Closing the loop: fully automated radiotherapy planning for head and neck cancer

Gao, R.

6

Summary and Future Work

6.1 Summary

We investigated automated radiotherapy planning for head and neck cancer, with a primary focus on deep learning–based dose prediction and the generation of clinically deliverable VMAT plans through dose mimicking. We also explored generalizable medical image segmentation methods that are relevant to radiotherapy and broader medical imaging applications. Across this thesis, we addressed key components of automated radiotherapy planning, including accurate dose prediction, clinically guided model optimization, and integration into a commercial treatment planning system (TPS).

Chapter 2 presented MCP-MedSAM, a lightweight and modality-agnostic medical image segmentation foundation model. To improve segmentation performance across different imaging modalities while maintaining computational efficiency, we introduced modality prompts and content prompts that provide additional task-relevant information to the model with minimal overhead. We further adopted a modality-based data sampling strategy to alleviate data imbalance across modalities. The results demonstrated that MCP-MedSAM achieved strong segmentation performance on a large-scale benchmark while remaining trainable on a single A100 GPU within one day, providing a practical and scalable framework for general medical image segmentation.

Chapter 3 investigated key factors influencing deep learning–based 3D dose prediction for head and neck photon radiotherapy. We systematically evaluated the effects of input and dose-grid resolution, input representation, loss function, model architecture, and robustness to noise using both a public dataset and an in-house clinical dataset. The study showed that high-resolution and anatomically informative inputs, including CT, PTVs, and OARs, substantially improved prediction accuracy. We also found that combining voxel-wise and DVH-oriented losses improved clinically relevant performance, while robustness analyses revealed that most models were relatively robust to Poisson noise but more sensitive to adversarial perturbations. These findings provide practical guidance for designing accurate and reliable dose prediction models.

Chapter 4 proposed a clinical DVH metric loss (CDM loss) for 3D dose prediction, aiming to better align model training with clinical plan evaluation criteria. Instead of relying only on voxel-wise supervision, the proposed loss directly incorporated differentiable formulations of commonly used clinical DVH metrics, including both *D–metrics* and surrogate formulations of *V–metrics*. In addition, we introduced a lossless bit-mask encoding scheme to efficiently represent multiple overlapping ROIs and thereby substantially reduce computational cost during training. The results showed that the CDM loss improved target coverage and satisfied all clinical constraints while maintaining comparable OAR sparing, demonstrating that clinically guided supervision can produce dose predictions that are better aligned with treatment planning requirements.

Chapter 5 integrated the trained dose prediction model into a commercial TPS (RayStation) and established a fully automated planning pipeline based on dose mimicking. Predicted dose distributions were imported into the TPS and used to guide dose mimicking optimization for the generation of clinically deliverable dual-arc VMAT plans. On an independent temporal test cohort, the automated plans achieved target coverage comparable to clinical plans, with similar or improved sparing for several OARs and no significant differences in NTCP estimates for xerostomia and dysphagia. In addition, the automatically generated plans showed lower modulation-related complexity for several plan complexity metrics. These results demonstrated the feasibility of combining deep learning–based dose prediction with dose mimicking to generate clinically deliverable head and neck radiotherapy plans in a fully automated manner.

6.2 Discussion and future work

The work presented in this thesis supports the feasibility of data-driven radiotherapy planning for head and neck cancer, from clinically guided dose prediction to deliverable VMAT plans. Taken together, these results point toward automated planning frameworks that more closely link patient anatomy, dose prediction, and delivery feasibility.

Although this work focuses on head and neck cancer, it has the potential to generalize to other treatment sites, particularly those with less complex anatomical and planning requirements. While such extensions would require additional data collection and validation, radiotherapy provides a favorable setting for data-driven approaches, as high-quality manual treatment plans have long been generated in routine clinical practice and can serve as training data. Even as automated planning becomes more widespread, clinician-reviewed and modified plans may continue to provide valuable supervision, enabling iterative model improvement over time.

