Deep-learning powered denoising of Monte Carlo dose distributions

Hannes A Loebner[1], Raphael Joost[1], Jenny Bertholet[1], Stavroula Mougiakakou[2], Michael K Fix[1], and Peter Manser[1]

[1]Division of Medical Radiation Physics and Department of Radiation Oncology, Inselspital, Bern University Hospital, and University of Bern, 3010 Bern, Switzerland
[2]ARTORG Center, University of Bern, Murtenstrasse 50, 3008 Bern, Switzerland

Abstract This work demonstrates the development and application of fast deep-learning models designed for the mitigation of noise in Monte Carlo dose distributions (MC-DDs) with high statistical uncertainty (SU) of radiotherapy treatment plans. Five 3D U-net models were developed, varying in input, input size and batch normalization. The models were trained on pairs of high/low SU MC-DD pairs of randomly generated clinically motivated treatment plans calculated on open source available computed tomography (CT) scans. Depending on the model, the CT was included as input. The accuracy of the models is evaluated on the test set using gamma passing rate (2% global, 2 mm, 10% threshold) comparing denoised and low SU MC-DD. The best performing model included the CT input and no batch normalization. It was retrained for detailed evaluation on inhouse data, consisting of high/low-SU MC-DD pairs generated from 106 clinically-motivated volumetric modulated arc therapy (VMAT) arcs for 29 CT scans. The dataset was augmented to encompass a total of 3074 pairs. On the test set, the denoised MC-DDs agree with low-SU MC-DDs, by an average (standard deviation) gamma passing rate of 82.9% (4.7%). Applied to 12 previously unobserved clinically-motivated cases originating from different treatment sites yields an average gamma passing rate of 91.0%. Denoised DDs are computed, on average, in 35.1 seconds, a 340-fold efficiency gain compared to the calculation of low-SU MC-DDs.

1 Introduction

Monte Carlo (MC) dose calculation is widely acknowledged as the gold-standard for radiation therapy dose calculation, mostly due to its ability to accurately simulate the particle transport in heterogeneous media [1]. However, the necessity for a substantial number of particle histories in MC simulations to achieve reasonable statistical uncertainty (SU) leads to prolonged calculation times [2] and poses challenges for clinical implementation. Current clinical practice often resorts to analytical or simplified MC dose calculation algorithms, balancing accuracy with calculation times typically in the range of several minutes [3].

Deep learning (DL) presents a promising avenue to circumvent long calculation times while maintaining promising accuracy. There exists several approaches, incorporating U-nets [4, 5] and general adversarial networks [6–8] to perform dose predictions, relying only on CT and/or structure set as input, for specific treatment techniques. However, these dose predictions approaches have inherent drawbacks. They may produce results substantially deviating from the ground truth, introducing errors such as predicting doses far away from the tumor [9]. Additionally, their employment is often constrained by training dataset specificity, rendering them dependent on particular cancer types, treatment sites, or techniques [5].

To address these limitations, a promising approach is to give an "informed" input: This involves first the computationally cheap and fast calculation of MC dose distributions (MC-DDs) with high SU using only a limited number of particle histories. This informed input potentially reduces the risk of introducing errors such as mentioned before. In a second step, these DDs are denoised using DL to infer DDs similar to MC-DD with low SU [10, 11]. This approach is expected of being applicable to different treatment sites and techniques, as the primary task is denoising instead of full dose prediction. Moreover, using this approach, the problem of limited training data can be adressed, as random MC-DDs with low/high SU can be generated if needed. In our research group, the Swiss Monte Carlo Plan (SMCP) [12] has been developed and presents a flexible framework to calculate MC-DDs of various treatment techniques and is able to generate the necessary MC-DDs for training such DL models.

The objective of this study is therefore to develop and test different flavors of these denoising models and to apply the best performing model to 12 unseen clinically motivated cases.

2 Materials and Methods

In this work five different DL-powered denoising models are developed and tested (table 1). The main architecture consists of a standard 3D U-net [13] with four layers, integrating skip connections between encoding and decoding layers. Convolutional kernel sizes are 3*3*3 voxels, and max-pooling dimensions are 3*3*1 for the first and 2*2*2 voxels for the subsequent layers. The input dimension is 192*192 voxels in x*y direction and varying size in z direction. Geometries exceeding 192 voxels in x- and y-direction are cropped centrally and geometries exceeding 64 voxels in z-direction are handled using a patched-based approach [14]. Resulting overlapping voxels are averaged using TorchIO [15]. Geometries with less voxels are padded with zero voxels to fit the model input size. The model has 1-2 input channels, depending whether the CT is given as an additional input to the MC-DD with high SU or not.

