RaD-Seg: Region-Aware Disentangled Learning for Plan-Guided Longitudinal Tumor Segmentation in Head-and-Neck Adaptive Radiotherapy

Hee Guan Khor[1][0000−0002−4935−7815], Fang Chen[2][0000−0003−1926−3807], Yingni Wang[1][0000−0001−9324−3172], Longfei Ma[1][0000−0001−6267−2177], and Hongen Liao[1,2](✉)[0000−0003−3847−9347]

[1] School of Biomedical Engineering, Tsinghua Medicine, Tsinghua University, Beijing, China
liao@tsinghua.edu.cn
[2] School of Biomedical Engineering and the Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

Abstract. Adaptive radiotherapy (ART) for head-and-neck (H&N) cancer requires reliable longitudinal delineation of primary and nodal gross tumor volumes (GTVp/GTVn) to update treatment plans under evolving anatomy. Clinically, mid-treatment (midRT) contouring anchors to stable anatomical landmarks while interpreting deviations in tumor-bearing regions as treatment response; however, many learning-based methods either decouple pre-treatment (preRT) from midRT or impose overly global temporal consistency, thereby under-modeling the heterogeneous, target-dependent evolution of GTVp/GTVn. We propose **RaD-Seg**, a clinically grounded framework that mirrors the plan–verify–adapt workflow via three modules: (i) **Region-Aware Dual-Stream Contrastive Learning** uses region prototypes to disentangle stable anatomical anchors from response-sensitive features, maintaining spatial reliability and sensitivity to regression; (ii) **Temporal-Difference Response Encoding** encodes cumulative radiation effects as an explicit preRT–midRT residual code to modulate midRT representations, rather than conflating response with incidental appearance variation; and (iii) **Plan-Guided Asymmetric Regularization** anchors midRT semantics to the preRT planning definition in anatomically stable regions while permitting clinically plausible target evolution. Patient-level cross-validation on HNTS-MRG 2024 shows significant gains in midRT GTVp/GTVn delineation over strong baselines.

Keywords: Adaptive Radiotherapy · Longitudinal Segmentation · Disentangled Representation · Head and Neck Cancer · Contrastive Learning

1 Introduction

Head-and-neck (H&N) adaptive radiotherapy (ART) follows a clinically structured plan–verify–adapt workflow, in which target volumes delineated on the

pre-treatment (preRT) planning scan must be re-assessed and re-delineated on the mid-treatment (midRT) verification scan to reflect interval anatomical deformation and therapy-induced response. In this setting, accurate longitudinal segmentation of the primary and nodal gross tumor volumes (GTVp/GTVn) is critical, as even small contour deviations may propagate into clinically meaningful target undercoverage or unnecessary irradiation of surrounding organs-at-risk, ultimately compromising dose conformity and adaptive decision-making. Importantly, this problem is inherently heterogeneous and asymmetric across tumor subregions: GTVp often undergoes substantial regression, shape deformation, and boundary migration during treatment, whereas GTVn typically exhibits more subtle changes and is confounded by radiation-induced edema, inflammation, and local tissue distortion. In routine practice, clinicians address this complexity by anchoring contour updates to relatively stable anatomical landmarks while interpreting localized deviations within tumor-bearing regions as treatment-response signals. However, this anatomy-guided yet response-aware reasoning process is only weakly captured by existing learning-based methods, which often emphasize global feature consistency but insufficiently model longitudinal response patterns and meaningful anchor-to-change relationships.

Recent deep models for ART segmentation are promising [1,7,14,9,8,12,11,6,5], yet many either treat preRT and midRT as separate domains or enforce global alignment / consistency that ignores the inherent asymmetry between stable anatomy and regressing targets [4,15,10]. While DIR supports contour propagation, correspondence can fail in regions with regression or appearance shifts, causing biased transfers [2]. Indiscriminate temporal invariance may suppress GTVp regression, whereas weak modeling of subtle GTVn evolution can lead to over-smoothing and missed contours [13]. Naive multi-timepoint fusion may further entangle treatment-induced appearance changes with true anatomical evolution, reducing interpretability [3,15,10]. These issues motivate preserving stable preRT plan semantics while enabling target-specific temporal adaptation.

