ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Fast denoising of low-count Monte Carlo proton therapy dose distributions with ResUNet

Pierre Merveille[1], Ana M. Barragán-Montero[2],
Kevin Souris[3], and John A. Lee[12] *

1- Université catholique de Louvain - ICTEAM/ELEN
Place du Levant 3 L5.03.02, 1348 Louvain-la-Neuve - Belgium

2- Université catholique de Louvain - IREC/MIRO
Avenue Hippocrate 55 B1.54.07, 1200 Brussels - Belgium

3- Ion Beam Applications SA
Chemin du Cyclotron 3, 1348 Louvain-la-Neuve - Belgium

Abstract. Monte Carlo (MC) dose calculation is the gold standard for proton therapy but remains computationally expensive due to statistical noise at low-count particle histories. We propose a deep learning approach based on a ResUNet architecture to denoise high uncertainty (5%) MC dose distributions and approximate low uncertainty (0.5%) reference doses. Using the MCsquare engine, dose pairs were generated for 150 patients from head-and-neck, prostate, and lung datasets. The proposed model achieved high structural similarity and dosimetric accuracy while reducing computation time by a factor of 30, enabling fast and accurate MC dose estimation for clinical proton therapy.

1 Introduction

Monte Carlo (MC) simulations are widely regarded as the gold standard for dose calculation in proton therapy because they can accurately model complex particle interactions and heterogeneous patient geometries. However, MC remains limited by the high computational cost required to achieve statistically stable dose distributions. The stochastic nature of MC methods introduces noise that scales inversely with the square root of the number of simulated particle histories, meaning that low-uncertainty doses (e.g., $< 1\%$) require long computation times as the time scales linearly with N. This trade-off between accuracy and runtime is a critical bottleneck for adaptive or online proton therapy workflows [1], where rapid but reliable dose estimation is essential.

Recent work has explored deep learning to accelerate proton dose computation. Direct-prediction approaches such as Dose Transformer algorithm (DoTA) [2] achieve very fast running times, but may struggle to generalize to new beam configurations or anatomical changes. This paper proposes a deep learning framework to denoise proton dose distributions generated with the MCsquare MC engine [3]. Previous MC-denoising approaches [4] use 2D or region-specific models and/or the number of simulated particles as a noise descriptor. In contrast, our method processes full 3D dose volumes and is trained directly on the

---

[*] J.A.Lee is a Research Director with the Belgian F.R.S.-FNRS. Ana M. Barragán-Montero is a Research Associate with the FNRS.

statistical uncertainty reported by the MC simulation. The 3D ResUNet learns to map high statistical uncertainty (5%) doses to low statistical uncertainty (0.5%) reference doses, effectively reducing the required particle histories while preserving MC accuracy.

The rest of this paper is organized as follows. Section 2 details the proposed methodology used to tackle the problem. Results are reported and discussed in Section 4. Section 5 concludes and sketches some perspectives.

2 Materials and Method

2.1 Data Generation

Paired dose distributions with different levels of statistical uncertainty were generated with MCsquare [3]. For each patient, two independent simulations were computed: a noisy dose with 5% statistical uncertainty and a nearly noiseless dose with 0.5% statistical uncertainty, serving as the reference. Each pair of doses has been computed on a common voxel grid (2x2x2mm).

Three datasets were used, totaling 150 patients and covering distinct tumour sites. The head-and-neck dataset (60 patients) was provided by the MIRO laboratory and included clinical proton plans. The prostate dataset (60 patients) was open-source [5], with plans created using two opposed beams at 90° and 270° gantry angles targeting the Gross Tumour Volume (GTV). The lung dataset (30 patients) originated from the open-source NSCLC-Radiomics collection [6], for which treatments were planned according to each individual patient anatomy. For both open-source datasets, the created plans were optimised using the beamlet-free method [7] to ensure tumour coverage and compliance with dose constraints on the organs at risk (OARs) in the regions of interest (ROIs). These three datasets were divided into training, validation, and test sets in proportions of 80%, 10%, and 10%, respectively.

2.2 Data Preprocessing

Each pair of noisy (5%) and reference (0.5%) dose distributions was zero-padded and divided into patches of 128^3 voxels, since the initial image definition differs between patients. Dose values were cropped between 0 and 100 Gy, which covers the clinically relevant range (average prescription of 70 Gy) and suppresses high intensity outliers.

Because large portions of the volume contain air or low-dose regions, patches with $< 1\%$ non-zero voxels were kept with a probability of 0.3, preserving the natural class imbalance while avoiding an excess of empty patches.

An Anscombe variance-stabilizing transform was then applied to approximate Gaussian noise before training.

2.3 Network Architecture

The denoising model is a 3D Residual U-Net [8, 9] with five levels, featuring channel dimensions $[32, 64, 128, 256, 512]$. Each encoder level contains a Res-

Block with two 3x3x3 convolutions, PReLU activations, GroupNorm, and residual connections. The spatial dimensions are halved by the first convolution of each ResBlock. The decoder mirrors the encoder with transposed convolutions and skip connections to the corresponding encoder features. The network outputs a single-channel dose map via a 1x1x1 convolution. The model predicts the noise contained in the noisy dose. The model was trained by reducing the mean square error (MSE) between the reference image and the result of subtracting the predicted noise from the noisy dose.

