Physics in Medicine & Biology

PAPER

Deep learning-based voxel sampling for particle therapy treatment planning

A Quarz[1,2] ⓘ, L Volz[1] ⓘ, C Hoog Antink[2] ⓘ, M Durante[1,3,4] ⓘ and C Graeff[1,2,*] ⓘ

1 Biophysics Department, GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt, Germany
2 Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt, Darmstadt, Germany
3 Department of Condensed Matter Physics, Technische Universität Darmstadt, Darmstadt, Germany
4 Department of Physics 'Ettore Pancini', University Federico II, Naples, Italy
* Author to whom any correspondence should be addressed.

E-mail: c.graeff@gsi.de

Abstract

Objective. Scanned particle therapy often requires complex treatment plans, robust optimization, as well as treatment adaptation. Plan optimization is especially complicated for heavy ions due to the variable relative biological effectiveness. We present a novel deep-learning model to select a subset of voxels in the planning process thus reducing the planning problem size for improved computational efficiency. *Approach.* Using only a subset of the voxels in target and organs at risk (OARs) we produced high-quality treatment plans, but heuristic selection strategies require manual input. We designed a deep-learning model based on *P*-Net to obtain an optimal voxel sampling without relying on patient-specific user input. A cohort of 70 head and neck patients that received carbon ion therapy was used for model training (50), validation (10) and testing (10). For training, a total of 12 500 carbon ion plans were optimized, using a highly efficient artificial intelligence (AI) infrastructure implemented into a research treatment planning platform. A custom loss function increased sampling density in underdosed regions, while aiming to reduce the total number of voxels. *Main results.* On the test dataset, the number of voxels in the optimization could be reduced by 84.8% (median) at <1% median loss in plan quality. When the model was trained to reduce sampling in the target only while keeping all voxels in OARs, a median reduction up to 71.6% was achieved, with 0.5% loss in the plan quality. The optimization time was reduced by a factor of 7.5 for the total AI selection model and a factor of 3.7 for the model with only target selection. *Significance.* The novel deep-learning voxel sampling technique achieves a significant reduction in computational time with a negligible loss in the plan quality. The reduction in optimization time can be especially useful for future real-time adaptation strategies.

1. Introduction

Charged particle therapy is a rapidly growing cancer treatment modality owing to the physical dose advantage of particles over photons (Durante and Paganetti 2016). Most charged particle therapy centers use protons to treat patients, but 14 centers in Asia and Europe treat tumors with high-energy carbon ions, and several more are under construction ((PTCOG) PTC-OG 2023), including one in the USA (Pompos *et al* 2022). Compared to proton therapy, carbon ion therapy benefits not only from reduced lateral scattering but also from a higher relative biological effectiveness (RBE) and other biological advantages (Tinganelli and Durante 2020).

Typically, radiotherapy is administered in fractions over the course of multiple weeks. For both carbon ion and proton therapy, adaptive therapy is needed to ensure dosimetric accuracy over the course of the fractionated treatment (Bobic *et al* 2021, Paganetti *et al* 2021, Jia *et al* 2022). Yet, workflows in clinical particle therapy centers are tightly packed to maximize patient throughput, and as such, adaptation methods

need to fit within a tight schedule to enable frequent replanning. Runtime efficiency for plan optimization can be a hurdle for online, and especially for future real-time adaptive workflows. For carbon ions, the high computational load, posed by the variable RBE when calculating biological doses, presents a hurdle in computational efficiency. The optimization time increases tremendously when complex plans must be calculated—e.g. robust (Wolf *et al* 2020), 4D (Graeff 2014) or carbon ion arc (Mein *et al* 2022) optimization—because of the large associated dose influence matrix. New solutions are required to make these particle therapy treatment planning strategies more efficient to yield the best patient treatments. Beyond clinical application, efficient, accurate and fast plan optimization is a key to enable fast turnover in treatment planning research in this growing field.

One option to increase computational efficiency of the treatment planning procedure typically applied in proton therapy is to reduce the problem size, for example by isotropic increase of the voxel size (Espãa and Paganetti 2011) or sampling the voxels used in the treatment planning process (Lu and Chin 1993, Martin 2007). However, these techniques cannot easily be translated to carbon ion therapy, due to the small straggling of heavy ions and consequently sharp dose gradients. Larger voxels have already been demonstrated that they lead to inferior dosimetric quality for proton therapy of lung cancer (Espãa and Paganetti 2011), and this problem is expected to be worsened by the steeper dose gradients of carbon ion therapy. In addition, there is no consensus on best practices for planning voxel sizes for different patient regions. Voxel sampling would in principle be a promising technique to keep the target coverage, while reducing the computational demand. However, for heuristic sampling techniques, the choice of sampling parameters, such as sampling densities within different volumes of interest (VOIs), can require manual input and patient-specific fine tuning which would negate the original purpose of voxel sampling.

Artificial intelligence (AI) is now benefiting nearly every scientific field and has found numerous applications in radiotherapy (Huynh *et al* 2020). Several works described the potential of AI to speed up or improve particle therapy in all stages of the radiotherapy workflow, such as segmentation, treatment planning, quality assurance and patient follow up (Graeff *et al* 2023). In this work, we developed a new deep learning model to create fully automatic patient-specific voxel sampling for carbon ion therapy of head and neck cancer patients. The model was integrated into the TRiP98 research treatment planning platform (Krämer and Scholz 2006, Hild *et al* 2014), to enable an efficient training and evaluation process.

2. Materials and methods

2.1. Treatment planning and initial problem size reduction
Treatment plans were generated and evaluated using the GSI Helmholtz Centre for Heavy Ion Research GmbH, Darmstadt, Germany, (GSI) in-house treatment planning software TRiP98 (Krämer and Scholz 2006, Hild *et al* 2014). TRiP98 features extensive infrastructure for robust, biological dose optimization and 4D-optimized treatment planning (Wolf 2019, Steinsberger *et al* 2023), and has recently been extended to carbon ion arc planning. To deal with the excessive computational demand of these advanced planning techniques, several methods to reduce the optimization problem size have been implemented that maintain the high-resolution dose grid necessary for sharp carbon ion beams. As the first step, voxels that are not expected to receive a dose greater than a user defined dose threshold q are removed from the optimization. After calculating the dose influence matrix for all target and OAR voxels, the matrix entries d_{ij} for all OAR voxels are analyzed. A specific OAR voxel k is removed from the optimization problem if it is unlikely to receive relevant doses during treatment planning according to the following equation:

$$\forall\, i, j : \max\left(d_{kj}\right)\, \mathrm{DC}\left(\mathrm{OAR}\right)\, q < \max\left(d_{ij}\right). \tag{1}$$

If the maximum contribution from any beam j to the OAR voxel k, multiplied with its dose constraint DC relative to the prescribed target dose and a given factor q is smaller than the maximum contribution to any target voxel i, then this OAR voxel is removed. For the remainder of this study, $q = 0.3$ is used for all dose calculations in this work, except where explicitly stated.