In this work, we present **RaD-Seg**, a clinically motivated framework for longitudinal H&N target delineation that explicitly aligns with the ART plan–verify–adapt workflow. We first introduce **Region-Aware Dual-Stream Contrastive Learning** (RDC), which leverages region prototypes to disentangle time-invariant anatomical anchors from response-sensitive target representations, preserving stability in anatomically reliable regions while remaining responsive to expected regression. We then propose **Temporal-Difference Response Encoding** (TDRE), modeling cumulative radiation effects as an explicit residual shift between preRT and midRT feature states, rather than as incidental appearance variation. Finally, we propose **Plan-Guided Asymmetric Regularization** (PGAR) to anchor midRT semantics to the preRT planning definition under deformation, while allowing clinically plausible target evolution. We evaluate RaD-Seg on the HNTS-MRG 2024 cohort[3] for midRT GTVp/GTVn segmentation, achieving significant gains over SOTA baselines by modeling region-dependent temporal evolution and plan-guided asymmetry in ART.

[3] `https://hntsmrg24.grand-challenge.org/overview/`

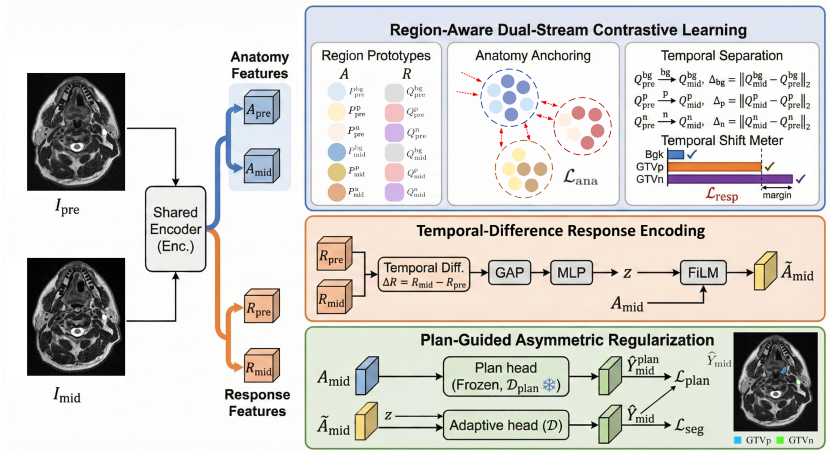

Fig. 1. Overview of RaD-Seg. Given paired preRT–midRT scans, RaD-Seg explicitly disentangles stable anatomical anchors from treatment-response representations, encodes longitudinal temporal residuals into a compact response code, and leverages plan-guided asymmetric regularization to model the distinct evolution patterns of GTVp and GTVn, thereby enabling clinically consistent midRT segmentation.

2 Methodology

2.1 Overview of RaD-Seg

In H&N ART, clinicians delineate planning targets on preRT and re-contour on midRT to account for regression and heterogeneous response; we formalize this as paired samples $\{(I^{(i)}_{\text{pre}}, I^{(i)}_{\text{mid}}, Y^{(i)}_{\text{pre}}, Y^{(i)}_{\text{mid}})\}_{i=1}^{N}$ with $Y_{\text{pre}}, Y_{\text{mid}} \in \{0,1\}^{K\times H\times W\times D}$ for K=3 classes (background, GTVp, GTVn), and predict $\widehat{Y}_{\text{mid}}$ from $(I_{\text{pre}}, I_{\text{mid}})$ while preserving stable anatomy and modeling target-dependent temporal dynamics consistent with the preRT plan. RaD-Seg integrates three components (Fig. 1): RDC (region-level disentanglement of anatomical anchors vs. response-sensitive features), TDRE (compact encoding of preRT–midRT response residuals), and PGAR (plan-guided asymmetric anchoring in reliable regions with flexibility for plausible target evolution). A shared encoder E extracts F_t=$E(I_t)$ for $t \in \{\text{pre}, \text{mid}\}$, projection heads produce (A_t, R_t)=$(g_{\text{ana}}(F_t),\ g_{\text{resp}}(F_t))$, TDRE derives a response code Z from $(R_{\text{mid}}-R_{\text{pre}})$ and modulates midRT anatomy to $\widetilde{A}_{\text{mid}}$, and the decoder outputs $\widehat{Y}_{\text{mid}}$=$D(\widetilde{A}_{\text{mid}}, Z)$; supervision is used only during training, while inference requires only $(I_{\text{pre}}, I_{\text{mid}})$.

