JULY 12–16 2020 VIRTUAL JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

Fast Monte Carlo Dose Estimation for Proton Therapy using a Dual-Pyramid Deep Learning Framework

Joseph Harms, Yang Lei, Serdar Charyyev, Alexander Stanforth, Jun Zhou, Liyong Lin, Walter J. Curran, Tian Liu and Xiaofeng Yang*

Department of Radiation Oncology and Winship Cancer Institute
Emory University, Atlanta GA, USA

EMORY UNIVERSITY

INTRODUCTION

The Monte Carlo (MC) algorithm has long been the gold standard for dose calculation in radiation therapy, especially for ion beams, and clinical treatment planning systems have recently added MC dose calculation models. Since computational time for MC varies with dose grid resolution and size, fine resolution calculations may take several hours to calculate, depending on the treatment field size. Pencil-beam algorithms can quickly calculate dose; however they can be inaccurate, especially near large heterogeneities. Thus, it is desirable to have a dose calculation algorithm with accuracy comparable to MC and computational time comparable to PB.

AIM

In this work, we adapt a dual-pyramid networks (DPNs) architecture to learn the differences between a pencil-beam (PB) based dose distribution and the more accurate MC calculation. This dual networks independently learns from the CT and input dose distribution and later selects the most significant features from each to generate an output dose distribution. Using two inputs, the proposed method can rely on the PB dose for a first approximation of the dose distribution and the structural information in the CT to fine-tune the dose distribution, achieving accuracy similar to the MC-calculated dose.

METHOD

- Dose from the clinically used ion beam plan for 20 prostate cancer patients being treated with intensity-modulated proton therapy (IMPT) were re-calculated using Pencil Beam (PB) algorithm in Raystation 9A with a 3x3x3 mm^3 dose grid and with MC using a 1x1x1 mm^3 dose grid at 0.5% statistical uncertainty
- The dual-pyramid network (DPN) architecture is shown in Fig. 1. Each individual pyramid network has a U-Net-like architecture for end-to-end synthesis
- The full dose distributions from each beam are fed into the network
- For increased variability in the training data, a script was developed to calculate dose for a series of shifts (5 mm ±x, ±y, and ±z) and rotations (3° ±roll and ±yaw) of the beam. In total, 10 plans, in addition to the nominal plans, were created per patient
- 2-fold cross-validation with 20 patient datasets were used for validation. The training took ~7.5 hours on on a NVIDIA Tesla V100 GPU with 32GB of memory, and the generation time for a new dose distribution was ~5 seconds

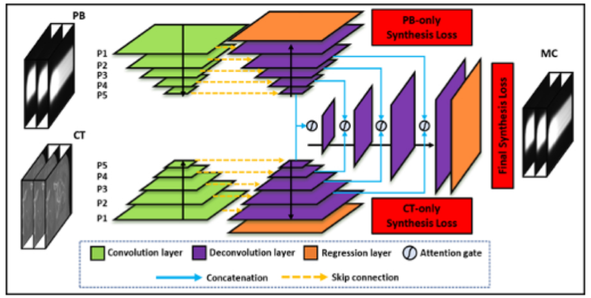

Fig. 1. Dual-pyramid networks architecture. The PB dose distribution calculated on a 3 mm^3 dose grid and CT image are the source data and the MC dose distribution with a 1 mm^3 dose grid is the target data. The pyramids on the left are trained on their corresponding input datasets independently, and this information is combined on the right in the late fusion network.

RESULTS

- DL-Generated and PB dose distributions were compared against MC-generated dose distributions, which were taken as the ground truth
- Metrics for evaluation:
  - Gamma analysis at 1%/1mm with a 10% threshold. Due to the sensitivity of gamma analysis to noise[1], the smoother dose distribution (either DL or PB) was used as the reference dose distribution
  - Structural similarity (SSIM), which has been suggested as a complement to Gamma for evaluation dose distributions[2,3]:

$$\mathrm{SSIM}(\mathrm{ref},\mathrm{eval}) = \frac{(2\mu_{\mathrm{ref}}\mu_{\mathrm{eval}}+C_1)(2\sigma_{\mathrm{ref,eval}}+C_2)}{(\mu_{\mathrm{ref}}^2+\mu_{\mathrm{eval}}^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}$$

  - where μ_{ref}, μ_{eval}, are local mean values for the reference and evaluated RSP maps, respectively, and σ_{ref}, σ_{eval}, and $\sigma_{ref,eval}$ and σ_{xy} are the standard deviations and cross-covariance for doses.
  - SSIM tests three quantitative image metrics: luminance, contrast, and structure. Peng et al showed that, when applied to dose distributions, these metrics can detect errors in absolute dose, gradient, and dose structures, respectively[3]
  - Mean error (ME) and mean absolute error (MAE)
  - All metrics were evaluated on a beam-by-beam and plan-by-plan basis
- The proposed DL-generated dose distributions very strongly reflected the MC dose distribution. Figs. 2 shows boxplots for all metrics, Fig. 3 shows dose distributions and gamma analysis for two sample cases, and Table 1 summarizes the numerical results.

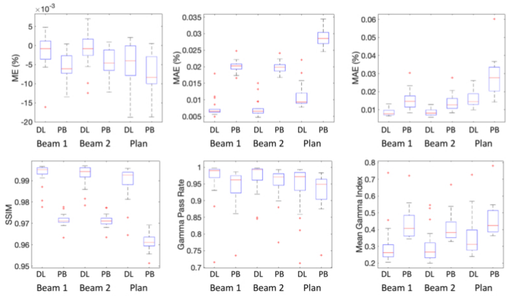

Fig. 2. Box and whisker plots for all comparison metrics. Both ME and MAE are expressed as percentage of the maximum dose, and MAE in the top right panel is calculated from only the high dose region (>90%).