Model	z-size	CT	Batch normalization
1	32	yes	no
2	64	yes	no
3	96	yes	no
4	64	no	no
5	64	yes	yes

Table 1: 3D U-net model flavours 1-5. The models differ by with varying input, input size and batch normalization.

During training, random patches with size 192*192*z are selected from DDs (and CTs, if specified), with additional data augmentation involving random mirroring and/or rotations in multiples of 90° in the xy-plane. The training employs a summed-squared-error loss function and a batch size of two. The ADAM optimizer [16] is used, starting with an initial learning rate of 10^{-4}. The learning rate decreases by a factor of 5, if the validation loss remains unchanged for four consecutive epochs. Training concludes when there is no improvement in the loss function over the last 30 epochs. The model is trained on a GeForce RTX 3090 GPU.

The dataset for training, validation and testing consisted of 5364 randomly generated volumetric modulated arc treatment (VMAT) plans: random multi-leaf collimator shapes are generated for a full gantry rotation, the collimator angle is randomly assigned and the monitor unit weights for each control point are assigned by randomly sampling numbers between 0 and 1 and sorting them in ascending order. Additionally, a random isocenter (within +- 5 cm in x, y and z of the center of the CT) is assigned. The beam energy is set to 6 MV. Always 12 plans were applied to 447 open source clinically motivated CTs to obtain a total of 5364 high/low SU MC-DD pairs. The high/low SU MC-DDs were calculated using SMCP [11], with $1.5*10^6$ particle histories (high SU >60%) and $3*10^8$ particle histories (low SU <2%), respectively. The dataset is divided into training, validation, and test set with 80%/10%/10% split.

The model accuracy is evaluated by comparing the low-SU MC-DD (reference) with the denoised DD by means of root-mean-squared error (RMSE) over all voxels and Gamma passing rate on the test set. Gamma evaluation with 2%/2 mm [17] and 3%/3 mm [18], 10% threshold is denoted as gamma-2 and gamma-3, respectively.

The best performing model out of those considered was re-trained on 106 clinically motivated VMAT applied to 29 CTs from our own data base. The plans were augmented by randomly changing the beam energy (6, 10, 15 MV), the isocenter (within +-5 cm in x, y, z direction), and the collimator angle, to a total of 3074 high/low SU MC-DD pairs. Training was completed within 28 h 19 min (149 epochs).

The re-trained model is applied to 12 previously unseen clinically motivated cases, including 4 head and neck, 4 non-small-cell lung, and 4 prostate cases. RMSE, gamma-2, and gamma-3 are evaluated. The noise reduction from high SU MC-DD to denoised DD, is quantified using the Improved Signal-to-Noise Ratio (ISNR) [11].

The computation time to calculate a high-SU MC-DD, the subsequent preprocessing, denoising and storing is recorded and compared to the calculation time of the low-SU MC-DD.

3 Results

The accuracy of the five models is shown in table 2. Including the CT information improved the model accuracy. Moreover, batch normalization had a negative impact on the model accuracy.

Model	Average gamma-2 [%]	Average gamma-3 [%]	RMSE [10^{-4} Gy/MU]
1	89.16±3.99	97.30±1.69	1.21±0.22
2	**89.74±4.00**	**97.50±1.66**	**1.16±0.20**
3	89.72±3.92	97.48±1.61	1.15±0.20
4	89.23±4.16	97.30±1.78	1.16±0.20
5	83.54±4.77	94.50±2.55	1.26±0.27

Table 2: Model accuracy including standard variation. The best performing model is marked in bold.

The model 2 had a mean (standard deviation) calculation time of 32.7 s (7.5 s) to obtain a high-SU MC-DD on a single CPU. Subsequent data preprocessing, denoising, and storage were completed on average (standard deviation) within 2.6 s (0.3 s). This contrasts an average (standard deviation) calculation time of 3 hours and 19 minutes (42 minutes) for low-SU MC dose calculation, including storage, on one CPU: an efficiency gain of factor 340. The application of the retrained model to 12 unseen cases is demonstrated for one head and neck and one lung case in figure 1.

