

A Hybrid Monte Carlo and Deep Learning Approach for Fast and Generic Photon Beam Dose Calculation

Maxime Rousselot¹, Chi-Hieu Pham¹, Dimitris Visvikis^{1*} and Julien Bert¹

¹ LaTIM, INSERM UMR1101, University of Brest, 29200 Brest, France

* corresponding author

E-mail: dimitris.visvikis@univ-brest.fr

Abstract. Dosimetry is an essential tool to provide the best and safest radiotherapies to a patient. In this field, Monte-Carlo simulations are considered to be the golden standard for predicting accurately the deposited dose in the body. Such methods are very time-consuming and simpler dose calculation engines, like dose kernel approaches, were created for cases where a fast estimation is necessary. In those approaches, dose distribution maps (or dose kernel) are learned for simple beams geometry, then, a combination of numerous kernels is used to simulate more complex beams. However, those methods often lack personalization and oversimplify the human body, especially for the secondary interactions dose deposition. In this article, we explore the possibility of learning the dose kernel using convolutional neural networks to improve their accuracy towards each different human body. We also highlight the limits of such approaches.

Keywords: Monte-Carlo, Dosimetry, Deep learning, Dose kernel

1. Introduction

Modeling the absorbed dose delivered to a patient during an X-ray medical imaging procedure is an essential tool for optimizing, monitoring and reducing exposure to ionizing radiation. If in conventional radiology, the DAP (Dose-Area Product) is used to calculate the dose to the skin, it is important for certain applications to estimate the specific dose to the patient and at the organ level. Within this context Monte-Carlo simulations (MCS) are considered to be the golden standard to estimate this deposited dose. Especially when the absorbed dose need to be estimated in 3D within the organs. They are considered more accurate than any other traditional deterministic dose calculation engine such as pencil beam (PB) [Mohan et al., 1986] or collapsed cones(CC) [Ahnesjö, 1989], etc. MCS provide more precision around tissue heterogeneities and air cavities [Krieger and Sauer, 2005].

Indeed, they model the comportment of numerous particles, taking into account density variations within the patient with the help of CT images and simulate particle transportation with high precision and score energy depositions accordingly. However, to enable sufficient statistical accuracy, the trajectories and physical processes of millions of particles have to be calculated through the patient voxelized phantom. This requires a considerable amount of execution time. To reduce the computation burden, several approaches were proposed. For example, as MCS are highly parallelizable, GPU implementations offer sensible gain in terms of speed [Bert et al., 2013] [Tian et al., 2015]. Other methods, such variance reduction techniques (VRT), proposed to modify the calculation of particles histories to increase the efficiency of MCS without introducing approximation. For example, the sampling of the particle tracking can be optimized [Behlouli et al., 2018], or the photon dose deposition along its path can be condensed with the track length estimator (TLE) method [Williamson, 1987].

In recent years, the fast evolution of deep learning opens a new way to quickly and accurately predict a dose distribution in various applications. Many specialised neural network were train to predict the dose to optimize particle therapy, for example, for helical tomotherapy [Liu et al., 2019], intensity-modulated radiation therapy (IMRT) [Kontaxis et al., 2020], brachytherapy [Villa et al., 2021], proton therapy [Neishabouri et al., 2021]. Deep learning approach were also widely used for medical imaging such as positron emission tomography (PET) [Lee et al., 2019], single photon emission computed tomography (SPECT) [Götz et al., 2020], and for x-ray imaging, CT scan [Roser et al., 2019] and Cone Beam CT [Villa et al., 2023].

Concerning X-ray imaging, the main deep learning drawback of the approach is the lack of genericity of the solution. In most cases the network model is trained for a given site (pelvis, head, thorax) and for a beam geometry (cone beam, fan beam). Under these conditions it is difficult to parameterize the beam with different aperture angles or collimation. A solution proposed by [Villa et al., 2023], in x-ray imaging of the thorax, has consisted of introducing and conditional network with the following imaging parameters: beam position, beam angulation and X-ray tube voltage. However, dynamic collimation was not considered since the combinatorial issue of having multiple beam geometry. In IMRT application, where the beam is collimated, [Kontaxis et al., 2020] has proposed encoding the beam into an image, obtained by raytracing, as an input of the network. Netherless, such approach still necessitates being sufficiently large training data to cover all the possible variations of the parameters. The learning processes is more time-consuming, but also requiring a network with more parameters. Therefore, most of the time each trained network can only be used for one specific application and is not very adaptable.

In this article, we explore a more versatile approach that combines Monte Carlo sampling and deep learning. The main idea is to decompose the x-ray beam into elemental small beams name beamlet, as it is done in radiation therapy inverse planning to optimize the beam aperture [Unkelbach et al., 2015]. A beamlet is a mono-energetic

pencil beam that represents virtually a simple element that compose any X-ray beam. Then by using a sampling approach, like Monte Carlo, it is possible to randomly draw any beam shape based on beamlets: fan beam, cone beam, collimated beam, etc. Therefore, it will be possible to estimate the patient dose for any beam configuration with a reasonable network and training data set. Only the dose response of a beamlet for different energies and different anatomical sites of the patient need to be learned.

2. Materials and methods

2.1. Database creation

For the training data set a total of 80 CT patients from different anatomical site (head and neck, thorax and pelvis) was used. These data were collected and anonymized by the Brest University Hospital. All images were re-scaled to a voxel resolution of $3 \times 3 \times 3$ mm. A total of 17000 beamlets were generated with random position, orientation, energy and patient. For each of these beamlets a 3D CT image patch from the corresponding patient were extracted following the direction of the beamlet. Then the oriented patch was transformed into axis aligned image, in order to be interpretable by the network. Each patch corresponds to a small region of the patient CT where the beamlet is passing through, and will be used to recovered the absorbed dose. A patch-size of $27 \times 27 \times 112$ voxels was defined. Since the beamlet is a pencil beam, the patch size along the particle direction was longer. Those dimensions are a trade-off between optimal data size and the relevance of the deposited dose gradient considering a maximum photon energy of < 2 MeV. The absorbed dose within each patch was calculated using Monte Carlo simulations with the GATE software [Sarrut et al., 2014]. Each patch, that contains Hounsfield value, was converted in materials labels using density conversion given by [Schneider et al., 2000]. A total of 8×10^7 particles was run for each beamlet in order to reach for every voxel a dose statistical uncertainty less than 5%. A TLE variance reduction technique [Williamson, 1987] was used to speed up the simulation. Some examples of extracted patches and the resulting dose can be found in Figure 1.