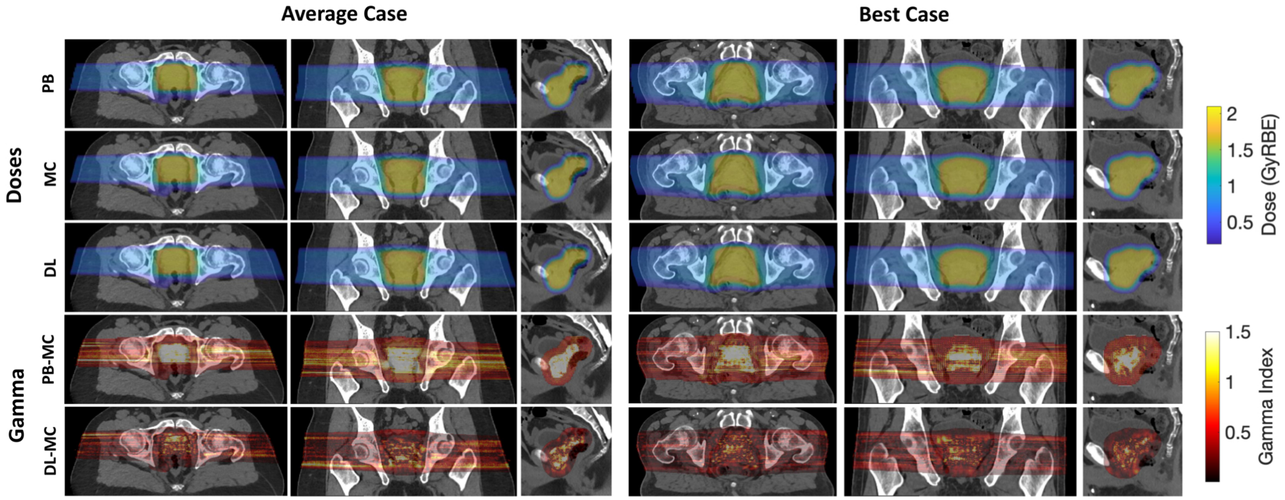

Fig. 3. Dose calculation and gamma distributions for a sample patient representing an average case (left 3 panes) and the best case (right 3 panes), based on passing gamma rate for the plan.

Table 1. Summary of quantitative analysis comparing both the DL-generated dose and PB-calculated dose to the MC-calculated dose. Both ME and MAE are expressed as percentage of the maximum dose, and MAE is calculated for only the high dose region (<90%) .

	Gamma pass rate				Mean gamma				SSIM				ME				MAE (high dose)			
	Per Beam		Per Plan		Per Beam		Per Plan		Per Beam		Per Plan		Per Beam		Per Plan		Per Beam		Per Plan	
	DL	PB	DL	PB	DL	PB	DL	PB	DL	PB	DL	PB	DL	PB	DL	PB	DL	PB	DL	PB
Mean	96.7%	94.5%	93.8%	92.4%	0.31	0.42	0.36	0.46	0.99	0.97	0.99	0.96	-1.4E-03	-4.7E-03	-4.5E-03	-7.7E-03	8.64E-3	1.5E-02	1.6E-02	2.8E-02
St. Dev.	5.7%	5.5%	7.6%	5.6%	0.11	0.09	0.13	0.10	0.01	0.00	0.01	0.00	4.5E-03	3.6E-03	5.4E-03	4.7E-03	2.0E-03	5.0E-03	4.8E-03	1.0E-02
Worst Case	71.6%	73.5%	71.2%	73.6%	0.74	0.72	0.24	0.36	0.98	0.96	0.96	0.95	-1.6E-02	-1.3E-02	-1.9E-02	-1.9E-02	1.3E-02	3.0E-02	2.6E-02	6.0E-02
Best Case	99.8%	99.3%	99.3%	98.3%	0.20	0.33	0.73	0.78	1.00	0.98	1.00	0.97	1.1E-04	1.3E-04	1.6E-04	4.9E-04	5.8E-03	8.3E-03	1.0E-02	1.4E-02
P-value	0.04		0.23		<0.0001		<0.0001		<0.0001		<0.0001		0.0004		0.04		<0.0001		<0.0001	

CONCLUSIONS

- The proposed network was able to learn the differences between PB-calculated dose distributions calculated with a 3x3x3 mm^3 dose grid and MC-calculated doses calculated with a 1x1x1 mm^3 dose grid
- The network generated highly accurate dose distributions, with an average gamma passing rate of 96.7% per beam, with a tight gamma criteria of 1% dose difference and 1 mm distance-to-agreement
- SSIM was 0.99 on average, indicating the shape and intensity of the dose distributions matched well
- Instead of traditional data augmentation techniques which may involve purely numerical operations, we augmented this dataset by re-calculating dose on a shifted or rotated dataset. We found this, in addition to training the network in a 3D fashion, greatly enhanced accuracy
- Future work will extend this framework to more geometrically varying treatment sites, such as head and neck

REFERENCES

1 Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med. Phys. 2003 Sep;30(9):2455-64.*

2 Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Imag. Proc.. 2004 Apr 13;13(4):600-12.*

3 Peng J, et al. Implementation of the structural SIMilarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection. *Med. Phys. 2020 Apr;47(4):1907-19.*

ACKNOWLEDGEMENTS

This research is supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA215718, Varian AI Research grant and Emory Winship Pilot Grant.

CONTACT INFORMATION

jmharms@emory.edu

*Senior author: xiaofeng.yang@emory.edu