2.2 Region-Aware Dual-Stream Contrastive Learning

Clinicians interpret longitudinal change by anchoring edits to stable anatomy while allowing target regions to evolve under treatment. To formulate this region-dependent asymmetry, we aggregate voxel-wise features into class-specific region prototypes for both the anatomical stream A_t and the response stream

R_t. For each class $c \in \mathcal{C} = \{\mathrm{bg}, \mathrm{p}, \mathrm{n}\}$ (background, GTVp, GTVn) and time $t \in \{\mathrm{pre}, \mathrm{mid}\}$, we compute mask-weighted average:

$$P_t^c \;=\; \frac{\sum\limits_x M_t^c(x)\, A_t(x)}{\sum\limits_x M_t^c(x) + \epsilon}, \qquad Q_t^c \;=\; \frac{\sum\limits_x M_t^c(x)\, R_t(x)}{\sum\limits_x M_t^c(x) + \epsilon}, \tag{1}$$

where x indexes voxels, M_t^c is the binary mask of class c at time t, and ϵ avoids division by zero. Importantly, these masks are used only during training to form region prototypes; no annotations are required at inference.

Stable anatomical structures serve as the reference for interpreting tumor evolution. We therefore enforce temporal consistency of background prototypes in the anatomical stream using an InfoNCE loss with in-batch negatives:

$$\mathcal{L}_{\mathrm{ana}} = -\log \frac{\exp\Big(\mathrm{sim}(\bar{P}^{\mathrm{bg}}_{\mathrm{pre}}, \bar{P}^{\mathrm{bg}}_{\mathrm{mid}})/\tau\Big)}{\sum\limits_{j=1}^{B} \exp\Big(\mathrm{sim}(\bar{P}^{\mathrm{bg}}_{\mathrm{pre}}, \bar{P}^{\mathrm{bg},(j)}_{\mathrm{mid}})/\tau\Big)}, \tag{2}$$

where $\bar{P}$ denotes ℓ_2-normalized prototypes, $\mathrm{sim}(\cdot,\cdot)$ is cosine similarity, τ is a temperature parameter, B is the mini-batch size, and negatives are background prototypes from other patients within the batch. This objective promotes a time-invariant anatomical embedding in A_{pre} and A_{mid}, reflecting clinical anchoring to stable anatomy. Since GTVp often regresses markedly while GTVn changes more subtly under edema/inflammation, the response stream should remain stable on anatomical anchors yet preserve non-trivial temporal variation within tumor regions; we therefore measure prototype residual magnitudes:

$$\Delta_{\mathrm{bg}} = \Big\|\bar{Q}^{\mathrm{bg}}_{\mathrm{mid}} - \bar{Q}^{\mathrm{bg}}_{\mathrm{pre}}\Big\|_2, \Delta_{\mathrm{p}} = \big\|\bar{Q}^{\mathrm{p}}_{\mathrm{mid}} - \bar{Q}^{\mathrm{p}}_{\mathrm{pre}}\big\|_2, \Delta_{\mathrm{n}} = \big\|\bar{Q}^{\mathrm{n}}_{\mathrm{mid}} - \bar{Q}^{\mathrm{n}}_{\mathrm{pre}}\big\|_2, \tag{3}$$

and impose a margin that encourages (i) low temporal drift for background anchors and (ii) non-trivial temporal signatures for both targets:

$$\mathcal{L}_{\mathrm{resp}} = \Delta_{\mathrm{bg}} + \frac{1}{2}\Big(\max(0,\, m - \Delta_{\mathrm{p}}) + \max(0,\, m - \Delta_{\mathrm{n}})\Big), \tag{4}$$

where $m > 0$ is a margin. Together, Eq. (2)–(4) realize region-aware disentanglement: the anatomical stream is anchored to stable background, while the response stream preserves learnable temporal change in target regions, aligning with the asymmetric nature of ART contour refinement.