2.2. Neural networks

2.2.1. Architecture For predicting the dose corresponding to each patch and each beamlet energy, a Convolutional Neural Network (CNN) was designed. The main idea was having an architecture with two inputs: the CT patch and the energy of the beamlet. The output of the network was the absorbed dose map. However, as emphasis in Figure 1, doses have two different areas of scale values. A first one, with high dose values, along the beamlet direction where primary interactions occurs. And a second one, with smaller dose values around the beamlet, and corresponding to the particles scattered within the patch. Between primary and scattered dose values a magnitude of two orders was measured. As in general CNN are less efficient to learn high frequencies,

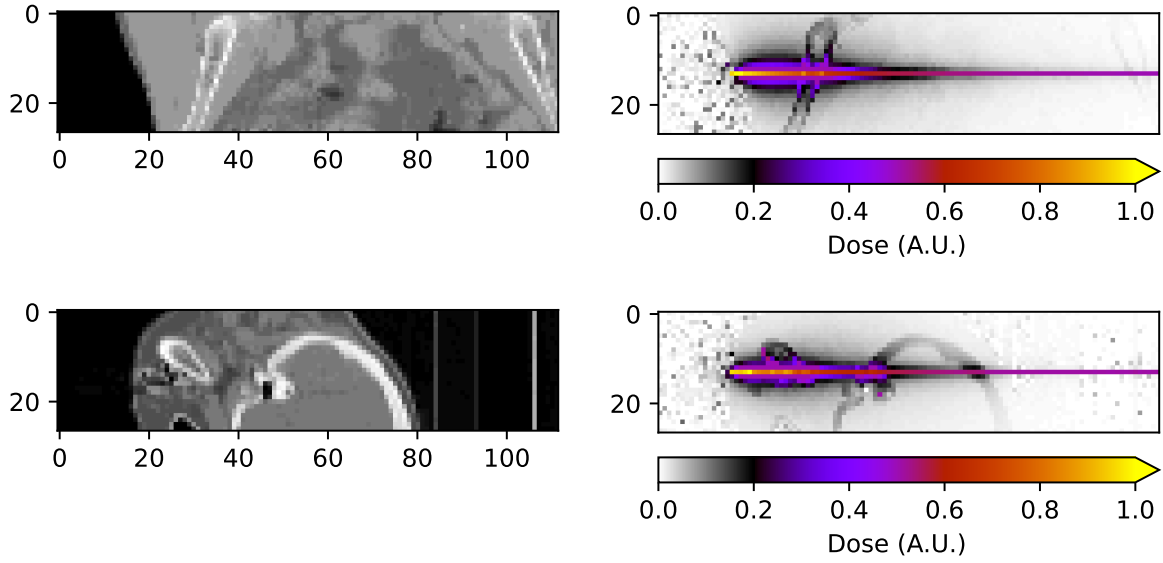


Figure 1. Examples of 3D patches for a photon beamlet of 89 keV (first row) and a photon beamlet of 428 keV (second row). The left column is the central slice of each CT image patch and the right column the corresponding absorbed dose map in normalized arbitrary unit estimated by Monte Carlo simulation.

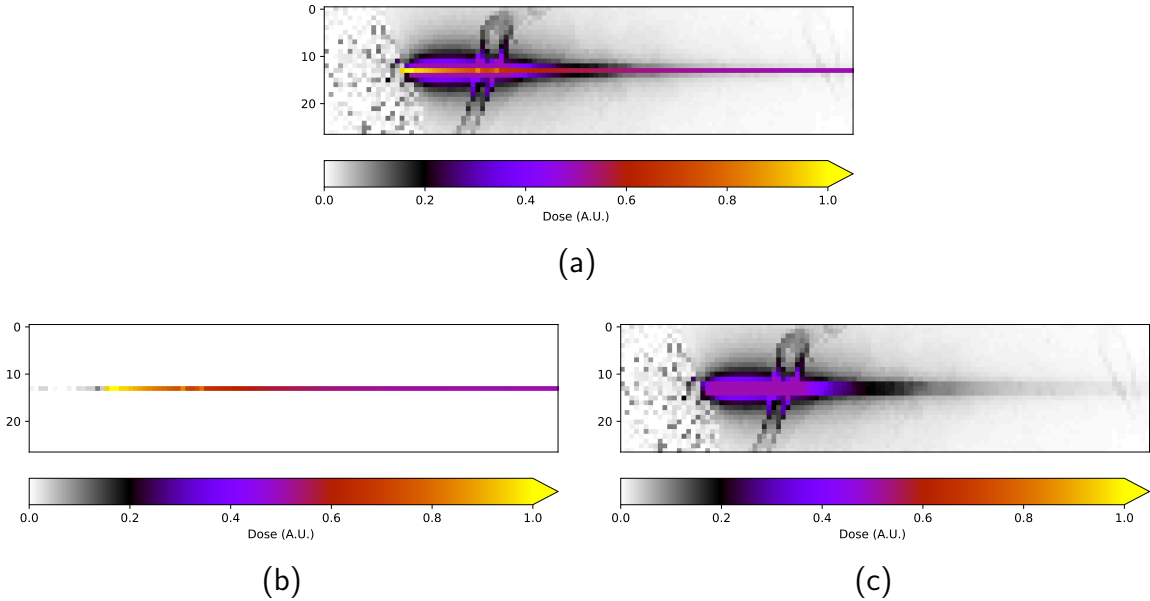


Figure 2. (a) Dose map from a beamlet Monte Carlo simulation, decomposed in (b) primary dose deposition and (c) scattered dose deposition.

we chose to split the dose map (Figure 2a) in two dose maps: the dose from the primary interaction (Figure 2b) and the dose from the scattering (Figure 2c).