In particular, the dose threshold option removes OAR voxels in the penumbra of the beams, while it preserves voxels closer to the beam center and the spread-out Bragg peak. If the OAR voxels closer to the beam fulfil the dose constraints, those further out will receive even less dose.

2.2. Heuristic voxel sampling
Furthermore, TRiP98 features a voxel sampling procedure. Prior work from our group found that a dense sampling of the voxels located in the shell of VOIs combined with less dense sampling in the VOI interior can achieve similar dosimetric plan quality while greatly reducing the computational cost (Wolf *et al* 2019). For this, both target and OAR VOIs are segmented into an outer shell and an inner core based on a 3D-distance

transform (Maurer *et al* 2003). The user can thereby set the sampling factor for the boundary and interior, as well as the boundary width. The sampling factor determines the likelihood for any given voxel to be included in the optimization problem, which is determined once prior to optimization via randomization. The concept utilizes the fixed relationship between the weight of a beam spot and the dose to neighboring voxels. As the dose given by a certain beam to individual neighboring voxels is not independent, it is sufficient to optimize the beam weight based on the dose in the relevant voxels, with the dose in the other voxels following accordingly.

In a region of homogeneous plan dose, if the voxel sampling is dense enough and isotropic, fulfilling the dose constraints in the subset of voxels selected by the voxel sampling translates to fulfilling the dose constraint in all voxels. At the target boundary, where the dose is no longer homogeneous, and the sampling can no longer be isotropic, a denser sampling of voxels is necessary.

Note that it is up to the user to define 'enough' in the above approach: not only does this depend on the specific patient, tumor size/shape and distance to OARs, a specific goal to be optimized also needs to be defined. Lower sampling automatically means less computational demand, but at the same time, depending on the specific problem at hand, may result in a suboptimal plan regarding the set dosimetric constraints. For example, tight dose constraints for close OARs may cause target coverage issues, if the sampling of the target in that region was set too coarse. The relevant parameters are the sampling density in the interior and boundary of the VOIs, as well as the boundary width. For this work, we used a sampling ratio of 1 for the boundary, and a sampling ratio of 1/8 for the interior of the VOIs, at a fixed width of 2 mm for the boundary when using the heuristic voxel sampling method. These values do not reflect the optimum for each patient and were found in preparation work from our group to be applicable to all the patients contained in this work without causing major issues for dose coverage in individual cases.

2.3. Patient data

Treatment planning data were taken from the GSI carbon ion pilot project for a cohort of head and neck cancer patients treated between 1997 and 2008 at GSI with carbon ions. The pilot project was approved was provided by the ethical committee of the University of Heidelberg. Anonymized treatment plans of all patients are stored for research purposes at the GSI. Informed consent is waived by the ethical committee of the University of Heidelberg for anonymized research on the data of the pilot project trial.

70 head and neck cancer patient plans were randomly selected from the database. The tumour histology comprised of adenoid cystic carcinoma, chondrosarcoma or chordoma, all located in different regions of the head or base of the skull. Optical nerves, chiasm, eyes, spinal cord, brainstem, and parotid (for two patients only) were included as OARs in the plan optimization. Overall, the patients displayed a large variability in tumour size (from 24.5 to 575 cm^3), position, and OARs. Therefore, the dataset represents a wide range of anatomies for the training data. We did not exclude difficult cases, where some of the OARs are wrapped in the tumour, in order not to compromise the generality of the model. Fifty patients were selected randomly for the training dataset and the remaining twenty were split into two equal groups for the test and validation datasets. In the training dataset there were two patients with specific anatomical variation: one had a titanium implant and one had only one eye. The datasets are summarized in table 1.

The original plans for these patients were generated with an earlier version of TRiP98 and clinically approved and delivered. For this work, all treatment plans were re-generated based on the original plan parameters with the latest TriP98 version with few modifications. For all the patients, two treatment directions were selected from the original treatment plans and a multi-field optimization was performed, i.e. both treatment fields were jointly optimized as opposed to a single-field uniform dose approach. The distance between adjacent beam spots was adjusted to 2×2 mm. A fixed, horizontal, but slightly downward tilted beam line (gantry angle was $-92.203°$) was assumed as per the specifications of the GSI medical research cave (Lis *et al* 2021), where the patients had originally been treated. The prescription target dose was adjusted to a uniform RBE-weighted dose of 3 Gy. All OAR constraints were maintained, but for comparability priority was given to target coverage. To enable an unbiased comparison, the plan optimization was terminated after 100 iterations.

2.4. Data pre-processing

The deep learning model requires every input to be of the same dimension. To fulfil this requirement the potential input was centred in a volume of size (40, 104, 114), corresponding to (*Z, Y, X*) dimensions of the distance transform provided by TRiP98. These values were chosen at the beginning of the development process as to be large enough to handle the input of all training and validation data, with additional margins accounting for the unknown maximum input size of the test dataset (table 1). Before inputting the data to the convolutional neural network (CNN), equation (1) was applied to all VOIs using a factor $q = 0.3$. Individual inputs were enlarged to the full input size by padding with zeros. The median occupancy of the

Table 1. Patient data.