The model was trained for 250 epochs on the training set using the Adam optimizer with an initial learning rate of 0.01, which decays by a factor of 2 every 7 epochs without improvement. The best model was selected according to validation performance. Training was performed on a Tesla V100 GPU 16Gb.

2.4 Evaluation Metrics

The trained network was evaluated on unseen test patients using both voxelwise and dosimetric criteria. The mean squared error (MSE) and mean absolute error (MAE), computed in gray (Gy), quantify absolute dose deviations between the predicted and reference distributions. In addition, the peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM) were used to assess perceptual similarity and image fidelity.

From a dosimetric perspective, dose-volume histograms (DVHs) were computed (dose CDF integrated from right to left in a given organ or volume) for the target volume and OARs. For the target volume, we extracted D95 (dose received by 95% of the target) and D5 (dose received by 5% of the target). For the OARs, we evaluated Dmean and D2, which respectively capture the overall organ exposure and near-maximum dose to sensitive subvolumes. These metrics provide a clinically meaningful evaluation of the dose after denoising.

Finally, the time taken to denoise a dose volume was compared with the durations of the corresponding MC simulations at 5% and 0.5% statistical uncertainty levels to quantify the computational gain.

2.5 Benchmark

The proposed ResUNet was tested on 15 patients: 6 head-and-neck cases, 6 prostate cases, and 3 lung cases. The model has been compared with BM4D (an adaptation of BM3D [10] for 3D volumes). The BM3D technique is a state-of-the-art denoising algorithm based on collaborative filtering in the transform domain, where similar image patches are grouped and jointly processed using a 3D transform to suppress noise while preserving structural details.

2.6 Workflow Integration

The proposed denoising model was used into the dose computation pipeline. After an initial with high statistical uncertainty (5%), the resulting dose volume was automatically divided into overlapping patches. Each patch was passed through the trained network and the denoised outputs were then cropped at their centers and reassembled into a full 3D dose volume. The final dose map was obtained by averaging the voxel values according to the number of times each voxel was included in overlapping patches. This process produced a denoised dose distribution that approximates the quality of a simulation with low statistical uncertainty (0.5%).

3 Results

Table 1 compares the voxel-wise and perceptual metrics on the test set among the original noisy dose, BM4D, and the proposed ResUNet model, which outperformed BM4D in all metrics. ResUNet reduced the MAE by a factor of 10 and the MSE by a factor of 80 compared to the noisy dose. It achieved a PSNR of approximately 60 dB and an SSIM above 0.99. These results demonstrate a strong voxel-wise correspondence and excellent preservation of fine spatial structures between the denoised and reference doses.

Method	MSE [Gy^2]	MAE [Gy]	PSNR [dB]	SSIM
Noisy dose	2.937	0.528	39.08	0.983
BM4D	0.103	0.128	53.66	0.995
ResUNet	**0.035**	**0.053**	**60.12**	**0.998**

Table 1: Quantitative comparison of denoising methods using MSE, MAE, PSNR, and SSIM on the test set.

Figure 1 illustrates the denoising performance of the model. Most of the errors are located within the irradiated regions, indicating that the model effectively targets noise in the dose distribution rather than in non-irradiated areas.

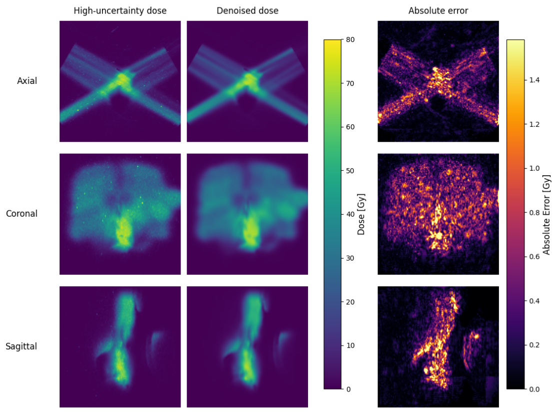

Fig. 1: Illustrative example for one head-and-neck test patient showing high-uncertainty dose, denoised dose, and absolute-error maps across axial, coronal, and sagittal slices.

Table 2 reports DVH metrics. The values of D_{95} and D_5 were extracted from the DVHs for the target and D_{mean} and D_2 have been averaged for OARs of each patient. For patient i and volume j, the absolute difference $\Delta D_{i,j} = |D_{i,j}^{GT} - D_{i,j}^{evaluated}|$ and the relative difference $\delta D_{i,j} = 100\% \times |D_{i,j}^{GT} - D_{i,j}^{evaluated}|/D_{i,j}^{GT}$ have been computed. The relative differences of the ResUNet were below 0.5% for the target and OARs. These differences are within clinically acceptable limits, confirming that the denoising process does not distort the overall dose coverage.

