Dose-Volume-Histogram Based Inverse Treatment Planning via Deep Reinforcement Learning in Intensity Modulated Radiation Therapy (IMRT) for Prostate Cancer

D.Sprouts[1], C.Shen[2], X.Jia[2] , Y.Chi[1]

[1]Dept. Of Physics, University of Texas at Arlington, Arlington, Tx 76019

[2]Dept. Of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75287

JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]
Improving Health QUALITY, Increasing Global IMPACT

INTRODUCTION

- In Intensity Modulated Radiation Therapy (IMRT), there can be millions of combinations of the multi-leaf collimator (MLC) positions to form a clinical acceptable dose distribution, which requires inverse treatment planning (ITP).
- ITP is technically challenging. Dose-to-volume-histogram(DVH) based ITP was found effective for the optimization and has been broadly used in modern treatment planning systems (TPSs).
- However, it can be still time consuming to form a clinically acceptable plan. Multi-rounds of parameter justifications with the TPSs would typically be needed.
- The quality of the obtained plan can heavily depend on the planning experience of the human planner and the allowed time length on it.

AIM

- To reduce the repetitive and time-consuming effect to fine tune the parameters by a human planner, in this study, we proposed to develop a deep neural network (DNN)-based virtual treatment planner(VTP) via end-end deep reinforcement learning (DRL). It is expected to mimic the behavior of human planners by automatically operating the DVH-based optimization engine for high-quality treatments plan.

METHOD

- The VTP was established with DRL using a Q-learning framework. Similar to human planners, VTP repetitively observes DVH of intermediate treatment plans and operates the in-house developed DVH-based TPS by adjusting treatment planning parameters (TPPs), such as changing volume and dose constraints, as well as the corresponding weights, to improve the plan quality (Figure 1).
- In an optimization engine with N TPPs, the VTP consists of N subnetwork (Fig. 2), with each subnetwork representing one TPP to adjust. In each step, only one TPP can be adjusted by the VTP. Based on the observations of the intermediate DVHs, VTP decided which TPP to adjust, together with the direction and magnitude of the adjustment.
- We trained VTP via end-to-end DRL with an experience replay mechanism under the Q-learning framework. Which allows for the VTP to look at the current stage which is the current DVHs and the DVHs after an action at M step and the action itself that was taken. Epsilon-greedy algorithm was implemented to explore the impacts of taking different actions for a large number of automatically generated plans. After the plans are generated a reward function is given to quantify the plan quality and this is incorporated to compare the plans before and after adjusting the TPPs. The VTP is updated to learn the appropriate TPP-tuning policy from which an optimal policy to improve plan quality can be learned.

- The Plan IQ score (ProKnow Systems,Stanford FL, USA) was used to quantify the plan quality. The developed TPS was designed to solve the following fluence map optimization problem:

$$\min_{x}\frac{1}{2}\|Mx-d_p\|_{-}^{2}+\frac{\lambda}{2}\|V_{ptv}(Mx-d_p)\|_{+}^{2}+\sum_{i}\frac{\lambda_i}{2}\|V_i(M_i x-\lambda_i^d d_p)\|_{+}^{2}, \text{s.t. } x \geq 0, D_{95\%}(Mx)=d_p.$$

  - M and M_i are the respected dose deposition matrices for the PTV and the considered OARs;
  - d_p is the prescription dose;
  - V_{PTV} and V_i are the volumes of the PTV and OARs.
- The testbed for the proposed framework was Intensity modulated radiotherapy (IMRT) for prostate cancer.

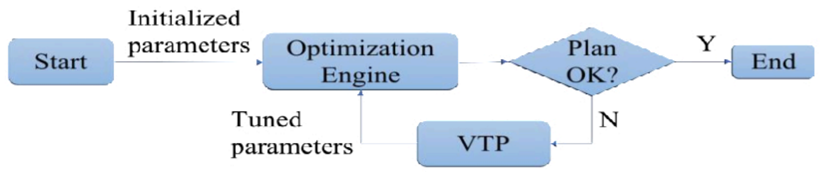

Figure 1. Overall workflow of the proposed human-like planning process. The workflow is similar to that of a human planner using a treatment planning system to develop a plan, but with VTP in lieu of the human planner.

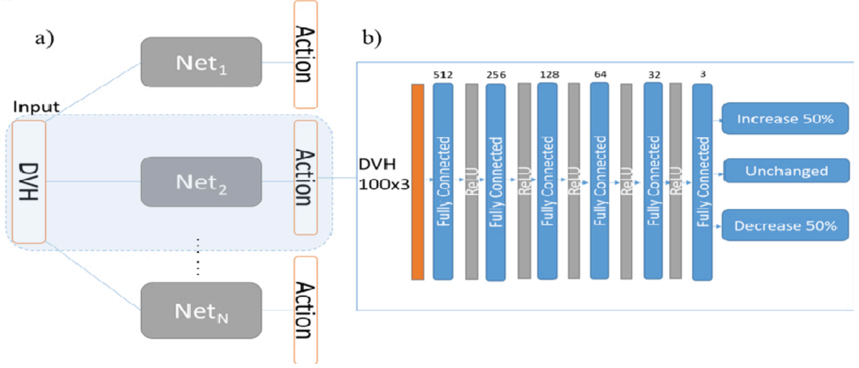

Figure 2 . Network structure of the VTP. (a) is the overall structure of VTP. The complete network consists of N subnetworks with identical structures. Each subnetwork corresponds to one TPP. The input is DVHs of a treatment plan. (b) Detailed structure for a subnetwork. Size after each fully connected layer is specified at the top. Output value of each network node is the Q function value of the corresponding action.

