tasks with a fixed 10% target-slice annotation budget, PMCL improves adaptation robustness across heterogeneous centers, with notable gains in geometric boundary accuracy.

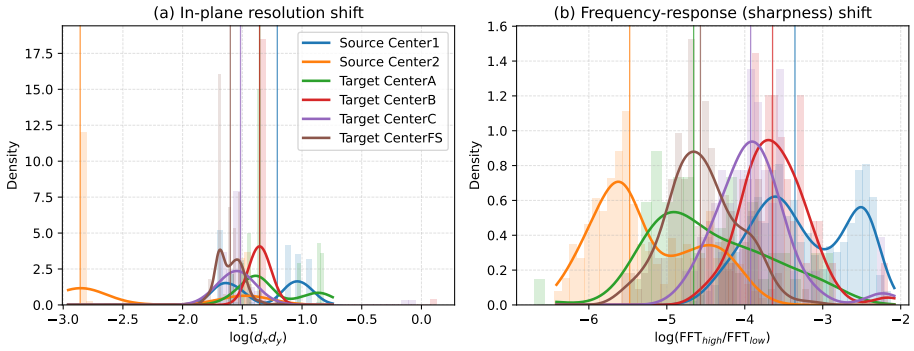


Fig. 1. Cross-center protocol mixture characterized by in-plane sampling geometry, $\log(dx\,dy)$, and image-frequency sharpness, $\log(\mathrm{FFT}_H/\mathrm{FFT}_L)$. Histograms and kernel-density estimates show partially overlapping but distinct distributions across the six centers.

2 Methods

2.1 Problem Formulation and Deployment Context

We study source-free active adaptation for NPC GTV segmentation under severe protocol mixture. For each transfer task, an initial model f_{θ_0} is trained on a fully annotated source dataset $\mathcal{D}_s$. At deployment, source images and source feature representations are unavailable because of privacy and data-sharing constraints; only an unlabeled target-slice pool $\mathcal{U}_t$ and annotation budget B are accessible [13]. The goal is to select $S \subset \mathcal{U}_t$, $|S| = B$, for expert annotation and then fine-tune f_{θ_0} on S to obtain f_θ. The PMCL pipeline is shown in Fig. 2.

2.2 Formulation of the Protocol Signature

To quantify acquisition differences, PMCL extracts a compact volume-level protocol signature from geometric and frequency-domain characteristics:

$$\phi(v) = \big[\log(A_v + \epsilon),\ \ \log(R_v + \epsilon)\big], \quad A_v = dx_v\, dy_v, \quad R_v = \frac{\mathrm{FFT}_H(v) + \epsilon}{\mathrm{FFT}_L(v) + \epsilon}, \tag{1}$$

where dx_v and dy_v are in-plane resolutions, A_v is the in-plane pixel area, and R_v is the ratio of high- to low-frequency FFT energy [17]. We use $\epsilon = 1 \times 10^{-8}$ for numerical stability. Because segmentation is performed slice by slice, z-spacing is excluded for compactness. The signature is assigned to all slices of volume v and standardized over target volumes as $\phi_z(v) = \frac{\phi(v) - \mu_t}{\sigma_t}$.

2.3 Boundary-Band Protocol Disagreement (PD)

To probe sensitivity to protocol-induced shifts, PMCL applies a fixed perturbation bank $\{T_k\}_{k=1}^{K}$ that simulates acquisition and reconstruction variability [11,

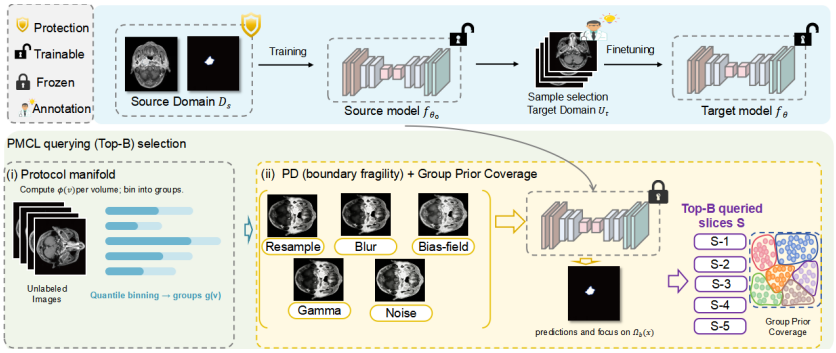


Fig. 2. Overview of the PMCL framework. **(Top)** Source model f_{θ_0} pre-training on $\mathcal{D}_s$ and source-free adaptation on target pool $\mathcal{U}_t$. **(Bottom)** Target queries are selected by: (i) volume grouping along a protocol manifold for prior coverage, and (ii) computing boundary-band Protocol Disagreement (PD) via the predictive variance of frozen f_{θ_0} under perturbations. The queried subset S is annotated to fine-tune f_θ.