	Train	Validation	Test	Total
Treatment plans	50	10	10	70
Adenoid cystic carcinoma	11	1	1	13
Chondrosarcoma	10	3	4	17
Chordoma	29	6	5	40
Median tumour volume, cm^3 (min–max)	134.4 (42.8–575.0)	72 (24.5–340.5)	78.0 (41.1–216.4)	115.5 (24.5–575.0)
Median number of voxels for optimisation $\times 10^3$ (min–max)	35.6 (13.0–136.5)	21.6 (13.3–54.3)	19.1 (8.1–93.9)	31.4 (8.1–136.5)
Median number of voxels in the tumour $\times 10^3$ (min–max))	30.8 (9.7–130.7)	17.72 (9.4–49.2)	16.4 (5.6–90.3)	27.2 (5.6–130.7)
Median fraction of tumour in input (min–max), %	86 (75–96)	82 (70–91)	85 (69–96)	85 (69–96)
Median size of input LR: X (min–max)	65 (45–98)	53 (36–102)	54 (42–90)	61.5 (36–102)
Median size of input AP: Y (min–max)	70.5 (46–102)	58 (42–96)	62 (48–96)	67 (42–102)
Median size of input SI: Z (min–max)	27 (16–39)	23 (16–40)	25 (18–32)	27 (16–40)

available input size was 54% in the *X*-axis, 64% in the *Y*-axis and 68% in the *Z*-axis. This step is necessary but makes the input sparse, especially taking into consideration the space between tumour and OARs as well. Within the final input volume, a median of 85% of the voxels corresponded to the target.

2.5. Model
A bespoke CNN architecture was designed based on TensorFlow (Abadi *et al* 2016) (version 2.9.1) and TensorFlow -probability (Dillon *et al* 2017) (version 0.17.0) for this work.
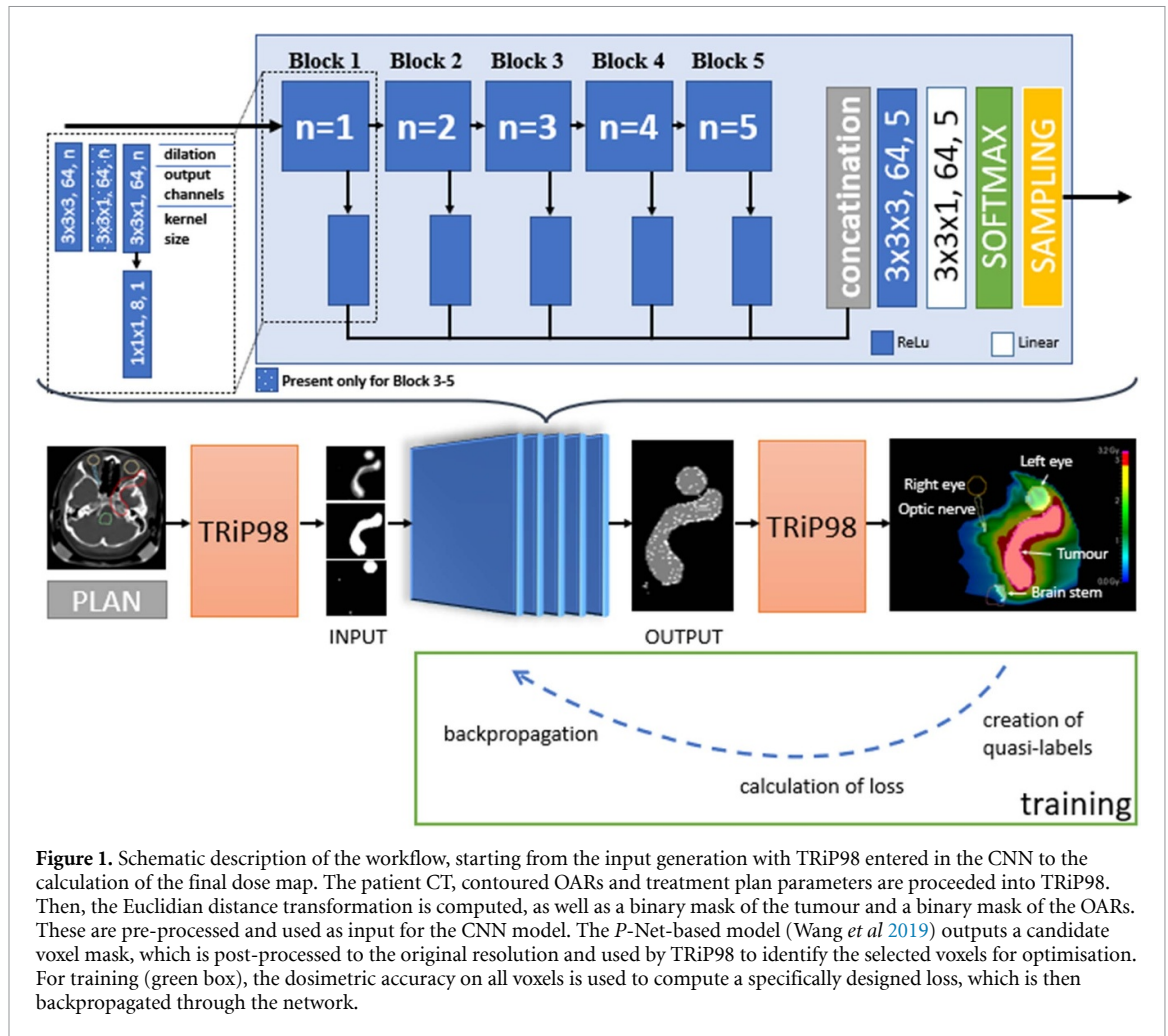
2.5.1. Architecture
We based the CNN architecture on *P*-Net (Wang *et al* 2019), which can capture high-level features through a dilation factor without lowering the resolution. The *P*-Net architecture showed impressive results in image processing (Furuta *et al* 2020) as well as in segmentation tasks (Yu and Koltun 2016, Liao *et al* 2020, Ma *et al* 2021). We opted for this approach because in our problem the distance between single voxels is crucial.

The model input was composed of a distance transformation, containing the 3D-Euclidian distance for each voxel to the boundary of its enclosing VOIs, as well as binary representations of the target and OAR (after applying equation (1) VOIs. This data is always computed within TRiP98 for plan optimization. Using only available data was a key design choice to avoid any need for user input.