2.3 Temporal-Difference Response Encoding

Clinicians attribute midRT deviations from the preRT plan to cumulative treatment effects; accordingly, we encode radiation response as a temporal residual in the response stream to calibrate midRT anatomical representations, computing a residual feature volume:

$$\Delta R \;=\; R_{\mathrm{mid}} - R_{\mathrm{pre}}, \tag{5}$$

and summarize it into a compact treatment-state representations:

$$Z \;=\; h(\mathrm{GAP}(\Delta R))\,, \tag{6}$$

where $\mathrm{GAP}(\cdot)$ is global average pooling and $h(\cdot)$ is a lightweight multilayer perceptron (MLP); despite ΔR being spatially distributed, the global code Z summarizes patient-level response context while preserving spatial structure in the anatomical stream. We inject Z via Feature-wise Linear Modulation (FiLM) by applying channel-wise affine modulation to the midRT anatomical features:

$$\widetilde{A}_{\mathrm{mid}} \;=\; \gamma(Z) \odot A_{\mathrm{mid}} + \beta(Z), \tag{7}$$

where $\gamma(\cdot)$ and $\beta(\cdot)$ are learned linear maps producing per-channel scale and shift, and $\odot$ denotes the Hadamard product. TDRE thus disentangles longitudinal cues into (i) a spatially structured anatomical baseline A_{mid} and (ii) a compact response code Z summarizing systematic treatment effects, enabling the decoder to interpret midRT appearance changes in a plan-consistent coordinate system.

2.4 Plan-Guided Asymmetric Regularization

In ART, the preRT plan provides canonical target semantics that midRT adaptation should respect despite deformation and appearance shifts; thus, we train a plan-optimized teacher on the anatomical stream and impose an asymmetric (plan→adapt) constraint on midRT predictions, first learning a planning decoder D_{plan} from A_{pre}:

$$\widehat{Y}^{\mathrm{plan}}_{\mathrm{pre}} = D_{\mathrm{plan}}(A_{\mathrm{pre}}), \qquad \mathcal{L}_{\mathrm{pre}} = \mathcal{L}_{\mathrm{seg}}(\widehat{Y}^{\mathrm{plan}}_{\mathrm{pre}}, Y_{\mathrm{pre}}), \tag{8}$$

and then freeze D_{plan} to serve as a fixed semantic reference. During joint optimization, the teacher produces a plan-consistent midRT distribution from the midRT anatomical stream:

$$\widehat{Y}^{\mathrm{plan}}_{\mathrm{mid}} = D_{\mathrm{plan}}(A_{\mathrm{mid}}). \tag{9}$$

The plan constraint should act strongly in regions whose semantics are expected to remain invariant (stable anatomy), while relaxing in tumor-bearing regions where genuine regression/migration is expected. To avoid using midRT labels to define this weighting, we derive an anchor mask from the preRT plan:

$$\Omega_{\mathrm{anc}} = \{x \in \Omega \mid x \notin \mathrm{Dilate}(\Omega^{\mathrm{pre}}_{\mathrm{tum}}, r)\}\,, \quad \Omega^{\mathrm{pre}}_{\mathrm{tum}} = \{x \in \Omega \mid Y_{\mathrm{pre}}(x) \neq \mathrm{bg}\}\,, \tag{10}$$

where Ω is the voxel grid and $\mathrm{Dilate}(\cdot, r)$ denotes morphological dilation with fixed radius r; thus, Ω_{anc} preserves reliable background while excluding a conservative margin around preRT targets to allow plausible evolution. We then regularize the student $\widehat{Y}_{\mathrm{mid}} = D(\widetilde{A}_{\mathrm{mid}}, Z)$ toward the frozen plan teacher only within Ω_{anc}:

$$\mathcal{L}_{\mathrm{plan}} = \sum_{x \in \Omega} w(x)\, \mathrm{KL}\Big(\widehat{Y}^{\mathrm{plan}}_{\mathrm{mid}}(x) \,\|\, \widehat{Y}_{\mathrm{mid}}(x)\Big)\,, \qquad w(x) = \mathbf{1}[x \in \Omega_{\mathrm{anc}}], \tag{11}$$

where $\mathrm{KL}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence. The PGAR constraint is designed to be asymmetric, anchoring midRT predictions to plan-derived semantics within stable anatomical regions without requiring preRT features to explain treatment-induced midRT variations, thereby allowing target regions to evolve during therapy. Importantly, PGAR is used only during training. The anchor region Ω_{anc} is derived from the preRT planning label to regularize reliable non-target areas, while Y_{pre}, Y_{mid}, Ω_{anc}, and the frozen plan teacher are not required at inference. After training, the student directly predicts $\widehat{Y}_{\mathrm{mid}}$ from $(I_{\mathrm{pre}}, I_{\mathrm{mid}})$, indicating that the plan prior is distilled into the learned representation and decoder while maintaining label-free deployment.