Two sub-networks were then designed for predicting the dose for each beamlet and patch. Both were based on U-Net [Ronneberger et al., 2015] architecture, which was intended for biomedical image segmentation but has already proven to be efficient for

dosimetric application [Villa et al., 2023]. That kind of network is composed of two symmetrical steps. The first step is a contracting path and the second is an expanding path, which gives the network its U-shape architecture. The contracting path is a typical CNN composed of repeated convolution, Rectified Linear Unit activation functions (ReLU), and maximum pooling layers that reduce dimensionality while capturing relevant features. The expanding path of the network gradually up-samples the images while concatenating, at each step, the contracting part features of the same dimension. We include the energy input between the compression and expansion part of the network using a dense layer. The aim was conditioning the network considering the beamlet energy. The energy value was not injected with the image patch at the input of the network, because during the training the single energy value will be mostly discarded compared to the thousand of voxels that compose the patch image. The first network (Figure 3) aims to predict the dose of the primary interactions. Then, has it was a pencil beam, the input and output images, as well as all layers filter, are uni-dimensional. We used a typical dimension for the different layer filters: 3 for the convolution filters and 2 for the max pooling and upsampling layers.

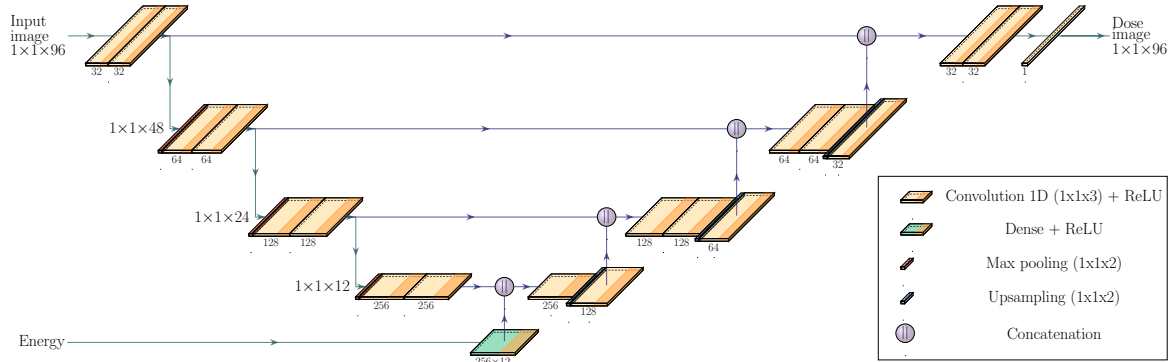


Figure 3. First 1D network that predict the dose of the beamlet primary interactions.

The second network that predict the 3D dose maps of the scattered particles is shown in Figure 4. Since the patch is in 3D, the convolution filters have also to be three-dimensional with a size of $3 \times 3 \times 3$. Moreover, to preserve the symmetry of the dose across the network, a maximum pooling and upsampling filter of size $3 \times 3 \times 2$ instead of the more conventional size $2 \times 2 \times 2$ was chosen.

2.2.2. Training and validation The dataset was divided into three sets, the training, the validation, and the test set, by randomly selecting, 13500, 1500, and 2000 patches, respectively. For the test set patches, different patients, which was not used for the training and the validation was used. The aim was to ensure that our network can generalize well on new data. To train the 3D CNN, a classical data augmentation technique was used. We applied three 90° rotations along the beamlet axis on the input and output images to quadruple the size of the training set. The dose values were

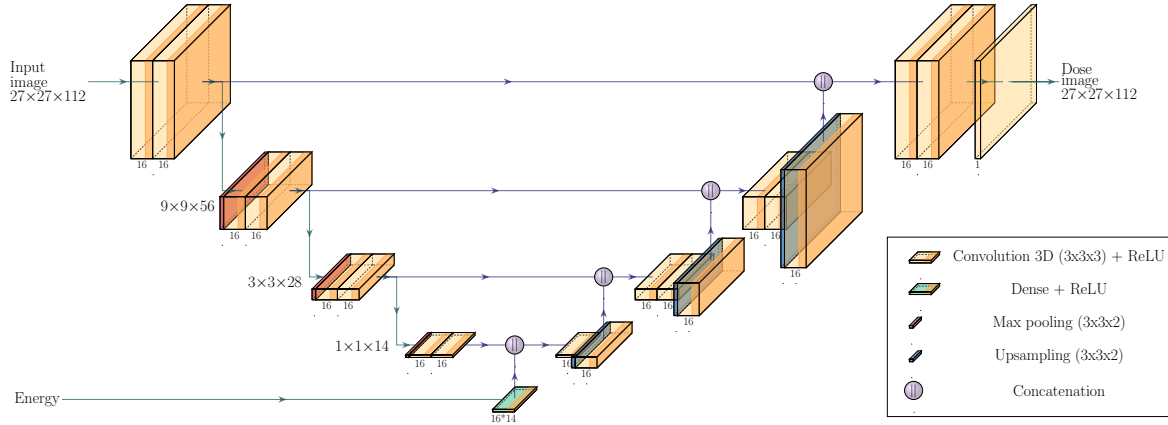


Figure 4. Second 3D network that predict the dose of the beamlet scattered interactions.

normalized between 0 and 1 to facilitate the training. In addition, the absorbed dose to voxel that contains air material (very small density) was set to zero avoiding too much attention to the network on very high dose value that are not relevant for patient dosimetry.

To train the two networks, the adaptive momentum algorithm (ADAM) [Kingma and Ba, 2014] was used to minimize the mean absolute error between the predicted dose and the MCS doses. During the training, both the validation and training loss at each epoch were monitored. The process was stopped when the validation loss ceased to decrease compared to the training loss. The two networks were implemented using Keras and Tensorflow on an Nvidia GeForce RTX 1080Ti GPU and with a processor Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz.