The CNN architecture was mostly equivalent to the 3D *P*-Net (Wang *et al* 2019), however, the input downsampling and output upsampling were not used to keep the original voxel resolution throughout. The network consisted of five blocks, where each block was a combination of two (blocks 1 and 2) or three (blocks 3–5) convolutional layers. The first two (three) convolutional layers formed a sub-block, the output of which was proceeded as input to the next block, as well as to a separated convolutional layer. The outputs of the separated convolutional layer of each block were directly proceeded further after the fifth block, where they were concatenated, and then processed with two convolution layers. A softmax activation was applied to the output of the final layer to yield the sampling probability for individual voxel positions, which was then used for a random sampling operation. This network design enables to gather all information from low to high features together to form the result. A schematic depiction of the deep-learning workflow, complete with the network architecture is provided in figure 1, indicating the kernel sizes, number of output channels, and dilation rate.

2.5.2. Loss function and labels
The key challenge in this case was the lack of pre-defined labels for training the model, as there is no ideal voxel selection that works for all patients. While fine tuning a heuristic selection scheme, such as outlined in section 2.2 for every patient would in principle be possible, it is not feasible to produce optimized labels for the number of patients required for training the model in a reasonable time. Moreover, it is not clear whether

Figure 1. Schematic description of the workflow, starting from the input generation with TRiP98 entered in the CNN to the calculation of the final dose map. The patient CT, contoured OARs and treatment plan parameters are proceeded into TRiP98. Then, the Euclidian distance transformation is computed, as well as a binary mask of the tumour and a binary mask of the OARs. These are pre-processed and used as input for the CNN model. The *P*-Net-based model (Wang *et al* 2019) outputs a candidate voxel mask, which is post-processed to the original resolution and used by TRiP98 to identify the selected voxels for optimisation. For training (green box), the dosimetric accuracy on all voxels is used to compute a specifically designed loss, which is then backpropagated through the network.

the heuristic chosen really represents the ideal case for the patients. Therefore, a reinforcement learning approach was taken, for which we created quasi-labels from the dosimetric evaluation of the plan optimized based on the candidate voxel sampling produced by the model. We differentiated between target and OAR voxels.

For target voxels, their pseudo-label was set to 1 if the dose coverage in that voxel was less than a user-defined threshold. The threshold was set to 98.5% of the prescription dose for results presented in this manuscript. The loss in this case was then computed as a binary cross entropy loss, forcing the model to give more attention to these voxels, if they were not previously selected. If the voxel coverage fulfilled the dose threshold, its pseudo-label was set to 0 and the loss was calculated as the binary cross-entropy loss multiplied with the ratio of currently selected and unselected voxels. This automatically results in the network reducing the number of selected voxels (SV) overall.

Loss function and quasi-labels in the target can be summarized as:

Algorithm: target pseudo-labels and voxel-wise loss

For i in all voxels in output:
    If i in target:
        if dose $<$ threshold:
            $\text{Label}_i = 1$
            $\text{Loss}_i = \text{CE}$
        else:
            $\text{Label}_i = 0$
            $\text{Loss}_i = \text{SV} * \text{CE}$
    else:
        $\text{Loss}_i = \text{CE}$

---

where SV is the rate of selected voxels in the target, CE is the cross-entropy loss in the voxel

$$CE = -\sum_{c=1}^{C} y_c \log \hat{y}_c,$$

$y -$ label probability, $\hat{y} -$ estimated probability, $C -$ number of classes (2).

The final loss was calculated as the sum losses in the target (T) and the rest of the matrix (R)

$$Loss = \frac{1}{N_T} \sum_{t=1}^{N_T} Loss_t + \frac{1}{N_R} \sum_{r=1}^{N_R} Loss_r,$$

where N is number of voxels, L is loss in the voxel.

For OAR voxels, we evaluated two different setups: we used either the input voxel selection (i.e. the voxels remaining after applying equation (1) as label (Setup 1) or a heuristic sampling procedure with different sampling probabilities for the OAR shell (100%) and interior (12.5%) (Setup 2), except for the chiasma and optical nerves, for which all voxels were used in both setups. The loss was then calculated as the cross-entropy loss. Dosimetric quality in the OARs was not explicitly considered in the training at this stage. Note, that while these OAR voxel samplings could also be used directly, without the need for the deep learning model, including the OARs into the model is anyways needed for adequately selecting target voxels close to critical structures where denser information may be needed. Therefore, the model architecture was kept general, to enable future inclusion of different OAR sampling strategies.

2.5.3. Training and synchronization

A deep-learning training framework was added to TRiP98 to achieve efficient model training. The framework handled the TRiP98 setup prior to optimization, then added in the voxel sampling generated by the current model as a masked 3D array, and then fed back the plan optimization result based on this voxel mask to compute the loss function as outlined above. Overall time efficiency of the plan optimization presented the key bottleneck for training. A major contributing factor is the setup of the dose influence matrix. Even if done only on the selected subset of voxels, recomputing the dose-influence matrix for each new voxel mask would be greatly limiting the training speed. Instead, we computed the complete dose influence matrix for each given patient and stored it internally. For plan optimization only the voxels selected in the sampling were used, which is equivalent to computing the dose influence matrix for only these voxels in the first place. As the full matrix is maintained in memory, subsequent iterations for each patient are much faster than recomputing the dose distribution at every step. Moreover, dose quality metrics such as D95% can be directly computed for all voxels after optimization.

For communication and synchronization of the training and treatment planning processes, a First In—First Out file was created, which becomes accessible only if one party opens the file for reading while the other one opens it for writing and vice versa. Through this file, the AI process provides an 3D binary map indicating voxels to be included in the next dose calculation, while TRiP98 provided the dose results in the form of 3D dose distributions and dose-volume histogram (DVH) metrics after completing this iteration.

The key bottleneck for training the model was the plan optimization procedure in TRiP98. Since TRiP98 is not ported to run on a GPU, the model had to be trained on CPU, for which we used a Linux server with 96 cores and 252 GiB RAM. A batch size of 1 was used for the training process, meaning each patient was processed individually. The reason for this was to increase the training computational performance because it let us process each individual patient multiple times in a row. This minimized the time delay associated with the setup of the dose influence matrix for the patient in TRiP98, while also aiding the system to recognize patient specific patterns. For a similar reason, the model was selectively validated during the training for selected epochs, were the training performed best. For each epoch, the training dataset was shuffled to avoid overfitting. We used the Adam optimizer (Kingma *et al* 2015). As a starting point, the learning rate was set to 10^{-3}. It was reduced to 10^{-4} after 5 epochs and halved with every additional 5 epochs, down to a minimum of 10^{-5}.