2.5 Segmentation Objective

We supervise midRT prediction using a standard composite segmentation loss:

$$\mathcal{L}_{\mathrm{seg}} = \mathrm{CE}(\widehat{Y}_{\mathrm{mid}}, Y_{\mathrm{mid}}) + \mathrm{Dice}(\widehat{Y}_{\mathrm{mid}}, Y_{\mathrm{mid}}), \tag{12}$$

where $\mathrm{CE}(\cdot,\cdot)$ is voxel-wise cross-entropy and $\mathrm{Dice}(\cdot,\cdot)$ is the soft Dice loss over K classes. The full objective combines segmentation supervision with the proposed regularizers:

$$\mathcal{L} = \lambda_{\mathrm{seg}}\mathcal{L}_{\mathrm{seg}} + \lambda_{\mathrm{ana}}\mathcal{L}_{\mathrm{ana}} + \lambda_{\mathrm{resp}}\mathcal{L}_{\mathrm{resp}} + \lambda_{\mathrm{plan}}\mathcal{L}_{\mathrm{plan}}. \tag{13}$$

Across all experiments, the four loss terms used equal weights ($\lambda_{\mathrm{seg}} = \lambda_{\mathrm{ana}} = \lambda_{\mathrm{resp}} = \lambda_{\mathrm{plan}} = 1.0$). We set $\tau = 0.1$ in Eq. (2), $m = 1.0$ in Eq. (4), and $r = 5$ voxels in Eq. (10) after resampling to 1-mm isotropic spacing. All hyperparameters were fixed across folds, ablations, and comparisons without test-set tuning.

3 Experiments

3.1 Experimental Setup

We evaluate RaD-Seg on the public HNTS-MRG 2024 longitudinal H&N MRI cohort[4] (N=150 paired preRT/midRT T2w scans with expert GTVp/GTVn labels) using patient-level K-fold cross-validation (default K=5) with fixed folds for fair comparison and paired testing. Volumes are resampled to 1 mm isotropic spacing and min–max normalized, without additional DIR beyond the provided pairs. We train on an RTX A6000 (48 GB) with AdamW (learning rate 2×10^{-4}, weight decay 1×10^{-5}), and cosine annealing for 500 epochs, following a two-stage schedule: pretrain D_{plan} on preRT with $\mathcal{L}_{\mathrm{seg}}$, then freeze it and optimize the full model on midRT; PGAR uses Ω_{anc} derived from the preRT plan (Eq. (10)), and inference requires only $(I_{\mathrm{pre}}, I_{\mathrm{mid}})$. All baselines share the same folds and inputs, and we report Dice similarity coefficient (DSC, %) and 95$^{\mathrm{th}}$ percentile Hausdorff distance (HD_{95}, mm) for GTVp/GTVn with two-sided Wilcoxon signed-rank tests on paired per-patient scores ($p < 0.05$). Results are reported as mean±std (median) across patients and then averaged over cross-validation folds.

[4] `https://hntsmrg24.grand-challenge.org/overview/`

Table 1. Ablation study. † indicates that RaD-Seg significantly outperforms the corresponding variant ($p < 0.05$). The best results are highlighted in **bold**.