3]. For slice x:

$$x^{(k)} = T_k(x), \quad p^{(k)} = f_{\theta_0}(x^{(k)}), \quad p^{(k)} \in [0,1]^{H\times W}. \tag{2}$$

PD is the predictive variance across views, aggregated only inside a dynamically predicted boundary band to avoid dilution by confident background:

$$\mathrm{PD}(x) = \frac{1}{|\Omega_b(x)|} \sum_{u\in\Omega_b(x)} \mathrm{Var}\big(\{p^{(k)}(u)\}_{k=1}^{K}\big), \tag{3}$$

The band is constructed from the mean prediction $\bar{p} = \frac{1}{K}\sum_k p^{(k)}$ using morphological operations [8]:

$$\Omega_b(x) = \mathrm{dilate}(\hat{y}; w) \setminus \mathrm{erode}(\hat{y}; w), \quad \hat{y} = \mathbb{1}[\bar{p} > \tau], \tag{4}$$

where $w = 3$ is the morphological kernel size and $\tau = 0.5$ is the binarization threshold. If catastrophic domain shift produces no foreground prediction ($|\Omega_b(x)| = 0$), global entropy is used as a fallback so zero-foreground slices remain eligible for manual inspection.

2.4 Enforcing Group Prior Coverage

Under protocol mixture, active querying may collapse onto redundant slices from a dominant protocol mode. PMCL therefore discretizes the two-dimensional standardized protocol space into non-empty quantile groups, assigning each volume v to $g(v) \in \{1,\dots,G\}$. If n_g is the number of target slices from group g, the global budget B is allocated proportionally with a minimum quota:

$$B_g \propto n_g, \quad B_g \geq B_{\min}, \quad \sum_{g=1}^{G} B_g = B. \tag{5}$$

Within each group, slices are ranked by a score that balances normalized Protocol Disagreement with spatial dispersion in the protocol manifold:

$$\text{Score}(x) = \lambda U(x) + (1-\lambda)\left\|\phi_z(v_x) - \overline{\phi_z}_g\right\|_2, \tag{6}$$

where $U(x)$ is min-max normalized PD, v_x is the originating volume, and $\overline{\phi_z}_g$ is the group centroid. We set $\lambda = 0.5$ and cap selection at $m = 5$ slices per volume to limit intra-volume redundancy.

2.5 Fine-Tuning Dynamics and Computational Complexity

After annotating S, the source model is fine-tuned with the same architecture and supervised loss used during source training. PMCL adds no learnable parameters or auxiliary networks; its overhead is limited to K one-off forward passes per slice during query scoring, and adapted-model inference cost is unchanged.

3 Experiments and Results

3.1 Experimental Configuration

We evaluate ten transfer tasks using Center1 and Center2 as source centers and the remaining centers as targets, with a fixed 7:1:2 volume-level split per center. A 2D U-Net [7] processes 256×256 z-score-normalized slices; Dice and HD95 are reported [9]. Query methods label 10% of target training slices and fine-tune for 1,500 iterations. Source and target-only models are trained for 12,000 iterations using SGD (batch 32, lr 0.03 with exponential decay 0.9). Baselines include Target-only, Source-only, ADVENT, AdaptSeg [12, 10], DPL, CPR [1, 2], Random, Entropy [13], and STDR [14]; all query methods use the same budget and fine-tuning protocol.

For PD, we use a fixed protocol-aware perturbation bank with $K = 5$ views that emulates realistic acquisition/reconstruction variations, combining down/up resampling (scale in $\{0.75, 0.85, 1.0, 1.15, 1.25\}$), Gaussian blur ($\sigma \in \{0, 0.5, 0.8, 1.2\}$), multiplicative bias field (strength in $\{0, 0.10, 0.15, 0.20, 0.30\}$) [11], gamma mapping ($\gamma \in \{0.90, 0.95, 1.0, 1.05, 1.10\}$), and additive Gaussian noise (relative std in $\{0, 0.01, 0.02, 0.03, 0.04\}$); all parameters are fixed for all datasets and tasks [3].