Each patient was processed 5 times per epoch, i.e. per epoch 250 treatment plans were generated and evaluated. Overall, the training took approximately four hours per epoch. We trained the network for 50 epochs, i.e. roughly 8 d.

2.5.4. Model evaluation

The treatment plans created with the different voxel sampling techniques were evaluated on all target and OAR voxels for their dosimetric accuracy and robustness against nine scenarios assuming 3 mm isotropic setup uncertainty. An identical robustness analysis on the full resolution CT was performed for the plan

derived on the full mask with all voxels included (called Reference throughout the rest of the manuscript) (Wolf *et al* 2020). We report the outcome for the 10 patients of the test dataset only. To visualize the dependency of the models on the sampling volume, the patients were sorted by tumor volume in all presented results. As the final set of SV involves random sampling, the effect the sampling variability has on the dose was assessed by repeating the sampling, plan generation, and evaluation 10 times for each patient with a different random seed. The results report the median of these dose computations, and the range over all patients as well as the standard deviation for the 10 samplings of each patient is given. In addition, the standard deviation of the robustness scenarios is reported.

The plan quality was evaluated for target coverage (D95%), the conformity number (CN) (Riet *et al* 1997), and homogeneity index (HI). OARs metrics included D0.1cc, i.e. the dose received by 0.1cc volume, for eyes and parotid; D1%, which can be understood as a robust representative of the maximum dose, for optic nerves, chiasma, brainstem, and spinal cord; and V20% for the brainstem and spinal cord. These metrics were chosen following typical metrics of interest in international dose volume constraint recommendations. Both setups were compared by reporting differences to the Reference for all metrics. A gamma index analysis with 3%/2 mm criteria was performed compared to the Reference, excluding those voxels with a dose lower than 10% of the maximum dose. Note, as the objective of the voxel sampling is to generate a clinically acceptable plan, not one that is identical to the Reference plan, a low gamma passing rate does not indicate a low plan quality. Instead, high gamma passing rates further strengthen the validity of a certain voxel sampling strategy and thus provides a further insight in the clinical potential of the models. Finally, a key objective of the voxel sampling is the efficiency benefit for the treatment planning process. The performance of the voxel sampling models was scored by the percentage of SV compared to the Reference, as well as the memory consumption and duration of the optimization under identical conditions on the same Linux server.

3. Results

Figures 2 and 3 show the outcome for the patients from the test dataset where the models performed best (Patient 3) or worst (Patient 1), respectively, as measured by the loss in target coverage. Shown are the dose distributions for the Reference, as well as for the heuristic sampling, and the deep-learning model Setup 1 and Setup 2. For both models, also the voxels selected for optimization are indicated. Plots of the relative difference between approaches are provided as figures S1 and S2 within the supplementary material to this article.
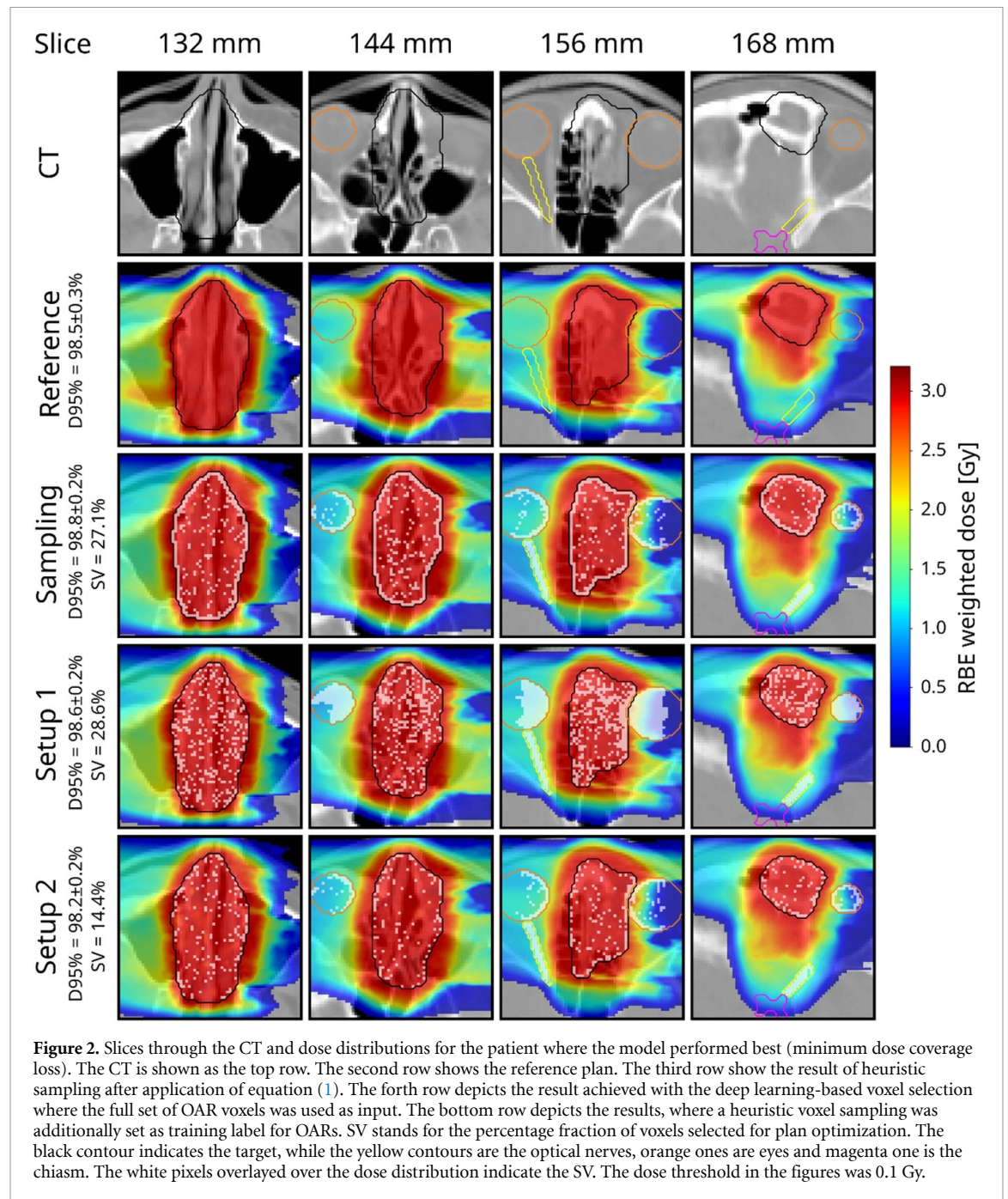
Patient 3 (Adenoid cystic carcinoma) had a median ΔD95% of $-0{,}4$% for Sampling, -0.03% for Setup 1 and 0.3% for Setup 2 at an SV of 27.1%, 28.6% and 14.4%, respectively. In this case, hardly any dose difference is visible, despite the drastically reduced number of SV.