2.2.3. Beam geometry sampling The second element of our method is the construction of the X-ray beam geometry using beamlets. In cases where the beam geometry can be analytically modeled, it is easy to sample its shape with beamlets. Simply generate beamlets randomly and uniformly, following the origin, direction, and energies of the source spectrum. For instance, in the case of a point source, the beamlet direction will be sampled according to isotropic emission angles $\{\theta, \psi\} \in 4\pi$, or $\{\theta, \psi\} \in [-\frac{\alpha}{2}, +\frac{\alpha}{2}]$ for a cone beam or, $\{0, \psi\} \in [-\frac{\alpha}{2}, +\frac{\alpha}{2}]$ for a fan beam, with α the aperture. In the case of a collimated sources, whether simple or complex, it is possible to use a rejection method with ray tracing. The beamlet is then represented as a parametric line (ray) and the intersection between the collimator and this line is calculated. Only beamlets that not hit the collimator are used to compute the dose to the patient. This allow to simulate square collimator in fluoroscopy for example, or a multi-leaf collimator in radiotherapy. If need the origin of the beamlet, i.e. position in space, can also be randomly generated to simulate a moving source. Regarding the beamlet energy, the source spectrum have to be respected. In order to random the beam energy according different probabilities,

a uniform random process on the cumulative density function (CDF) derived from the spectrum can be used.

2.2.4. Patch extraction In order to limit the size of the patch and the network, only pixels around the beamlet are considered in the dose prediction. As the beamlets are oriented, it is necessary to be able to extract the patch from the CT image based on the beamlet's orientation. This is mandatory since the network architecture used only axis-aligned image. Therefore, the oriented patch image $\mathbf{I}_{op}$ has to be transformed into an image axis-aligned patch $\mathbf{I}_{aap}$. Similarly, the axis-aligned dose patch after prediction has to be transformed and placed back into the dose image at the right beamlet position.

This was done by first calculating the entry point of the beamlet into the patient's CT image. To achieve this, the beamlet is treated as a parametric line as mention previously and defined as follow,

$$\mathbf{b} = \mathbf{b}_o + \mathbf{d}l \quad (1)$$

with $\mathbf{b} = \{b_x, b_y, b_z\}^T$ a point along the beamlet path (L), $\mathbf{b}_o = \{b_{ox}, b_{oy}, b_{oz}\}^T$ the origin position of the source, $\mathbf{d} = \{d_x, d_y, d_z\}^T$ the direction of the beamlet, and l the parametric distance along the line (L). CT images are naturally axis-aligned images. Then the bounding geometry of the CT phantom is defined as an Axis Aligned Bounding Box (AABB). AABB objects are aligned with the axis $\{u_x, u_y, u_z\}$ of the simulation. Therefore, the intersection between the beamlet and the CT image was solved using an efficient ray/box intersection algorithm [Smits, 2002]. The interaction point between the beamlet and the AABB was determined by considering each intersection of the beamlet with the slabs that compose the AABB. A slab is defined as the surface between a pair of parallel planes or lines. Every distance between the ray and the minimum and maximum boundary slabs were calculated using their respective plane equations. The final intersection distance $l_{\square}$ with the AABB, was given by the minimum positive value between all slabs intersections. The point $\mathbf{p}_{\square}$ that intersect the phantom was subsequently calculated using the parametric line equation.

Based on this intersection point and the beamlet direction an affine transformation matrix $\mathbf{M}$, that transform the beamlet oriented patch to be extracted $\mathbf{I}_{op}$ in an axis-aligned patch $\mathbf{I}_{aap}$, was defined. This matrix was composed of a 3×3 sub-matrix for the rotation $\mathbf{R}_{xyz}$ and a 1×3 sub-matrix for the translation $\mathbf{T}_{xyz}$, as follow,

$$\mathbf{M} = \left\{ \begin{array}{cc} \mathbf{R} & \mathbf{T} \\ \mathbf{0} & 1 \end{array} \right\} \quad (2)$$

A backward sampling approach was used to align the patch. This avoids aliasing and holes in the images (voxel without values) due to numerical approximation. Instead of transforming directly the oriented patch image into axis-aligned $\mathbf{I}_{op} \xrightarrow{\mathbf{M}} \mathbf{I}_{aap}$ with the matrix $\mathbf{M}$, the inverse process was used. For each pixel of the targeted image $\mathbf{I}_{aap}$, the inverse matrix $\mathbf{M}^{-1}$ was used to find the corresponding voxel position in $\mathbf{I}_{op}$. The value of this voxel, using a linear interpolation, was then copied into the targeted image $\mathbf{I}_{aap}$.

After predicting the dose of the current CT patch, the corresponding dose is placed back, by following the same backward sampling process by directly using the transformation $\mathbf{M}$.

2.3. Evaluation study

2.3.1. Beamlet dose prediction The proposed deep learning approach was first evaluated at the beamlet level. The aim was estimated the level of accuracy of the CNN dose model. For each beamlet contains in the test set the corresponding CT image patch was extracted and used to estimate the dose with our networks. The predicted doses were compared against the beamlet dose from the Monte Carlo simulation consider here as gold standard. The proposed approach requires two inferences, one to recover the primary dose contribution and another one to get the dose from the scattering. Since absorbed doses are cumulative, the two dose maps were then sum up to obtain the final dose map that will be used for comparison.

Two metrics, the Absorbed Dose Error (ADE) and the Mean Absolute Error (MAE) were calculated in order to estimate the robustness of the networks. The ADE correspond to the total dose absorbed by each patch and was calculated as:

$$ADE = \frac{|\sum_{i=0}^N D_i - \sum_{i=0}^N \tilde{D}_i|}{\sum_{i=0}^N D_i} \quad (3)$$

where D_i and $\tilde{D}_i$ are respectively the dose at voxel i by the gold standard Monte Carlo simulation and by the CNN. N is the number of voxels that compose the patch. The MAE was calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=0}^N |D_i - \tilde{D}_i| \quad (4)$$

2.3.2. Cone Beam CT imaging The heart of the proposed method was to combine a sampling approach, like Monte Carlo, with a deep learning based dose engine, in order to gain in versatility. For example, in Monte Carlo simulation a cone beam CT is generated by randomly emitting particle that follow the beam geometry, like the aperture angle or the beam direction. In the proposed method the philosophy was the same. Instead of emitting particles, beamlets that follow the characteristics of any desired complex beam were randomly produced. And the absorbed dose of each beamlet that reach the patient was estimated by using the deep learning method. As Monte Carlo, by sampling enough beamlet, the dose accumulated will converge to a dose map that will reveal the dose distribution of the complex X-ray beam.