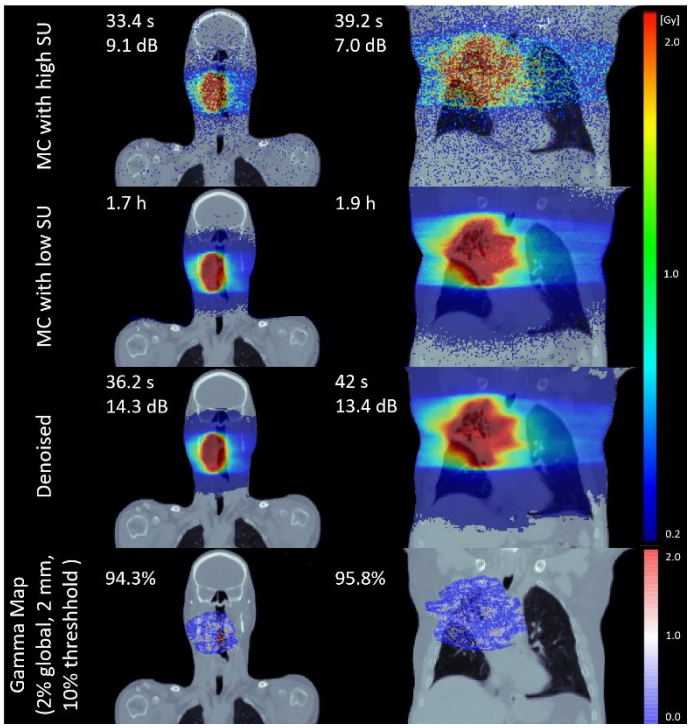

Figure 1: High SU, low SU and denoised DDs are visualized with 10% dose threshold for a head and neck (left) and a lung (right) case. Their computation times and signal to noise ratio as compared to the MC-DD with low SU are shown on the respective upper left corner. The gamma evaluation (low SU DD as reference and denoised DD as evaluation) is shown on the bottom. In the gamma map, voxels failing the gamma criteria (> 1.0) are shown in red tones with the gamma passing rate shown in the top-left corner of the gamma map.

For the 12 cases, the average [range] RMSE is $1.1*10^{-4}$ Gy/MU [$0.5*10^{-4}$, $1.7*10^{-4}$]. The average [range] gamma-3 and gamma-2 is 97.5% [93.4, 98.7] and 91.0% [79.6, 95.9], respectively. The average [range] ISNR is 9.9 dB [4.2, 15.2].

4 Discussion

In this work, we have developed and tested DL-powered models to denoise MC-DDs.

Generally, the differences in accuracy between the models are small: the different z input sizes had little impact on the gamma-2, gamma-3 and RMSE. However, the patched borders could be identified visually due to “steps” in the dose distribution. Keeping in mind, that the GPU has a limited memory [19], the results suggest choosing a z-size conforming to the availability and specifications of the available GPU model. Similarly, no substantial improvement in accuracy due to the additional CT input was observed. However, in the context of explainable results and minimizing the risk of false dose predictions, the CT input is potentially beneficial. Including batch normalization substantially degraded the accuracy of the denoising. Although batch normalization is used in many segmentation tasks and similar U-nets [20], it was not beneficial for the denoising task. This might be due to the fact that batch normalization tends to lose the information of the absolute dose values, which is critical for predicting accurate dose distributions. This effect can be magnified when using patching, as the information of the previous patch is not available for the next one.

Denoised DDs generated with the best performing model considered exhibit an average RMSE of $1.1*10^{-4}$ Gy/MU when compared to low-SU MC-DDs, aligning with findings in literature using similar methodologies [10, 19]. For a representative VMAT arc with 300 MU and a prescription of 50 Gy, delivered in 2 Gy fractions, this leads to an average RMSE of 0.9 Gy for the whole treatment. Although these results are quite promising, they do not conform to the dose calculation accuracy levels recommended by AAPM [21].

For the 12 unseen clinically motivated cases slightly lower gamma passing rates are observed when compared to previous works using DL-powered denoising or dose prediction methods [10, 18]. This can be partly explained by the slightly different task of denoising the DD of entire VMAT arcs rather than beams from single directions, resulting in a 9.9 dB average improvement in SNR, a factor 2 smaller than in other studies [11]. Moreover, the other studies used less/more particle histories for the high/low SU MC-DDs.

The denoising computation time is comparable to the findings in literature [10]. The entire workflow, encompassing the generation and denoising of high-SU MC-DDs, is completed within less than a minute, surpassing the speed of the majority of clinically applied dose calculation algorithms [3, 22].

5 Conclusion

We developed and tested DL-powered denoising model for MC-DD. The most accurate model showed an impressive efficiency gain (factor 340), substantial noise reduction and promising accuracy.

Acknowledgements

This work was supported by Varian, a Siemens Healthineers Company. Dose calculations were performed on UBELIX (www.id.unibe.ch/hpc), the high-performance cluster at the University of Bern.

References

[1] Fauzi A, Zahra F, Miftahul Munir M, Krieger T, Sauer OA. Monte Carlo- versus pencil-beam-/collapsed-cone-dose calculation in a heterogeneous multi-layer phantom. Phys Med Biol 2005;50:859. https://doi.org/10.1088/0031-9155/50/5/010.