Patient 1 (Chondrosarcoma) had a median ΔD95% of 3.3% in the target for heuristic sampling, 3.7% for Setup 1 and 5.4% for Setup 2 and an SV of 28.2%, 24.0% and 14.6%, respectively. Close to the optical nerve, the target was underdosed. This is likely related to the close vicinity of this OAR to the target, and the comparatively large number of voxels corresponding to the OAR relative to those of the target volume in that region.

In these examples, it can be observed qualitatively that Setup 1 achieved overall better dosimetric outcomes than Setup 2, but at a larger fraction of SV. The worst performing case was the patient with the smallest tumor volume. For this patient, dose coverage was the lowest among the test dataset (D95% = 97.2%) and both voxel sampling setups as well as heuristic sampling did not achieve the clinical coverage criteria of D95% larger than 95% of the prescribed dose.

There were additional differences in the lateral penumbra of the irradiation fields, that can be seen in the supplementary material figures S1 and S2. Both the heuristic and deep-learning voxel sampling methods led to a broader irradiation field in the shown patient cases, resulting in additional dose in the field periphery, while the central region in the entrance channel received less dose. An explicit avoidance ring around the target can in principle be used to reduce this effect during plan optimization, if deemed clinically necessary. For the deep-learning voxel sampling, this could be included also as additional objective during training, but this was beyond the scope of this work.

Table 2 and figure 4 present detailed dosimetric analyses for the entire test dataset. Values preceded by Δ represent the difference of Reference—heuristic sampling, Setup 1 or 2. For both Setup 1 and 2, median dose coverage (D95%) was reduced in comparison to Reference, by 0.5% and 0.9%, respectively. Both setups also reduced the number of selected target voxels in the optimization, by a factor of 5 and 8, respectively. Kolmogorov–Smirnov test shows the similarity of the distribution of dose coverage for heuristic sampling and Setup 1 ($D = 0.3$, *p*-value = 0.7591), while Setup 1 include 5.35% less voxels in the target. The gamma

Figure 2. Slices through the CT and dose distributions for the patient where the model performed best (minimum dose coverage loss). The CT is shown as the top row. The second row shows the reference plan. The third row show the result of heuristic sampling after application of equation (1). The forth row depicts the result achieved with the deep learning-based voxel selection where the full set of OAR voxels was used as input. The bottom row depicts the results, where a heuristic voxel sampling was additionally set as training label for OARs. SV stands for the percentage fraction of voxels selected for plan optimization. The black contour indicates the target, while the yellow contours are the optical nerves, orange ones are eyes and magenta one is the chiasm. The white pixels overlayed over the dose distribution indicate the SV. The dose threshold in the figures was 0.1 Gy.
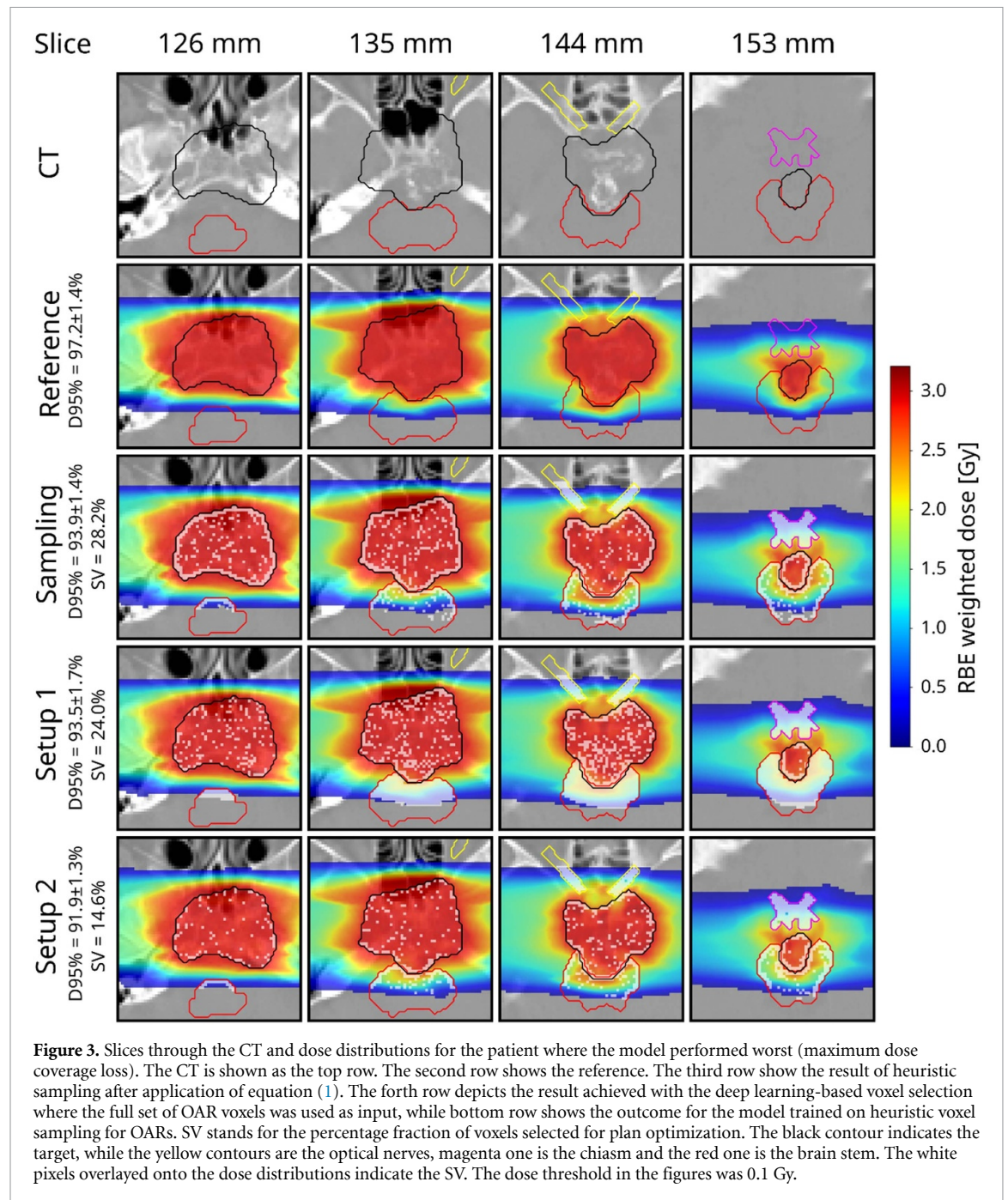
passing rate (3%/2 mm) compared to Reference was 89.4% for heuristic sampling, 93.2% for Setup 1 and 86.8% for Setup 2.