In order to illustrate, and evaluate the proposed mechanism a cone beam CT simulation was performed. More precisely, a beam with a rectangular source of $0.5 \times 0.3 \text{ mm}^2$ and with an aperture of 10° were considered. A classical tube voltage of

125 kVp and a 2 mm aluminum filter was defined. The X-ray spectrum was obtained using the TASMIP model [Boone and Seibert, 1997]. For this simulation, we used a CT scan of the abdominal region that was not contained in the training data set. The CT scan was resized to have a resolution of $3 \times 3 \times 3 \text{ mm}^3$ and a size of $123 \times 81 \times 80$ voxels. The X-ray source was placed at 42 cm of the patient back. The ground truth was obtained with MCS using 8×10^7 particles (see Figure 5).

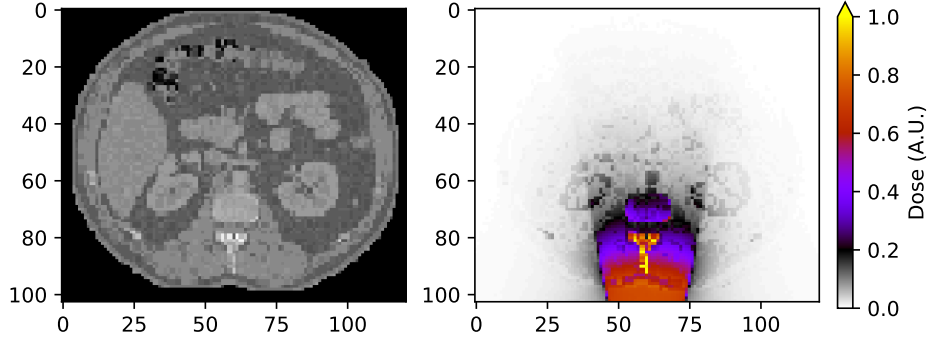


Figure 5. Cone Beam CT imaging application, with a slice showing the patient CT image (left) and the associated absorbed dose map of the beam calculated by MCS (right)

Beamlets were randomly generated using the cone beam specification, in term of energy and geometry. The entry point of the beamlet within the patient was calculated using a ray casting approach. Based on this position and the beamlet direction, the corresponding image patch was extracted on-the-fly. This involved a rotation to realign the patch horizontally how it was train to the network. Subsequently, this patch and the beamlet energy was fed to the network and the corresponding dose map was replace back by adding dose values into the final 3D dose map, which has the same size that the patient CT image. Image rotation used a fast algorithm of re-sampling with a Nearest Neighbor approximation to reduce the computational burden and to preserve the sharpness of the primary dose beam. The evaluation was performed by comparing MCS and the proposed method for different number of beamlets. Each time, the MAE and ADE metrics were calculated considering the dose within the X-ray primary beam, where the values are higher and the dose outside the beam, where the scattered dose are smaller.

3. Results

3.1. Beamlet dose prediction

The first and the second network were trained for approximately 4 hours, and 22 hours respectively. The distribution of ADE, MAE, and MAE only within the primary beam for the test set were shown in Figure 6. Across the 2000 patches that compose the test set, the ADE was $1.2 \pm 3.87\%$, $1.90 \pm 1.73 \times 10^{-6}$ for the MAE calculated on whole

patches and $4.6 \pm 9.4 \times 10^{-4}$ for the MAE calculated only on the primary dose deposition maps.

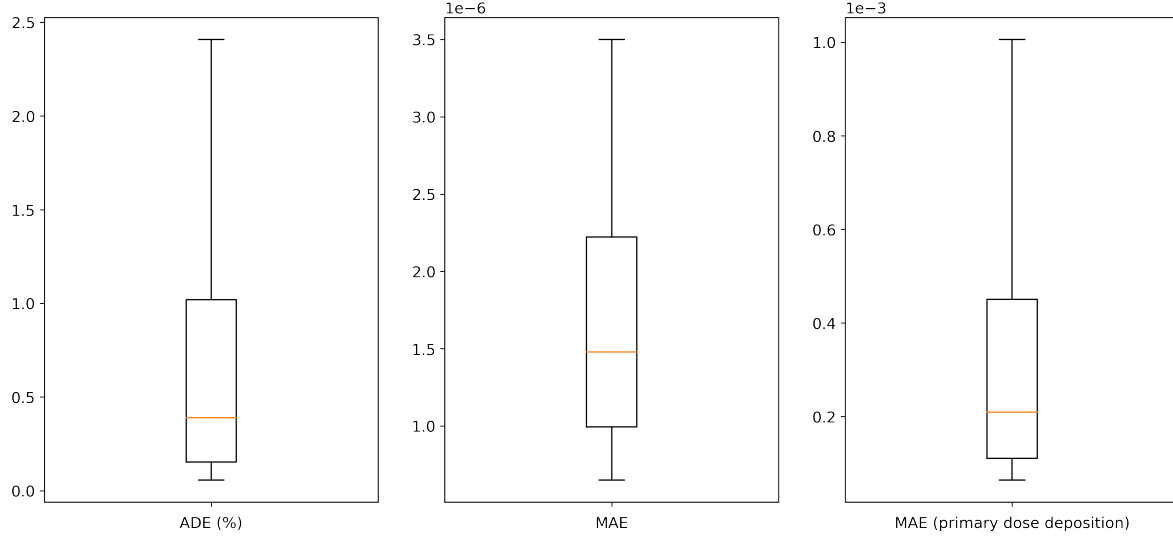


Figure 6. Distribution of the ADE, MAE and MAE only within the primary beam, for the 2000 patches that compose the test set.

Overall, the proposed networks has provided satisfying dose prediction map. Three examples of predicted dose map are shown in Figure 7, the best and worst case as well as a median case. For the three patches, the ground truth from MCS, the prediction from the network, and the absolute error map were provided. Even in the worst case, the prediction was mainly in a good agreement.