Method Variant	GTVp		GTVn	
	DSC (%) ↑	HD_{95} (mm) ↓	DSC (%) ↑	HD_{95} (mm) ↓
Baseline (concat. preRT+midRT)	74.43 ± 12.53 (74.17)†	7.99 ± 5.31 (7.27)†	66.12 ± 15.47 (68.78)†	11.52 ± 6.73 (10.57)†
+ RDC (anatomy anchoring)	76.13 ± 12.87 (77.52)†	6.14 ± 3.56 (5.23)†	67.12 ± 17.58 (69.84)†	10.15 ± 5.42 (8.97)†
+ TDRE (response residual)	74.82 ± 13.91 (76.53)†	7.23 ± 3.94 (6.42)†	68.23 ± 16.82 (70.14)†	9.21 ± 4.83 (8.12)†
+ PGAR (plan semantics)	75.46 ± 13.14 (76.92)†	6.93 ± 3.82 (6.15)†	67.51 ± 17.06 (69.57)†	9.82 ± 5.14 (8.63)†
TDRE + PGAR	75.92 ± 12.54 (77.26)†	6.53 ± 3.31 (5.82)†	69.13 ± 14.82 (70.83)†	8.05 ± 4.12 (7.14)†
RDC + TDRE	76.84 ± 11.23 (77.93)†	5.26 ± 2.94 (4.52)†	68.91 ± 15.12 (70.52)†	8.13 ± 4.24 (7.23)†
RDC + PGAR	77.12 ± 10.57 (78.24)†	4.82 ± 2.64 (4.13)†	67.92 ± 16.23 (69.91)†	9.15 ± 4.63 (8.24)†
RaD-Seg (Ours)	**77.35 ± 9.81 (78.43)**	**4.25 ± 2.12 (3.51)**	**69.48 ± 13.20 (71.13)**	**7.55 ± 3.50 (6.20)**

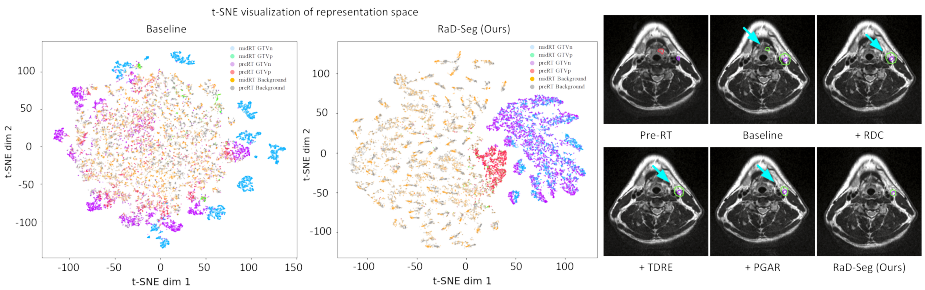

Fig. 2. Ablation analysis of RaD-Seg. Left: t-SNE on the test set shows improved feature disentanglement, with clearer separation of anatomy-anchored and response-sensitive embeddings. **Right:** MidRT GTVp/GTVn ablations show progressive contour refinement from RDC, TDRE, and PGAR. Arrows highlight reduced boundary leakage and false positives, yielding more clinically plausible delineation under deformation and edema-related shifts (GT GTVp: red; GT GTVn: purple; predicted GTVn: green).

4 Results

4.1 Ablation Study

Table 1 and Fig. 2 show that RDC, TDRE, and PGAR provide complementary gains. Using the same 3D ResNet encoder–decoder with concatenated preRT–midRT inputs and only $\mathcal{L}_{\text{seg}}$, the ablation baseline isolates module effects without changing capacity. RDC mainly improves geometric accuracy through anatomy anchoring (GTVp HD_{95}: 7.99 → 6.14 mm; GTVn: 11.52 → 10.15 mm), whereas TDRE enhances temporal response modeling and nodal delineation under edema-confounded changes (GTVn DSC: 66.12 → 68.23%; HD_{95}: 11.52 → 9.21 mm). PGAR further stabilizes midRT predictions by preserving reliable plan semantics while allowing plausible target evolution. Among two-module variants, TDRE + PGAR gives the best nodal results, while RDC+PGAR achieves the best primary-target geometry. The full RaD-Seg produces more structured embeddings, tighter contours, and significantly outperforms all ablated variants (paired Wilcoxon, $p < 0.05$).

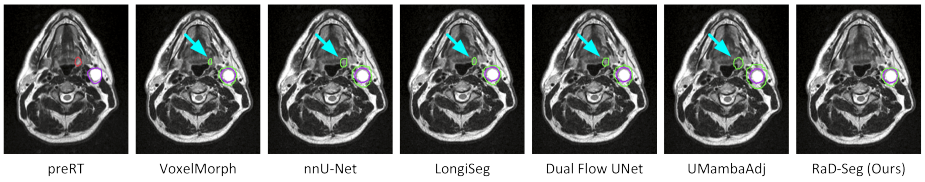

Fig. 3. Qualitative midRT segmentation comparison. SOTA methods show over-delineation in arrow-marked regions, whereas RaD-Seg yields more clinically plausible contours by modeling treatment response and enforcing plan-guided asymmetry. PreRT images are shown only as longitudinal references; all predictions are evaluated on midRT segmentation (GT GTVp: red; GT GTVn: purple; predicted GTVn: green).