[2] Chetty IJ, Curran B, Cygler JE, DeMarco JJ, Ezzell G, Faddegon BA, et al. Report of the AAPM Task Group No. 105: Issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning. Med Phys 2007;34:4818–53. https://doi.org/10.1118/1.2795842.

[3] Das IJ, Moskvin V, Johnstone PA. Analysis of Treatment Planning Time Among Systems and Planners for Intensity-Modulated Radiation Therapy. Journal of the American College of Radiology 2009;6. https://doi.org/10.1016/j.jacr.2008.12.013.

[4] Dong P, Xing L. DoseNet: A Deep Neural Network for Accurate Dosimetric Transformation between Different Spatial Resolutions and/or Different Dose Calculation Algorithms for Precision Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics 2019;105:S255–6. https://doi.org/10.1016/J.IJROBP.2019.06.2471.

[5] Gronberg MP, Gay SS, Netherton TJ, Rhee DJ, Court LE, Cardenas CE. Technical Note: Dose prediction for head and neck radiotherapy using a three-dimensional dense dilated U-net architecture. Med Phys 2021;48:5567–73. https://doi.org/10.1002/MP.14827.

[6] Kearney V, Chan JW, Wang T, Perry A, Descovich M, Morin O, et al. DoseGAN: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation. Sci Rep 2020;10. https://doi.org/10.1038/s41598-020-68062-7.

[7] Babier A, Mahmood R, McNiven AL, Diamant A, Chan TCY. Knowledge-based automated planning with three-dimensional generative adversarial networks. Med Phys 2020;47. https://doi.org/10.1002/mp.13896.

[8] Zhan B, Xiao J, Cao C, Peng X, Zu C, Zhou J, et al. Multi-constraint generative adversarial network for dose prediction in radiotherapy. Med Image Anal 2022;77:102339. https://doi.org/10.1016/J.MEDIA.2021.102339.

[9] Teng L, Wang B, Xu X, et al.. Beam-wise dose composition learning for head and neck cancer dose prediction in radiotherapy (2024). Medical Image Analysis. https://doi.org/10.1016/j.media.2023.103045.

[10] Fornander H. Denoising Monte Carlo Dose Calculations Using a Deep Neural Network (2019). https://WwwDiva-PortalOrg/Smash/Get/Diva2:1366439/FULLTEXT01Pdf 2019 (accessed August 16, 2023).

[11] Javaid U, Souris K, Dasnoy D, Huang S, Lee JA. Mitigating inherent noise in Monte Carlo dose distributions using dilated U-Net. Med Phys 2019;46:5790–8. https://doi.org/10.1002/MP.13856.

[12] Fix MK, Manser P, Frei D, Volken W, Mini R, Born EJ. An efficient framework for photon Monte Carlo treatment planning*. Phys Med Biol 2007;52:N425. https://doi.org/10.1088/0031-9155/52/19/N01.

[13] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[14] Ghimire K, Chen Q, Feng X. Patch-Based 3D UNet for Head and Neck Tumor Segmentation with an Ensemble of Conventional and Dilated Convolutions. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12603 LNCS, 2021. https://doi.org/10.1007/978-3-030-67194-5_9.

[15] Pérez-García F, Sparks R, Ourselin S. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput Methods Programs Biomed 2021;208. https://doi.org/10.1016/j.cmpb.2021.106236.

[16] Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings 2014.

[17] Venselaar J, Welleweerd H, Mijnheer B. Tolerances for the accuracy of photon beam dose calculations of treatment planning systems. Radiotherapy and Oncology 2001;60. https://doi.org/10.1016/S0167-8140(01)00377-2.

[18] Tsekas G, Bol GH, Raaymakers BW. Robust deep learning-based forward dose calculations for VMAT on the 1.5T MR-linac. Phys Med Biol 2022;67:225020. https://doi.org/10.1088/1361-6560/AC97D8.

[19] Wang L, Ye J, Zhao Y, et al. SuperNeurons: Dynamic GPU Memory Management for Training Deep Neural Networks. Arxive 2018. https://doi.org/10.48550/arXiv.1801.04380

[20] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[21] Smilowitz, J.B., Das, I.J., Feygelman, et al. (2015), AAPM Medical Physics Practice Guideline 5.a.: Commissioning and QA of Treatment Planning Dose Calculations — Megavoltage Photon and Electron Beams. Journal of Applied Clinical Medical Physics, 16: 14-34. https://doi.org/10.1120/jacmp.v16i5.5768

[22] Failla GA, Wareing T, Archambault Y, Thompson S. Acuros® XB Advanced Dose Calculation for the EclipseTM Treatment Planning System 2015. https://varian.widen.net/s/n2hmf2pshr/acuros_clinicalperspectives_rad10156a_april2015 (accessed August 16, 2023).