Both setups showed a similar trend of the fraction of SV (figure 4(a)), while Setup 2 overall resulted in a smaller fraction of SV. For larger targets (patients 3 through 10), the dose coverage (figure 4(b)) was comparable between the two setups and Reference, but was clearly reduced for smaller targets, even more so for Setup 2. Except Patient 3 and 4 both setups included less voxels in the target in comparison to heuristic sampling.

For most patients, the voxel sampling plans yielded a worse CN compared to Reference, but better compared to heuristic sampling (figure 4(c)) and reduced target dose homogeneity (figure 4(d)), although it remained within acceptable limits. Overall, Setup 2 produced slightly lower quality plans in all metrics compared to Setup 1. In terms of mean dose coverage, the minimum reduction with Setup 2 was 0.1% (Patient 5) and the maximum reduction was 1.7% (Patient 1) compared to Setup 1.

The dosimetric analysis of OARs is presented in figure S3 and shows no considerable difference with respect to the reference values of the full mask for most OARs. Table S1 reports the median value and the range of the mean values in the OARs for each setup. As expected, the outcomes depended highly on the

Figure 3. Slices through the CT and dose distributions for the patient where the model performed worst (maximum dose coverage loss). The CT is shown as the top row. The second row shows the reference. The third row show the result of heuristic sampling after application of equation (1). The forth row depicts the result achieved with the deep learning-based voxel selection where the full set of OAR voxels was used as input, while bottom row shows the outcome for the model trained on heuristic voxel sampling for OARs. SV stands for the percentage fraction of voxels selected for plan optimization. The black contour indicates the target, while the yellow contours are the optical nerves, magenta one is the chiasm and the red one is the brain stem. The white pixels overlayed onto the dose distributions indicate the SV. The dose threshold in the figures was 0.1 Gy.

location of the OAR with respect to the target. In addition, the difference between the voxel sampling and the reference plans shows a strong patient specific variation for most OAR's.

Figure 5 shows the memory requirement and duration of the optimization of the patient in the test dataset. Computations were performed on a Linux server using 10 threads (more information can be found in figure S4). Especially for larger tumors, the benefit in memory requirement and runtime becomes apparent. The median memory requirement for the full voxel setup was 15.6 GiB (14.8 GiB after applying equation (1) for the patients from the test dataset, while it was only 4.0 GiB for heuristic sampling, 4.3 GiB for Setup 1 and 2.3 GiB for Setup 2 (table S2). When running with 10 threads, this translated to median plan optimization times of 3 min, 40, 49 and 24 s for Reference, heuristic sampling, Setup 1 and Setup 2, respectively. Setup 2 outperformed Setup 1 as expected from the reduced number of voxels included in the optimization. The runtime per iteration in the optimization is linearly correlated with the required memory (figure S5).

Equation (1) does not show any significant effect applying Kolmogorov-Smirnov test on Reference or heuristic sampling for both nominal scenario ($D = 0.2$, *p*-value $= 0.9883$ for both) and uncertainties

Table 2. Target dose metrics achieved with the deep-learning voxel selection and their comparison to the reference plan. Standard deviations from different generated masks are reported in parenthesis. Values are given for the nominal scenario, where stated, 'uncertainty' reflects the standard deviation from the robustness evaluation with nine uncertainty scenarios. Arrows indicate whether higher values (↑) or lower ones (↓) are desired. SV—percentage of the selected voxels in the target; HI—homogeneity index; CN -conformation number.

Criteria	Type	Sampling Median	Sampling Min–max	Setup 1 Median	Setup 1 Min–max	Setup 2 Median	Setup 2 Min–max
D95% [%]	Nominal↑	98.3	94.0–98.8	98.2 (0.06)	93.5–98.6 (0.03–0.30)	97.6 (0.05)	91.8–98.3 (0.02–0.46)
	Uncertainty	0.29	0.14–1.14	0.47 (0.03)	0.14–1.77 (0.01–0.06)	0.40 (0.03)	0.09–1.47 (0.01–0.12)
	Δ	0.32	−0.03–3.3	0.5	−0.03–3.67	0.9	0.33–5.33
Selected voxels	Nominal↓	6523	4748-13856	7663(52)	5621-13626 (33–83)	3821 (42)	3079-7355 $(24–92)10^{-3}$
SV in the target [%]	Nominal↓	25.65	22.91–28.22	20.30 (0.29)	17.36–28.12 (0.15–0.43)	12.51 (0.22)	10.32–14.49 (0.13–0.34)
SV [%]	Nominal↓	25.42	23.17–30.80	28.40 (0.19)	23.34–40.06 (0.13–0.26)	**15.18** (0.16)	12.60–21.09 (0.09–0.24)
HI [%]	Nominal↓	2.80	2.40–7.94	3.26 (0.09)	2.71–8.75 (0.06–0.47)	3.95 (0.11)	3.12–10.51 (0.04–0.64)
	Uncertainty	0.26	0.13–1.41	0.40 (0.05)	0.12–1.60 (0.02–0.30)	0.37 (0.04)	0.10–1.32 (0.01–0.29)
	Δ	−0.38	−3.44–0.56	−0.42	−4.39–1.84	−0.89	−6.16–0.62
CN	Nominal↑	68.54	64.34–71.50	72.25 (0.29)	67.46–74.63 (0.16–0.71)	69.47 (0.36)	66.80–72.95 (0.19–0.52)
	Uncertainty	0.51	0.25–0.96	0.46 (0.03)	0.30–1.15 (0.02–0.07)	0.60 (0.04)	0.36–0.93 (0.02–0.09)
	Δ	4.79	3.01–6.93	1.30	0.35–2.69	3.14	1.01–5.01
Gamma (3%/2 mm)	Nominal↑	89.4	43.6–97.8	93.2 (1.1)	63.1–98.0 (0.4–2.4)	86.8 (1.1)	54.1–95.1 (0.7–3.5)