Table 2. Quantitative comparison with SOTA methods. † indicates that RaD-Seg significantly outperforms the corresponding method ($p < 0.05$).

Method	GTVp		GTVn	
	DSC (%) ↑	**HD_{95} (mm) ↓**	**DSC (%) ↑**	**HD_{95} (mm) ↓**
VoxelMorph [2]	76.54 ± 12.30 (78.13)†	6.82 ± 3.15 (5.98)†	67.20 ± 15.40 (69.54)†	10.45 ± 5.20 (9.11)†
nnU-Net [3]	73.12 ± 14.84 (75.25)†	8.15 ± 4.52 (7.21)†	64.89 ± 17.11 (66.49)†	12.30 ± 6.14 (10.87)†
LongiSeg [4]	75.43 ± 13.10 (77.05)†	7.17 ± 3.81 (6.41)†	66.56 ± 16.20 (68.15)†	11.05 ± 5.50 (9.84)†
Dual Flow UNet [15]	77.15 ± 11.52 (78.35)†	6.45 ± 2.91 (5.57)†	68.17 ± 14.80 (70.21)†	9.81 ± 4.85 (8.53)†
UMambaAdj [10]	74.21 ± 14.51 (76.12)†	7.81 ± 4.22 (6.91)†	66.54 ± 18.20 (69.36)†	10.97 ± 5.97 (9.54)†
RaD-Seg (Ours)	$\mathbf{77.35 \pm 9.81}$ (**78.43**)	$\mathbf{4.25 \pm 2.12}$ (**3.51**)	$\mathbf{69.48 \pm 13.20}$ (**71.13**)	$\mathbf{7.55 \pm 3.50}$ (**6.20**)

4.2 Comparison to State-of-the-Art

We compare RaD-Seg with representative ART/longitudinal baselines, including VoxelMorph [2], nnU-Net [3], LongiSeg [4], Dual Flow UNet [15], and UMambaAdj [10]. VoxelMorph is used for registration-based contour propagation by warping preRT images and GTVp/GTVn labels to midRT space. As shown in Table 2, RaD-Seg achieves the best overall performance, with the highest DSC and lowest HD_{95} for both GTVp (77.35%/4.25 mm) and GTVn (69.48%/7.55 mm). Compared with the strongest baseline, Dual Flow UNet, RaD-Seg improves both overlap and boundary accuracy, reflecting its anatomy anchoring and response-aware modeling. Fig. 3 shows reduced drift and leakage with more plausible contours. All improvements are significant (paired Wilcoxon, $p < 0.05$).

5 Conclusion

Overall, RaD-Seg follows the plan–verify–adapt paradigm by integrating stable anatomy anchoring, explicit response modeling, and plan-guided asymmetric regularization to stabilize midRT segmentation while preserving plausible target evolution. Experiments on HNTS-MRG 2024 show statistically significant SOTA performance for both GTVp and GTVn, with ablations confirming each module's contribution. Future work will extend RaD-Seg toward uncertainty-aware clinician-in-the-loop adaptation and robust multi-sequence, multi-center deployment.

Acknowledgments. This study was funded by National Key Research and Development Program of China (2022YFC2405200), National Natural Science Foundation of China (U22A2051, 62271246), and Natural Science Foundation of Beijing Municipality (L252201, L252126).