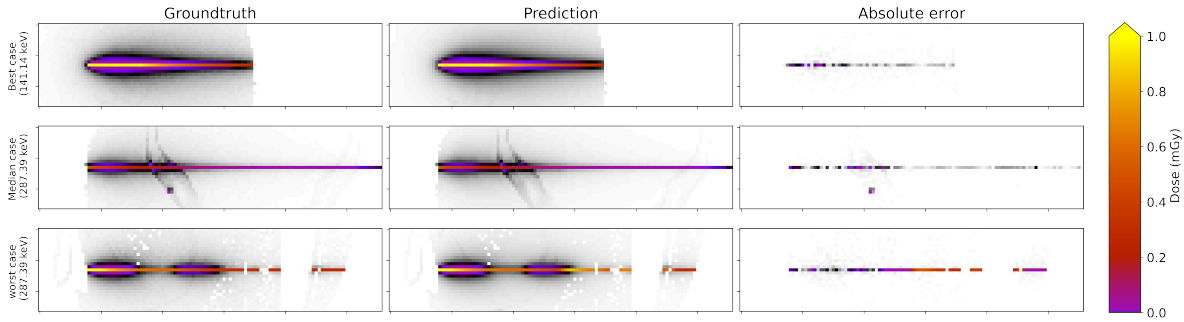


Figure 7. Three examples of predicted dose map: best (top), median (middle) and worst case (bottom). The ground truth from MCS (left), the prediction from the network (center), and the absolute error map (right) are provided.

3.2. Cone Beam CT imaging

The performance of the proposed method was studied according the number of simulated beamlets. Each time, using the same computer, a MCS was performed with the same amount of execution time that the proposed solution. Then both dose map were

compared. Results for different number of beamlets are presented in Figure 8 with a comparison with the ground truth which is also a MCS but with a large number of particles. The convergence of the proposed method was faster compare to MCS. Which is normal since the beamlet fill the dose values all along the path of the beamlet. For 9000 beamlets a run time of 1.5 min was needed. For the same among of time, MCS was still remaining noisy due to the lack of sample. To recover an equivalent result than the proposed method for 1.5 min, MCS has needed at least x15 more running time (22 min). However, the proposed method has shown no significant improvement with more beamlets, which is not the cas with MCS where more sample clearly decrease the discrepancy with the ground truth. One element was also emphasis with this evaluation is the fact that the proposed beamlet method has some difficulty to fully recover scattered dose outside the main primary beam. For example the left column in Figure 8, even with more beamlet it was not possible to distinguish the shape of the patient, which was not the case with MCS. In addition, some artefact appears on the border of the beam while using the proposed method. Regarding the execution time, 38% was used by the inference of the network and 62% for the patches/dose rotation.

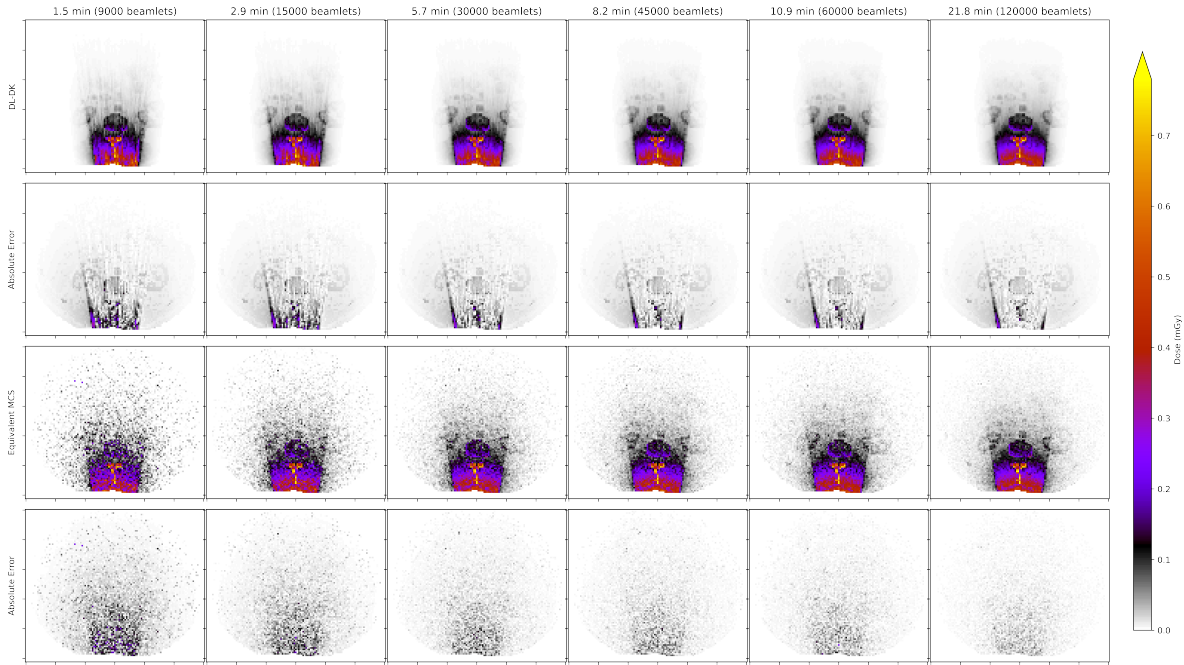


Figure 8. Example of transversal slices for the dose map prediction of the proposed method (first row), its absolute error compare to the ground truth (second row), the equivalent dose map from MCS for a among of run time (third row), and the absolute error of the current MCS with the ground truth.

In addition, we can observe in Figure 8 that the noise created by the proposed method was different than the noise from MCS. Indeed, the dose map tends to be blurred compare to the ground truth, since the deep learning approach are composed of convolution. It is particularly visible on the vertebral column (see right column Figure 9).