3.2 Cross-Institutional Transfer

Across ten transfer tasks (Table 2), PMCL improves Dice and HD95 over Entropy in every task under the same target pool, 10% budget, checkpoint, and evaluation protocol. Relative to STDR, mean Dice increases from 0.624 to 0.725 and mean HD95 decreases by 3.865 mm; paired Wilcoxon tests give $p < 0.05$. On CenterFS with Center2 as source, HD95 decreases from 24.09 to 9.31 mm despite the minimal annotation budget. The protocol analyses in Figs. 1 and 4 show a clearer degradation trend for HD95 than Dice, consistent with boundary-dominant protocol effects. Representative CenterFS cases are shown in Fig. 3.

Table 2. Cross-center results. Each cell gives Dice (0–1, first line) and HD95 (mm, second line), mean±std. Target-only uses 100% target labels; query methods use 10%. Bold denotes the best practical query method for each metric.

Method	C2	A	B	C	FS
Source: Center1					
	0.79 ± 0.08	0.85 ± 0.04	0.74 ± 0.12	0.78 ± 0.09	0.66 ± 0.14
Target-only	3.68 ± 2.31	5.95 ± 6.87	13.81 ± 19.93	12.04 ± 25.37	8.54 ± 5.71
	0.33 ± 0.21	0.29 ± 0.21	0.16 ± 0.10	0.36 ± 0.18	0.38 ± 0.20
Source-only	15.42 ± 10.99	20.22 ± 13.31	24.98 ± 10.13	26.30 ± 27.45	13.94 ± 6.77
	0.58 ± 0.15	0.37 ± 0.20	0.24 ± 0.11	0.43 ± 0.17	0.45 ± 0.19
ADVENT [12]	6.70 ± 2.64	17.55 ± 11.04	21.08 ± 8.97	24.86 ± 27.63	12.55 ± 5.60
	0.60 ± 0.15	0.38 ± 0.20	0.26 ± 0.12	0.44 ± 0.17	0.46 ± 0.19
AdaptSeg [10]	6.49 ± 2.65	17.05 ± 10.68	19.62 ± 7.28	24.38 ± 27.74	12.63 ± 5.65
	0.60 ± 0.14	0.37 ± 0.19	0.31 ± 0.11	0.49 ± 0.14	0.43 ± 0.20
DPL [1]	6.45 ± 2.62	16.61 ± 9.67	18.21 ± 6.60	23.41 ± 27.84	13.12 ± 6.12
	0.53 ± 0.17	0.46 ± 0.18	0.50 ± 0.10	0.57 ± 0.10	0.46 ± 0.20
CPR [2]	7.76 ± 3.44	19.85 ± 11.73	22.82 ± 10.12	44.89 ± 32.05	13.06 ± 6.05
	0.73 ± 0.12	0.69 ± 0.15	0.55 ± 0.14	0.60 ± 0.12	0.57 ± 0.16
Random [13]	5.04 ± 3.31	15.43 ± 11.28	18.67 ± 5.88	25.90 ± 33.19	12.08 ± 6.83
	0.74 ± 0.10	0.68 ± 0.10	0.38 ± 0.13	0.69 ± 0.11	0.58 ± 0.16
Entropy [13]	5.03 ± 4.15	25.72 ± 19.93	60.28 ± 21.69	15.82 ± 21.99	14.76 ± 21.92