From a practical perspective, multiple technical pathways exist for automated planning, including approaches such as iCycle [118] and ECHO [119]. The approach

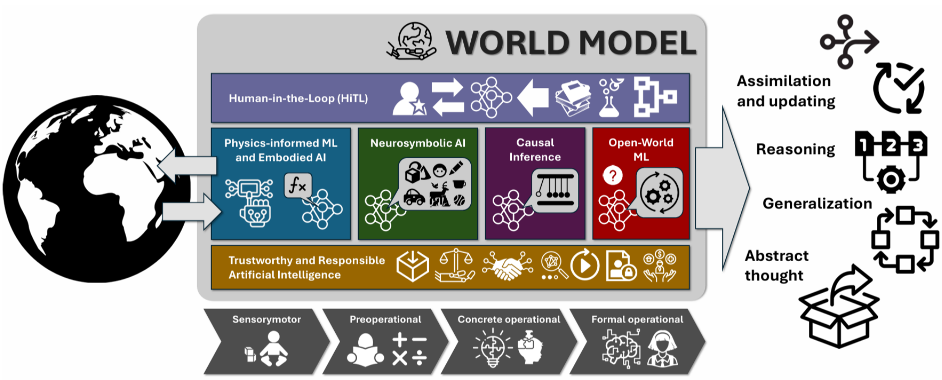

Figure 6.1: A world model built upon six research areas: (1) embodied AI and physics-informed machine learning, allowing AI systems to interact with the physical world; (2) neuro-symbolic approaches that integrate statistical learning with symbolic representations; (3) causal inference to capture action–effect relationships in observed data; (4) open-world machine learning to enable adaptation to unknown situations; (5) human-in-the-loop approaches that incorporate human knowledge and oversight; and (6) trustworthy and responsible AI to ensure accountability and reliability. Together, these research areas illustrate possible components of future reasoning-capable and human-centered AI systems. Figure reproduced from [117].

explored in this thesis, based on dose prediction and dose mimicking within a commercial TPS, offers advantages in terms of computational efficiency and clinical integration. In the head and neck cases studied here, fully automated plan generation can be achieved in approximately 15 minutes, suggesting that such methods may provide a pragmatic balance between plan quality, robustness, and speed in real-world clinical workflows.

From a broader artificial intelligence perspective, future planning systems may be viewed through the lens of "world model", that is, integrated representations of the environment, its constraints, and the consequences of possible actions [117]. As illustrated in Figure 6.1, such systems may combine embodied and physics-informed learning, neuro-symbolic reasoning, causal inference, open-world learning, human-in-the-loop interaction, and trustworthy and responsible AI. This perspective provides a useful framework for interpreting the future directions discussed below.

In radiotherapy, the relevant "world" includes patient anatomy, the physical processes governing radiation transport and dose deposition, the mechanical constraints of treatment delivery systems, and the clinical judgments involved in balancing target coverage against organ sparing [3, 7, 18]. Within this context, end-to-end planning and personalized, interactive planning can be broadly viewed as complementary steps toward a radiotherapy-specific world model, with particular relevance to embodied and physics-informed learning, human-in-the-loop guidance, and trustworthy and responsible AI.

6.2.1 Toward end-to-end radiotherapy planning

The framework developed in Chapter 5 reduces the need for manual trial-and-error, but it still consists of sequential components: although dose prediction can be substantially improved through appropriate input design and clinically guided loss functions (Chapters 3 and 4), the predicted dose distributions must still be translated into deliverable plans through dose mimicking within a commercial TPS. This separation limits the integration of learning-based models with the physics of treatment delivery. A natural next step is therefore the development of more unified end-to-end planning frameworks.

Radiotherapy planning is fundamentally governed by well-established physical models. In conventional systems, dose calculation is based on beamlet–voxel formulations, and optimization algorithms search for machine parameters that satisfy clinical objectives while respecting delivery constraints [18, 19]. These models are reliable and interpretable, but they are also computationally expensive and typically rely on iterative optimization. Integrating physics-based planning with modern machine learning remains a major challenge, because accurate dose calculation, deliverability constraints, and clinical objective trade-offs must be modeled jointly within a computationally efficient and clinically reliable framework.

The results in Chapters 3 and 4 suggest that learned dose prediction models can capture clinically relevant relationships between anatomy and achievable dose, particularly when anatomically informative inputs and clinically motivated supervision are used. A natural next step is to extend such models so that they not only predict voxel dose, but also reason about the physical and mechanical constraints of treatment delivery. This may involve combining image-derived anatomical information with representations of dose deposition and delivery constraints. Examples include differentiable or surrogate dose calculation models [120], physics-informed neural networks [26, 27], and learned optimization policies [8, 121]. Explicit modeling of delivery parameters such as MLC trajectories, gantry motion, and dose-rate modulation may allow future systems to reason jointly about dose quality and delivery feasibility, thereby reducing reliance on intermediate steps such as dose mimicking and enabling faster, potentially sub-minute, end-to-end plan generation.

A practical limitation, however, is that the pipeline in Chapter 5 depends on a commercial TPS, where key components such as dose calculation and optimization engines are implemented as closed-source software. As a result, researchers can interact with these systems mainly through predefined scripting interfaces and can influence the optimization process only indirectly through objective settings and constraints. This limits the extent to which learning-based methods can be integrated with the underlying physical planning process.