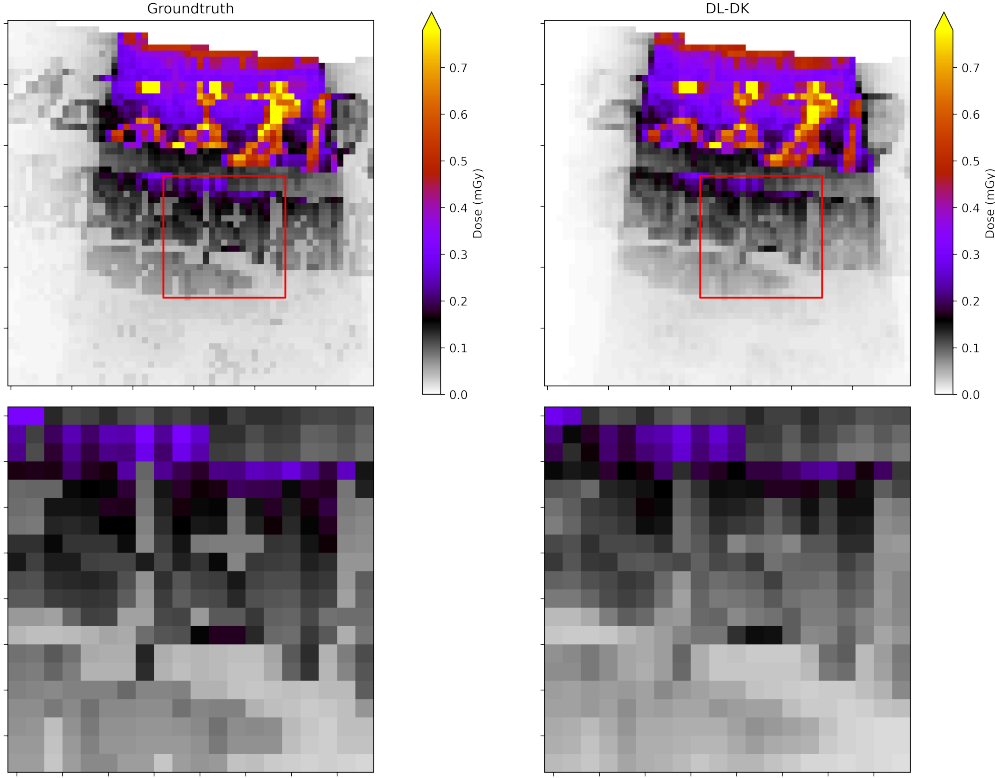


Figure 9. Detail of the sagittal slice of the dose map from the ground truth (MCS) and the proposed method for 45000 beamlets.

Table 1. Results for X-ray beam prediction using 9000 beamlets.

Sampled with beamlets			MCS
Inside the beam	MAE	0.0111	0.0169
	ADE	9.8%	6.9 %
Full dose map	MAE	0.0071	0.0097
	ADE	25%	6.9%

The next results have considered two different simulations, using 9000 and 45000 beamlets for the proposed method and two MCS equivalent in terms of execution time. The different metrics considered were reported in Table 1 for the dose predicted using 9000 beamlets and in Table 2 for the dose predicted using 45000 beamlets. We can observe that the proposed method have provided a better MAE inside the beam trajectory, with an improvement of 34% compared to equivalent MCS for simulations of 9000 beamlets and 22% for simulations of 45000 beamlets. However, the ADE was better for MCS compared to the beamlet approach. When we consider the whole dose map, the MAE decrease, showing that the average dose value outside the beam was less accurate. Same conclusion with the ADE which was 23% for the proposed method.

Table 2. Results for X-ray beam prediction using 45000 beamlets.

		Sampled with beamlets	MCS
Inside the beam	MAE	0.0078	0.0090
	ADE	6.9%	1.8 %
Full dose map	MAE	0.0061	0.0056
	ADE	23%	1.8%

4. Discussions

In this work, an hybrid method using random sampling and deep-learning approaches were explored for dose application in X-ray imaging. The aim was improving the MCS efficiency by using deep learning but with a more generic solution that only predicted dose map from a given X-ray application. The versatility of the proposed solution allows multiple X-ray system simulations without retraining the deep learning model with new beam specificity. Any geometry of X-ray beam can be simulated without the need of large training data set and complex network architecture. The same trained model, learned using different anatomical sites, is able to calculate the 3D dose within the patient for helical CT, cone beam CT, fan beam CT or any colimated beam shape.

The dose prediction from the network has shown a relative absorbed dose error of $1.2 \pm 3.87\%$ compare to MCS. For a full example in cone beam CT, the proposed method has shown a faster convergence compare to standard Monte Carlo simulation. Although the results have shown a good agreement between the beamlet approach and the ground truth from Monte Carlo simulation, some artefacts have appeared on the close border of the X-ray beam. Therefore, relative absorbed dose error within the beam reach almost 7%. More investigation should be achieved to verify if this is bring by the ray casting or the patch extraction. One more limitation of the proposed method was the difficulty to recover dose that are outside the X-ray primary beam. This correspond to the dose deposited to the patient from the scattered particle. This come from the nature of the method that use patch-based approach. The very small dose values outside the patch are then not considered. However those dose values are two order of magnitude smaller than the dose inside the X-ray beam. One solution to solve this issue will be to predict the dose of a beamlet using the entire CT image, and then considering far scattered particle. This will solve the missing scattering dose but will slightly increase the computation time since the network will need to predict a large dose map. Regarding the current results, numerous improvement can be explored for improving the accuracy. This first study was a proof-of-concept showing promising result for solution that mix sampling approach and deep-learning based dose calculation for more versatile MCS.

Execution time and comparison where performed using MCS with an equivalent run time of the beamlet approach. This was not the most suitable solution for timing comparison, but it was more easier since both methods not provide the same statistical

uncertainty. Although the proposed method was faster to converge compare to MCS, more improvement is possible since 62% of the running time was used by the patch extraction. This include the rotation of the oriented patch into axis aligned image, because the network has learned image in this frame, and the reverse operation that consist to rotate back the predicted dose into the final dose map. Two possible way can be used to improve the computation time. One consisting of developing fast image rotation using also deep learning or GPU programming. Another method will consist to directly learn an oriented patch image by the network with a specific architecture or using an image representation which is invariant in rotation.

Regarding dedicated solution that predict directly the dose map such in [Villa et al., 2023], the execution time of the proposed method was slower. Dedicated deep learning model may predict in less than 1s a 3D dose map. However, the proposed beamlet approach is a solution between full MCS, which is versatile but very slow, and full deep-learning, which is trained for a given application but very fast. The proposed solution try to mix both advantages by combining MCS and deep learning approach.