	0.62 ± 0.13	0.71 ± 0.10	0.64 ± 0.10	0.67 ± 0.09	0.59 ± 0.13
STDR [14]	13.26 ± 9.34	$\mathbf{7.77 \pm 4.30}$	14.47 ± 8.95	$\mathbf{14.27 \pm 24.67}$	11.03 ± 5.90
	$\mathbf{0.76 \pm 0.09}$	$\mathbf{0.78 \pm 0.06}$	$\mathbf{0.75 \pm 0.06}$	$\mathbf{0.74 \pm 0.07}$	$\mathbf{0.66 \pm 0.15}$
PMCL (Ours)	$\mathbf{4.75 \pm 3.38}$	8.42 ± 6.33	$\mathbf{7.87 \pm 2.38}$	15.54 ± 24.77	$\mathbf{9.48 \pm 5.84}$
Source: Center2					
Method	C1	A	B	C	FS
	0.76 ± 0.06	0.85 ± 0.04	0.74 ± 0.12	0.78 ± 0.09	0.66 ± 0.14
Target-only	3.55 ± 3.86	5.95 ± 6.87	13.81 ± 19.93	12.04 ± 25.37	8.54 ± 5.71
	0.61 ± 0.12	0.52 ± 0.15	0.40 ± 0.16	0.55 ± 0.12	0.58 ± 0.15
Source-only	6.46 ± 4.52	12.65 ± 6.80	16.34 ± 4.61	23.72 ± 30.50	11.05 ± 5.95
	0.63 ± 0.10	0.52 ± 0.16	0.41 ± 0.15	0.56 ± 0.12	0.59 ± 0.15
ADVENT [12]	6.25 ± 4.47	12.82 ± 6.96	15.46 ± 5.65	23.73 ± 30.73	10.38 ± 5.09
	0.63 ± 0.10	0.52 ± 0.15	0.42 ± 0.15	0.56 ± 0.12	0.58 ± 0.14
AdaptSeg [10]	6.24 ± 4.45	12.46 ± 6.74	15.29 ± 5.61	23.56 ± 30.69	10.40 ± 5.14
	0.64 ± 0.10	0.54 ± 0.14	0.43 ± 0.15	0.56 ± 0.11	0.59 ± 0.15
DPL [1]	5.99 ± 4.38	13.07 ± 7.89	15.34 ± 6.27	29.63 ± 31.76	10.49 ± 5.11
	0.64 ± 0.10	0.56 ± 0.14	0.47 ± 0.13	0.57 ± 0.10	0.58 ± 0.15
CPR [2]	6.02 ± 4.40	12.78 ± 8.05	17.14 ± 9.15	29.91 ± 31.98	10.69 ± 5.15
	0.72 ± 0.08	0.68 ± 0.14	0.57 ± 0.14	0.63 ± 0.10	0.57 ± 0.18
Random [13]	4.29 ± 4.09	23.16 ± 15.50	16.66 ± 6.38	$\mathbf{13.51 \pm 11.48}$	10.59 ± 6.68
	0.73 ± 0.08	0.72 ± 0.11	0.42 ± 0.18	0.64 ± 0.14	0.55 ± 0.19
Entropy [13]	9.61 ± 19.14	13.69 ± 11.73	33.10 ± 35.53	23.24 ± 27.39	28.71 ± 43.39
	0.69 ± 0.09	0.65 ± 0.20	0.46 ± 0.21	0.59 ± 0.06	0.62 ± 0.15
STDR [14]	4.50 ± 4.10	$\mathbf{9.33 \pm 4.72}$	17.20 ± 8.87	15.25 ± 23.63	24.09 ± 36.90
	$\mathbf{0.75 \pm 0.07}$	$\mathbf{0.75 \pm 0.14}$	$\mathbf{0.73 \pm 0.09}$	$\mathbf{0.70 \pm 0.17}$	$\mathbf{0.64 \pm 0.14}$
PMCL (Ours)	$\mathbf{3.86 \pm 3.98}$	9.95 ± 7.95	$\mathbf{8.62 \pm 4.23}$	13.61 ± 27.42	$\mathbf{9.31 \pm 5.38}$