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

References

1. Andrearczyk, V., Oreiller, V., Boughdad, S., Rest, C.C.L., Elhalawani, H., Jreige, M., Prior, J.O., Vallières, M., Visvikis, D., Hatt, M., et al.: Overview of the hecktor challenge at miccai 2021: automatic head and neck tumor segmentation and outcome prediction in pet/ct images. In: 3D head and neck tumor segmentation in PET/CT challenge, pp. 1–37. Springer (2021)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
3. 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)
4. Kächele, J., Zenk, M., Rokuss, M., Ulrich, C., Wald, T., Maier-Hein, K.H.: Enhanced nnu-net architectures for automated mri segmentation of head and neck tumors in adaptive radiation therapy. In: Challenge on Head and Neck Tumor Segmentation for MRI-Guided Applications, pp. 50–64. Springer (2024)
5. Khor, H.G., Chen, F., Liao, H.: Cfm-sam2: Cnn-prompt-guided segment anything 2 with cross-modal fusion and contextual feature assimilation for gross target volume segmentation in nasopharyngeal carcinoma. In: Proceedings of the 2025 2nd International Conference on Smart Healthcare and Wearable Intelligent Devices. pp. 149–154 (2025)
6. Khor, H.G., Yang, X., Sun, Y., Huang, S., Wang, Y., Wang, J., Wang, S., Bai, L., Ma, L., Liao, H.: Rt-sam: Visual-prompt fusion and uncertainty enhancement for nasopharyngeal carcinoma radiotherapy target delineation. IEEE Journal of Biomedical and Health Informatics (2026)
7. Khor, H.G., Yang, X., Sun, Y., Wang, J., Huang, S., Wang, S., Lu, B., Ma, L., Liao, H.: Unified prompt-visual interactive segmentation of clinical target volume in ct for nasopharyngeal carcinoma with prior anatomical information. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 659–669. Springer (2024)
8. Li, Z., Zhang, W., Li, B., Zhu, J., Peng, Y., Li, C., Zhu, J., Zhou, Q., Yin, Y.: Patient-specific daily updated deep learning auto-segmentation for mri-guided adaptive radiotherapy. Radiotherapy and Oncology **177**, 222–230 (2022)
9. McDonald, B.A., Cardenas, C.E., O'Connell, N., Ahmed, S., Naser, M.A., Wahid, K.A., Xu, J., Thill, D., Zuhour, R.J., Mesko, S., et al.: Investigation of autosegmentation techniques on t2-weighted mri for off-line dose reconstruction in mr-linac workflow for head and neck cancers. Medical physics **51**(1), 278–291 (2024)
10. Ren, J., Hochreuter, K., Kallehauge, J.F., Korreman, S.S.: Umamba adjustment: Advancing gtv segmentation for head and neck cancer in mri-guided rt with umamba and nnu-net resenc planner. In: Challenge on Head and Neck Tumor Segmentation for MRI-Guided Applications, pp. 123–135. Springer (2024)

11. Ren, J., Hochreuter, K., Rasmussen, M.E., Kallehauge, J.F., Korreman, S.S.: Gradient map-assisted head and neck tumor segmentation: a pre-rt to mid-rt approach in mri-guided radiotherapy. In: Challenge on Head and Neck Tumor Segmentation for MRI-Guided Applications, pp. 36–49. Springer (2024)
12. Tie, X., Chen, W., Huemann, Z., Schott, B., Liu, N., Bradshaw, T.J.: Deep learning for longitudinal gross tumor volume segmentation in mri-guided adaptive radiotherapy for head and neck cancer. In: Challenge on Head and Neck Tumor Segmentation for MRI-Guided Applications, pp. 99–111. Springer (2024)
13. Vizitiu, A., Mohaiu, A.T., Popdan, I.M., Balachandran, A., Ghesu, F.C., Comaniciu, D.: Multi-scale self-supervised learning for longitudinal lesion tracking with optional supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 573–582. Springer (2023)
14. Wahid, K.A., Dede, C., El-Habashy, D.M., Kamel, S., Rooney, M.K., Khamis, Y., Abdelaal, M.R., Ahmed, S., Corrigan, K.L., Chang, E., et al.: Overview of the head and neck tumor segmentation for magnetic resonance guided applications (hnts-mrg) 2024 challenge. In: Challenge on Head and Neck Tumor Segmentation for MRI-Guided Applications, pp. 1–35. Springer (2024)
15. Wang, L., Liao, W., Zhang, S., Wang, G.: Head and neck tumor segmentation of mri from pre-and mid-radiotherapy with pre-training, data augmentation and dual flow unet. In: Challenge on Head and Neck Tumor Segmentation for MRI-Guided Applications, pp. 75–86. Springer (2024)