RESULTS

- The VTP was successfully trained using ten prostate cancer patient cases and then tested on another 12. The average initial plan score was 5 while the average planning score was 9 out of 10.
- Figure 3 shows the highest scoring patient's DVHs and dose map.
  - The Initial plan shows bad sparing for bladder and rectum, and violating some of the dose constraints that applied to the training.
  - In the VTP generated plan, the rectum and bladder is better
  - spared, with a planscore of 5. After the VTP training, a planscore of 10, maximum score, was obtained.
- Table 1 shows that the average volume constraints for all 12 cases
  - Bladder was well under the volume constraint. With only the V94.7 being closes to the volume constraints. All 12 cases were below the dose constraints. PAT 12 was the closes as in terms of violates but still was 18% volume lower than the constraint.
  - Rectum Volume percentage were closer to the constraint, but at the same time were well off constraints. All 12 test cases once again past the dose constraints. This time PAT 16 was the closest to violation of the dose constraints. It was 22% irradiated volume form violation.
  - PTV had good dose coverage of the PTV.

Table 1. The mean and standard deviation (std.) of the dose volume constraints (%) for OARs and the Prostate PTV over the 12 test cases computed based on the DVH obtained from the trained network. The requirement from the PlanIQ score is also listed.

	Bladder				Rectum				PTV
	V101	V94.7	V88.4	V82.1	V94.7	V88.4	V82.1	V75.8	V100
Mean	0.15	4.13	5.59	7.29	1.26	2.29	3.29	5.97	95.1
Std.	0.4	7.25	9.55	12.6	3.18	4.83	5.06	6.68	0.29
Required	<20	<30	<40	<55	<20	<30	<40	<55	>95

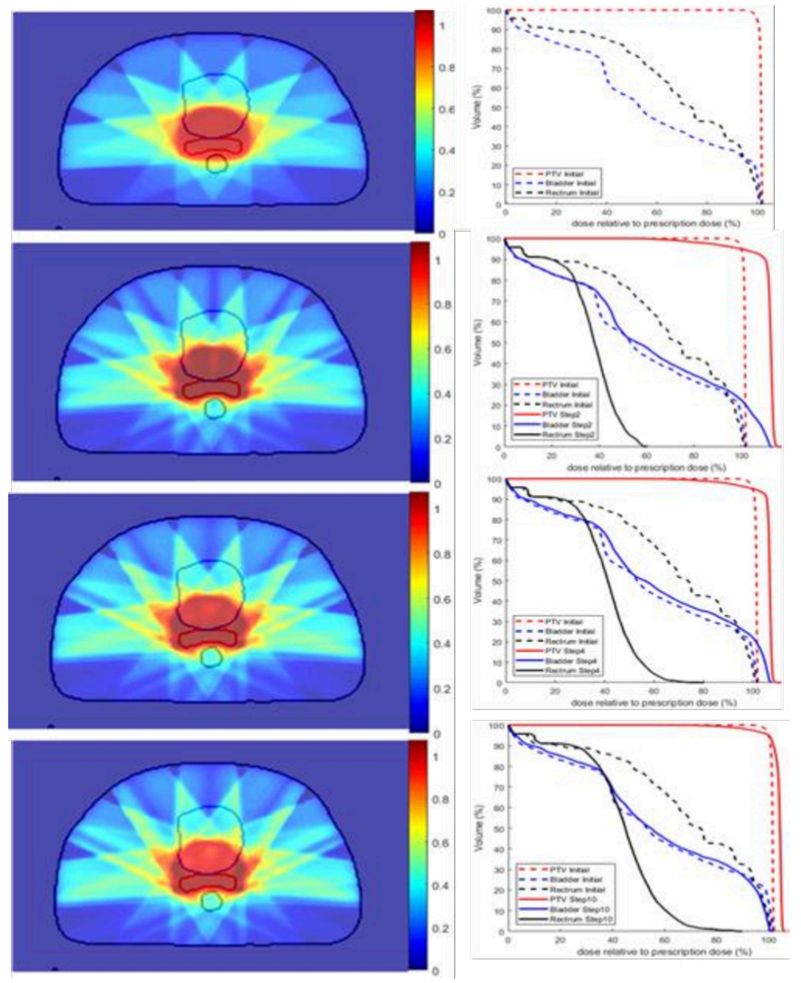

Figure 3. An illustration for the training of the VTP. 1st row: the initial dose maps and DVH distribution for the PTV and OARs. 2nd-4th rows: the dose maps of the same slice and the DVHs for plans generated by the VTP in steps 2, 4 and 10 at the trained epoch.

CONCLUSIONS

- The trained VTP is capable of automatically producing high-quality IMRT plans for prostate cancer via adjusting the TPPs intelligently in a human manner.
- The proposed approach is generally applicable to other cancer sites and treatment techniques.
- It has the potentially to be incorporated into the commercial TPSs to fully automate the treatment planning process.

REFERENCES

1. Oelfke U and Bortfield T, *Inverse photon planning for photon and proton beams.* Medical dosimetry, 2001.**26**:p.113-124
2. Mnih, V., et al., *Human-level control through deep reinforcement learning.* Nature, 2015. **518**(7540): p. 529.
3. Chenyang S., et al. *Intelligent inverse treatment planning via deep reinforcement learning, a proof-of-principle study in high dose-rate brachytherapy for cervical cancer,* Physics in Medicine & Biology, 2019.**64**

ACKNOWLEDGEMENTS

The research is partially supported by the UT system Rising Stars grant.

CONTACT INFORMATION

If interested, please contact Yujie Chi: Yujie.chi@uta.edu

or Chenyang Shen: Chenyang.Shen@utsouthwestern.edu