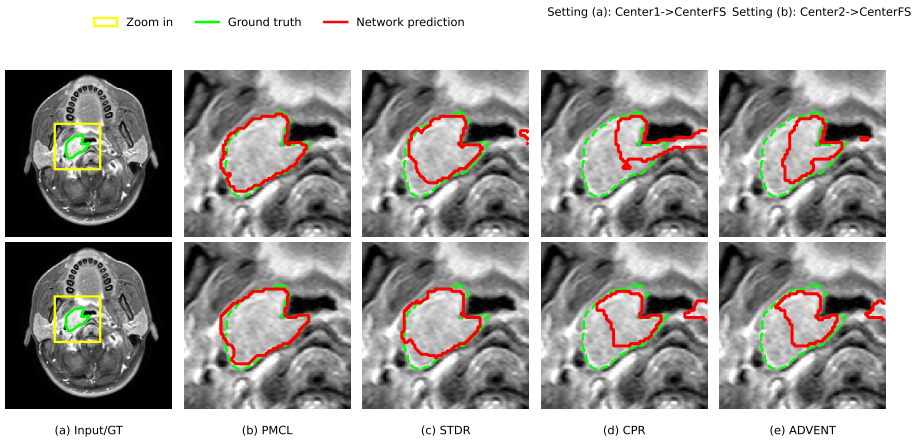


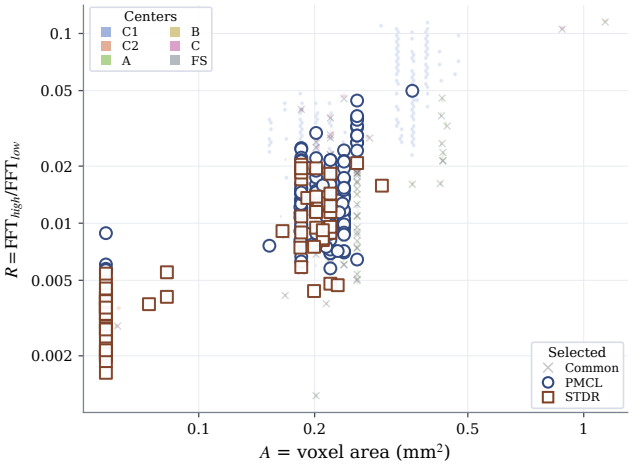
Fig. 3. Representative CenterFS cases for Center1→CenterFS (top) and Center2→CenterFS (bottom). Green and red contours denote ground truth and prediction, respectively.

3.3 Query Distribution and Ablation

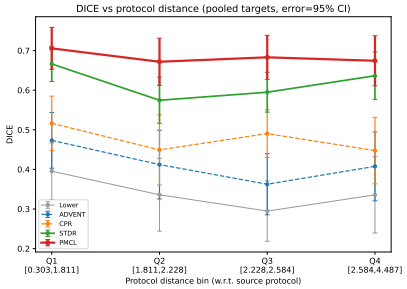
Figure 4(a) shows that PMCL and STDR share some selections in dense protocol regions but allocate method-specific queries differently; PMCL also covers relatively sparse protocol regions. Figures 4(b–c) show clearer protocol-distance variation for HD95 than Dice, suggesting that boundary accuracy is more sensitive to protocol shift than volumetric overlap. In Table 3, removing Group Prior Coverage, protocol perturbations, or boundary-band aggregation generally reduces Dice and/or increases HD95, supporting their complementary roles. Thus, the query plot and ablation should be read together: PMCL couples boundary instability with coverage-constrained allocation rather than relying on PD alone.

Table 3. Ablation results under a fixed 10% target-slice annotation budget (mean±std).

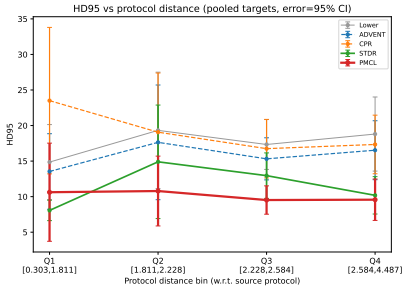
	C1→C2		C1→FS	
Method	Dice↑	HD95↓	Dice↑	HD95↓
STDR [14]	0.62±0.13	13.26±9.34	0.59±0.13	11.03±5.90
PMCL (full)	**0.76±0.09**	**4.75±3.38**	**0.66±0.15**	**9.48±5.84**
w/o Group Prior	0.75±0.11	6.47±10.15	0.61±0.15	9.95±4.36
w/o Protocol Perturb.	0.76±0.09	5.32±4.24	0.60±0.15	10.08±5.71
w/o Boundary Band	0.74±0.11	6.68±7.51	0.65±0.14	9.73±4.49



(a) Query distribution



(b) Dice vs. distance



(c) HD95 vs. distance

Fig. 4. Protocol-space analysis. (a) PMCL and STDR query distributions for pooled Center1-source tasks. (b–c) Dice and HD95 across protocol-distance quartiles; error bars denote 95% confidence intervals.

4 Discussion and Conclusion

Multi-center CE-T1 protocol mixture can produce boundary-dominant errors that are not fully captured by volumetric overlap. PMCL addresses this failure mode by combining protocol-space coverage with perturbation-based boundary instability. Across 1,289 volumes and ten transfer tasks, it improves Dice and HD95 under a fixed 10% slice-labeling budget without source data, source representations, or pseudo-labels during querying. Protocol-distance analyses and comparisons with Entropy further support the value of protocol-aware coverage beyond single-view uncertainty sampling. PMCL is intended for low-annotation, pre-deployment adaptation and does not replace clinician review, correction, or approval. Limitations include evaluation at a single budget, incomplete protocol metadata, and 2D slice-level querying without explicit 3D context. Future work

should evaluate multiple budgets, incorporate richer acquisition metadata, and extend querying to 3D volumes or anatomically meaningful regions.

Data and Code Availability CenterFS is publicly available [4]; CenterA/B/C follow the M-NPC access procedure [14]; Center1 and Center2 remain private under institutional data-governance restrictions. Code is available at `https://github.com/seatrees104/PMCL2026`.