5. Conclusion

A versatile approach that combines Monte Carlo sampling and deep learning dose calculation was explored. The proposed idea was decomposing any x-ray beam into elemental small beams name beamlet. Then by using a sampling approach it is possible to simulate any beam geometry and energy. In this case only a neural network that predict the dose of a beamlet within patient CT was required. This allows multiple X-ray system simulations without retraining the model with new beam specificity. Results have shown that the dose engine based on deep learning have leaded a relative dose error about $1.2 \pm 3.87\%$ compare to the reference dose. For a more realistic simulation, in cone beam CT, dose results have shown a relative error within the beam of 7% compare to a full MCS. The convergence of the proposed method was faster compare to MCS, and need at least x20 more time for equivalent dose results. This method is promising and required more investigation, especially to reduce artefact on the beam border and also to speed up image processing required to read and write ptach within the CT and dose map.

Acknowledgement

This work was partially funded by the French National Research Agency through the MoCaMed project (ANR-20-CE45-0025).

References

- [Ahnesjö, 1989] Ahnesjö, A. (1989). Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Medical physics*, 16(4):577–592.

- [Behloul et al., 2018] Behloul, A., Visvikis, D., and Bert, J. (2018). Improved woodcock tracking on monte carlo simulations for medical applications. *Physics in Medicine & Biology*, 63(22):225005.
- [Bert et al., 2013] Bert, J., Perez-Ponce, H., Bitar, Z. E., Jan, S., Boursier, Y., Vintache, D., Bonissent, A., Morel, C., Brasse, D., and Visvikis, D. (2013). Geant4-based monte carlo simulations on GPU for medical applications. *Physics in Medicine and Biology*, 58(16):5593–5611.
- [Boone and Seibert, 1997] Boone, J. M. and Seibert, J. A. (1997). An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kv. *Medical physics*, 24(11):1661–1670.
- [Götz et al., 2020] Götz, T. I., Schmidkonz, C., Chen, S., Al-Baddai, S., Kuwert, T., and Lang, E. W. (2020). A deep learning approach to radiation dose estimation. *Physics in Medicine & Biology*, 65(3):035007.
- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [Kontaxis et al., 2020] Kontaxis, C., Bol, G., Lagendijk, J., and Raaymakers, B. (2020). Deepdose: towards a fast dose calculation engine for radiation therapy using deep learning. *Physics in Medicine & Biology*, 65(7):075013.
- [Krieger and Sauer, 2005] Krieger, T. and Sauer, O. A. (2005). Monte carlo- versus pencil-beam-/collapsed-cone-dose calculation in a heterogeneous multi-layer phantom. *Physics in Medicine and Biology*, 50(5):859–868.
- [Lee et al., 2019] Lee, M. S., Hwang, D., Kim, J. H., and Lee, J. S. (2019). Deep-dose: a voxel dose estimation method using deep convolutional neural network for personalized internal dosimetry. *Scientific reports*, 9(1):1–9.
- [Liu et al., 2019] Liu, Z., Fan, J., Li, M., Yan, H., Hu, Z., Huang, P., Tian, Y., Miao, J., and Dai, J. (2019). A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. *Medical Physics*, 46(5):1972–1983.
- [Mohan et al., 1986] Mohan, R., Chui, C., and Lidofsky, L. (1986). Differential pencil beam dose computation model for photons. *Medical Physics*, 13(1):64–73.
- [Neishabouri et al., 2021] Neishabouri, A., Wahl, N., Mairani, A., Köthe, U., and Bangert, M. (2021). Long short-term memory networks for proton dose calculation in highly heterogeneous tissues. *Medical Physics*, 48(4):1893–1908.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.
- [Roser et al., 2019] Roser, P., Zhong, X., Birkhold, A., Strobel, N., Kowarschik, M., Fahrig, R., and Maier, A. (2019). Physics-driven learning of x-ray skin dose distribution in interventional procedures. *Medical Physics*, 46(10):4654–4665.
- [Sarrut et al., 2014] Sarrut, D., Bardès, M., Boussion, N., Freud, N., Jan, S., Létang, J.-M., Loudos, G., Maigne, L., Marcatili, S., Mauxion, T., Papadimitroulas, P., Perrot, Y., Pietrzyk, U., Robert, C., Schaart, D. R., Visvikis, D., and Buvat, I. (2014). A review of the use and potential of the gate monte carlo simulation code for radiation therapy and dosimetry applications. *Medical Physics*, 41(6Part1):064301.
- [Schneider et al., 2000] Schneider, W., Bortfeld, T., and Schlegel, W. (2000). Correlation between CT numbers and tissue parameters needed for monte carlo simulations of clinical dose distributions. *Physics in Medicine and Biology*, 45(2):459–478.
- [Smits, 2002] Smits, B. (2002). Efficient bounding box intersection. *Ray Tracing News*, 15.
- [Tian et al., 2015] Tian, Z., Shi, F., Folkerts, M., Qin, N., Jiang, S. B., and Jia, X. (2015). A GPU OpenCL based cross-platform Monte Carlo dose calculation engine (goMC). *Physics in Medicine and Biology*, 60(19):7419–7435.
- [Unkelbach et al., 2015] Unkelbach, J., Bortfeld, T., Craft, D., Alber, M., Bangert, M., Bokrantz, R., Chen, D., Li, R., Xing, L., Men, C., Nill, S., Papp, D., Romeijn, E., and Salari, E. (2015). Optimization approaches to volumetric modulated arc therapy planning. *Medical Physics*,

42(3):1367–1377.

- [Villa et al., 2021] Villa, M., Bert, J., Valeri, A., Schick, U., and Visvikis, D. (2021). Fast monte carlo-based inverse planning for prostate brachytherapy by using deep learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, pages 1–1.
- [Villa et al., 2023] Villa, M., Nasr, B., Benoit, D., Padoy, N., Visvikis, D., and Bert, J. (2023). Fast dose calculation in x-ray guided interventions by using deep learning. *Physics in Medicine and Biology*, 68(16).
- [Williamson, 1987] Williamson, J. F. (1987). Monte carlo evaluation of kerma at a point for photon transport problems. *Medical Physics*, 14(4):567–576.