These limitations are particularly relevant for applications such as online adaptive radiotherapy, where rapid replanning is required to account for daily anatomical variation [80, 9]. In such settings, the need for reliable and scalable anatomical delineation also becomes especially important, linking future end-to-end planning not only to the dose prediction work in Chapters 3–5, but also to the segmentation ideas developed in Chapter 2. Existing optimization pipelines remain computationally demanding, and many TPS engines still rely heavily on CPU-based iterative solvers. Advances in GPU-accelerated dose calculation, differentiable optimization, and hybrid physics-informed learning approaches may help make real-time or near-real-time automated replanning more practical. At the same time, any move toward end-to-end planning must still ensure physical consistency, interpretability, safety, and regulatory compliance.

6.2.2 Personalized and interactive planning through human preferences and human–AI collaboration

Although automation can reduce manual workload, radiotherapy planning remains inherently dependent on clinical judgment and preference variability. Different clinicians may prioritize target coverage, organ-at-risk sparing, toxicity risk, or quality-of-life considerations differently depending on patient condition, prognosis, and institutional practice [20, 21]. These variations are difficult to encode in a fixed set of optimization objectives, motivating the need for systems that can both represent clinician intent and support interactive decision-making.

Although Chapters 3 and 4 did not directly address inter-clinician preference variability, they show that automated planning performance depends not only on anatomy, but also on how clinical priorities are represented. Chapter 3 highlighted the importance of input and supervision design, while Chapter 4 showed that explicitly encoding clinical evaluation criteria in the loss function can improve the clinical relevance of predicted dose distributions.

One possible direction is preference-conditioned or prompt-based planning, in which clinicians provide high-level guidance to steer automated plan generation [122, 123]. Such prompts could include, for example, prioritizing specific organs-at-risk, specifying acceptable toxicity thresholds, selecting planning templates, or defining trade-offs between competing objectives (e.g., improved parotid sparing at the expense of target homogeneity). Extending this paradigm may enable more flexible and user-adaptive planning systems.

Another direction is learning from clinician interaction and feedback. In routine workflows, automated plans are often reviewed and further refined by planners and other clinical staff involved in treatment planning. These refinements can be interpreted as signals of mismatch between model outputs and clinical expectations. Analogous to autonomous driving systems, where human interventions are treated as indicators of model failure and used to improve model performance, a similar feedback loop could be established in radiotherapy. Modifications to auto-generated plans could be systematically recorded and incorporated into model retraining, enabling continuous alignment with clinical practice.

Because radiotherapy planning is fundamentally a multi-criteria optimization problem, multiple Pareto-optimal solutions may be clinically acceptable [18]. A natural extension is therefore to generate multiple clinically plausible candidate plans that reflect different trade-offs between competing objectives. However, presenting multiple solutions introduces practical challenges. Clinicians are already subject to information overload, and displaying too many alternatives may reduce usability. Important questions include how many candidate plans should be presented, how to ensure that solutions are meaningfully diverse, and how to enable efficient navigation of the solution space. Possible directions include generating a limited set of representative solutions, using diversity-aware selection criteria, and developing interfaces that allow clinicians to navigate trade-offs efficiently without excessive cognitive burden [124].

To support effective human–AI collaboration, future systems may incorporate tools such as real-time DVH steering, rapid re-optimization under updated constraints, and uncertainty-aware plan comparison. Uncertainty may arise from multiple sources, including anatomical and delineation uncertainty, dose prediction model uncertainty, and optimization-related variability. Explicitly quantifying and communicating these uncertainties may improve interpretability and support more informed clinical decision-making.

Overall, future automated planning systems should not only generate high-quality plans, but also function as clinical decision-support tools that allow clinicians to explore, refine, and personalize treatment strategies in a controlled and interpretable manner.

6.3 General conclusions

This thesis demonstrates the potential of data-driven methods for automated radiotherapy planning in head and neck cancer, particularly for deep learning–based dose prediction and the generation of clinically deliverable VMAT plans through dose mimicking. It also highlights the broader relevance of generalizable medical image segmentation for scalable radiotherapy and medical imaging workflows.

Overall, the results show that clinically meaningful dose distributions can be learned from patient anatomy and translated into deliverable treatment plans within a commercial TPS. These findings support the feasibility of planning workflows that are more efficient and consistent, with less reliance on manual trial-and-error. Further progress will depend on continued advances in machine learning, physics-informed modeling, and human–AI collaboration, together with close collaboration among clinicians, medical physicists, and researchers.