Acknowledgments The authors thank Dr. Hongqiu Wang and the M-NPC team for providing access to the M-NPC dataset.

Disclosure of Interests The authors declare no competing interests.

References

1. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.A.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science, vol. 12905, pp. 225–235. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_22
2. Huai, Z., Ding, X., Li, Y., Li, X.: Context-aware pseudo-label refinement for source-free domain adaptive fundus image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Lecture Notes in Computer Science, vol. 14226, pp. 618–628. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-43990-2_58
3. Isensee, F., Jaeger, P.F., Kohl, S.A.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). https://doi.org/10.1038/s41592-020-01008-z
4. Li, Y., Chen, Q., Li, M., Si, L., Guo, Y., Xiong, Y., Wang, Q., Qin, Y., Xu, L., van der Smagt, P., et al.: A dataset of primary nasopharyngeal carcinoma MRI with multi-modalities segmentation. Scientific Data **12**(1), 1450 (2025). https://doi.org/10.1038/s41597-025-05815-x
5. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 6028–6039. PMLR (2020)
6. Lin, S.J., Guo, Q.J., Liu, Q., Ng, W.T., Ahn, Y.C., AlHussain, H., Chan, A.W., Chow, J., Chua, M.L.K., Corry, J., et al.: International consensus guideline on delineation of the clinical target volumes at different dose levels for nasopharyngeal carcinoma (2024 version). International Journal of Radiation Oncology*Biology*Physics **123**(2), 415–431 (2025). https://doi.org/10.1016/j.ijrobp.2025.05.019
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

8. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer Berlin, Heidelberg, 2 edn. (2004). https://doi.org/10.1007/978-3-662-05088-0
9. Taha, A.A., Hanbury, A.: Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging **15**, 29 (2015). https://doi.org/10.1186/s12880-015-0068-x
10. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7472–7481 (2018). https://doi.org/10.1109/CVPR.2018.00780
11. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging **29**(6), 1310–1320 (2010). https://doi.org/10.1109/TMI.2010.2046908
12. Vu, T.H., Jain, H., Bucher, M., Cord, M., P'erez, P.: ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2517–2526 (2019). https://doi.org/10.1109/CVPR.2019.00262
13. Wang, H., Jin, Q., Li, S., Liu, S., Wang, M., Song, Z.: A comprehensive survey on deep active learning in medical image analysis. Medical Image Analysis **95**, 103201 (2024). https://doi.org/10.1016/j.media.2024.103201
14. 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. IEEE Transactions on Medical Imaging **43**(12), 4078–4090 (2024). https://doi.org/10.1109/TMI.2024.3412923
15. Wang, H., Zhang, S., Chen, J., He, Y., Xu, J., Huang, H., Xiao, J., Li, L., Liao, W., Zhang, S., Luo, X.: Versatile source-free active domain adaptation for multi-center and multi-rater medical image segmentation. Information Fusion **126**, 103586 (2026). https://doi.org/10.1016/j.inffus.2025.103586
16. Wang, Z., Yu, X., Wang, C., Chen, W., Wang, J., Chu, Y.H., Sun, H., Li, R., Li, P., Yang, F., Han, H., Kang, T., Lin, J., Yang, C., Chang, S., Shi, Z., Hua, S., Li, Y., Hu, J., Zhu, L., Zhou, J., Lin, M., Guo, J., Cai, C., Chen, Z., Guo, D., Yang, G., Qu, X.: One for multiple: Physics-informed synthetic data boosts generalizable deep learning for fast MRI reconstruction. Medical Image Analysis **103**, 103616 (2025). https://doi.org/10.1016/j.media.2025.103616
17. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4085–4095 (2020). https://doi.org/10.1109/CVPR42600.2020.00414
18. Yang, Z., Li, Y., Liu, Y.: Continual source-free active domain adaptation for nasopharyngeal carcinoma tumor segmentation across multiple hospitals. Neural Networks **192**, 107869 (2025). https://doi.org/10.1016/j.neunet.2025.107869
19. Yuce, M., Erdogan, S.B., Akin, A., Ozyar, E., Gungor, G., Guvendiren, B., Tanoren, B., Dumlu, S.N.: Multi-institutional deep learning for GTV segmentation and survival prediction in nasopharyngeal carcinoma. BMC Medical Imaging **26**(1), 154 (2026). https://doi.org/10.1186/s12880-026-02195-5