Method	D_{95} Target		D_5 Target		D_{Mean} OARs		D_2 OARs	
	Abs [Gy]	Rel [%]	Abs [Gy]	Rel [%]	Abs [Gy]	Rel [%]	Abs [Gy]	Rel [%]
Noisy	3.191	4.944	3.576	5.069	**0.032**	**0.419**	**0.063**	**0.100**
BM4D	0.227	0.392	0.386	0.556	0.049	0.626	0.259	0.427
ResUNet	**0.215**	**0.364**	**0.257**	**0.368**	0.071	0.908	0.283	0.428

Table 2: Absolute and relative differences in D_{95} and D_5 for the target and in D_{Mean} and D_2 OARs, compared with the 0.5% reference MC dose.

Figure 2 shows that the model reduces noise while closely approximating the low-uncertainty dose distributions across target volumes and pre-selected OARs.

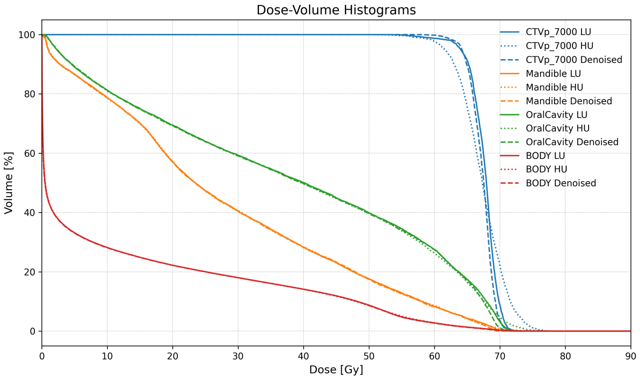

Fig. 2: Dose-volume histograms (DVHs) for representative structures of a head and neck test patient. Solid lines: Low-uncertainty dose (0.5% uncertainty, LU), dotted lines: High-uncertainty dose (5% uncertainty, HU) and dashed lines: denoised dose predicted by the proposed ResUNet model.

The average computation time for each stage of the workflow was computed on a CPU (AMD Ryzen 7 5800X). At 5% statistical uncertainty, a MCsquare simulation required approximately 65 seconds, whereas the 0.5% reference simulation took 2410 seconds. For denoising, BM4D lasted for 300 seconds per patient while ResUNet took only 41 seconds. When denoising with a Tesla V100 GPU 16Gb, execution lasts less than a second resulting in a total effective running time hardly longer than the fast simulation while producing dose distributions of reference quality.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

4 Conclusion

This work demonstrates the feasibility of using deep learning to accelerate Monte Carlo proton dose calculation. By training a ResUNet model to denoise high-uncertainty (5%) MCsquare simulations into low-uncertainty (0.5%) counterparts, we achieved accurate dose reconstruction across multiple tumour sites while dividing computation time by 30. The resulting doses preserved both voxelwise accuracy and clinically relevant DVH metrics for targets and organs at risk. The proposed approach can be seamlessly integrated into existing MC dose computation workflows, offering a practical solution for rapid, high-quality dose computation. Future work will investigate uncertainty quantification strategies to evaluate the confidence in the denoised dose predictions. We will also explore the limits of the model under more extreme noise levels, including higher statistical uncertainty regimes (e.g., 10% or beyond) to assess robustness outside the current 5% training domain.

References

[1] Harald Paganetti, Pablo Botas, Gregory C Sharp, and Brian Winey. Adaptive proton therapy. *Physics in Medicine Biology*, 66(22):22TR01, nov 2021.

[2] O. Pastor-Serrano and Z. Perkó. Millisecond speed deep learning based proton dose calculation with monte carlo accuracy. *Physics in Medicine and Biology*, 67(10):105006, 2022.

[3] Kevin Souris, John Aldo Lee, and Edmond Sterpin. Fast multipurpose monte carlo simulation for proton therapy using multi- and many-core cpu architectures. *Medical Physics*, 43(4):1700–1712, 2016.

[4] Umair Javaid, Kevin Souris, Sheng Huang, and John A. Lee. Denoising proton therapy monte carlo dose distributions in multiple tumor sites: A comparative neural networks architecture study. *Physica Medica*, 89:93–103, 2021.

[5] Richard F. Thompson, Arjun Kanwar, Benjamin Merz, Evan Cohen, Hannah Fisher, Shaan Rana, Christopher Claunch, and Alexander Hung. Stress-testing pelvic autosegmentation algorithms using anatomical edge cases (prostate anatomical edge cases) (version 1). The Cancer Imaging Archive (TCIA), 2023. Dataset.

[6] Hugo J. W. L. Aerts, Leon Wee, Emmanuel Rios Velazquez, Ralph T. H. Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, Rene Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, Floris Hoebers, Maarten M. Rietbergen, C. Rene Leemans, Andre Dekker, John Quackenbush, Robert J. Gillies, and Philippe Lambin. Data from nsclc-radiomics (version 4). The Cancer Imaging Archive (TCIA), 2014. Dataset.

[7] Danah Pross, Sophie Wuyckens, Sylvain Deffet, Edmond Sterpin, John A. Lee, and Kevin Souris. Technical note: Beamlet-free optimization for monte-carlo-based treatment planning in proton therapy. *Medical Physics*, 51(1):485–493, 2024.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[9] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[10] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.