($D = 0.1$, *p*-value $= 1$ for both) reducing the number of OAR voxels by 11% on median out of the full number of voxels in the test dataset.
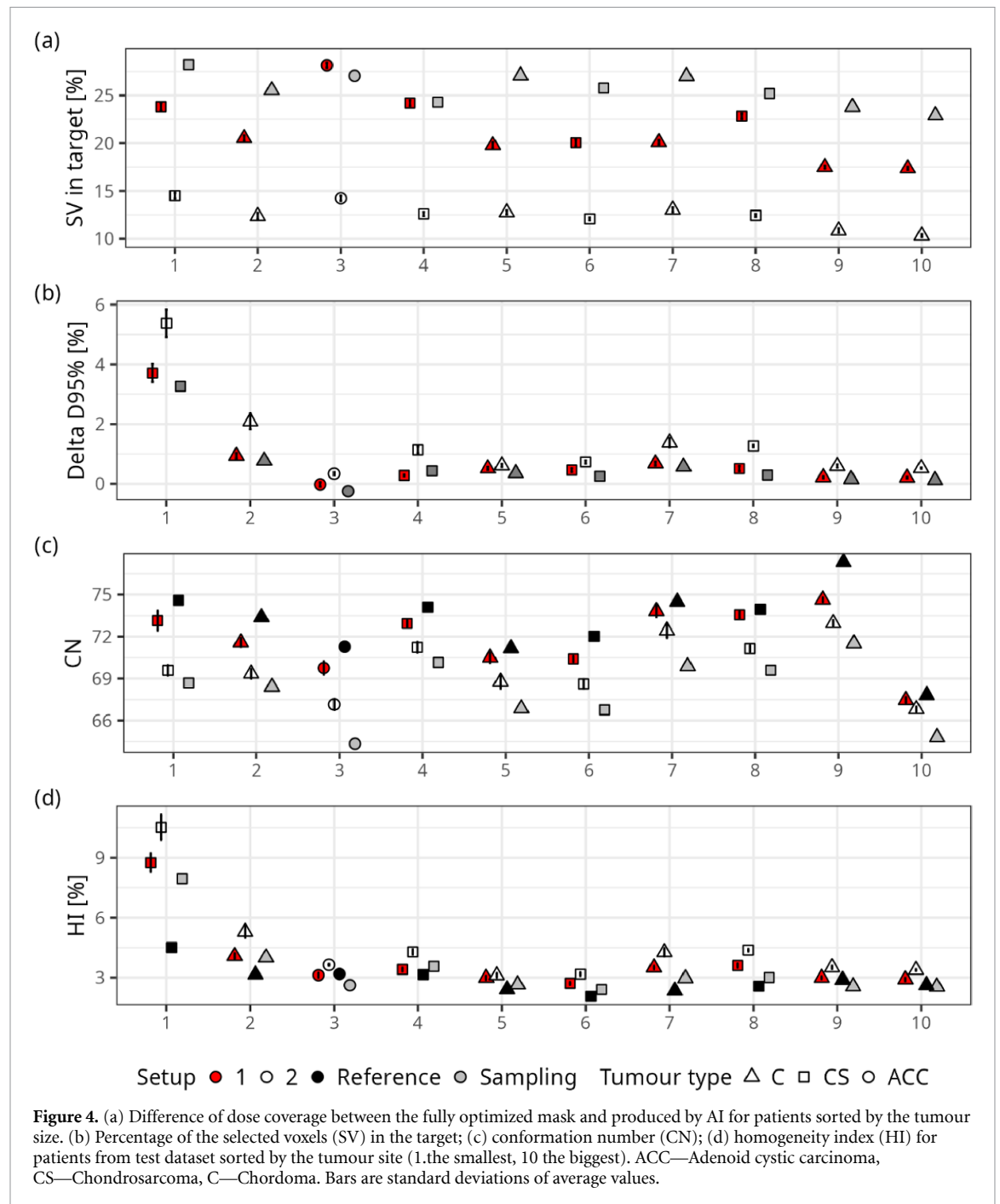
4. Discussion

In this work, we investigated different options for voxel sampling in carbon ion therapy. Two general approaches were tested: randomised voxel selection based on a heuristic understanding of the optimisation process, and a deep-learning model, which was trained on a set of 50 head and neck cancer patients. In general, both approaches performed well in terms of reducing the problem size of carbon ion therapy optimization, which is relevant in the context of online and especially future real-time adaptive therapy workflows. The general optimization problem in carbon ion therapy is complicated over that encountered in proton therapy, due to the need to consider the carbon ions' variable RBE along their depth-dose profile. The associated computational load of the plan optimization process hinders access to many methods that are now commonly available with proton therapy treatment planning, such as the generation of pareto-optimal plans. Reducing the computational load is a key necessity to enable advanced treatment planning features.

The heuristic voxel selection strategy was based on the assumption that voxels corresponding to the boundary region of the organ, and those in the interior need to be sampled differently for best results. In the interior of the volume, a coarser sampling rate can be chosen, utilizing the overlap of neighbouring pencil beams. At the organ boundary, a finer sampling is needed, to accommodate the desired strong dose gradients to the surrounding healthy tissue. The three parameters of this approach (boundary width, boundary sampling rate, and interior sampling rate) need to be user defined. In practice, it is often the best approach to choose a conservative sampling rate to avoid dosimetric issues and avoid painstaking parameter optimization. Different options for voxel selection can also be envisioned, such as choosing a functional dependence of the sampling rate with the distance transformation, instead of the here used separation into boundary and interior. However, finding the right function for the patient is also time consuming and quickly becomes complicated when including other conditions such as nearby OARs.
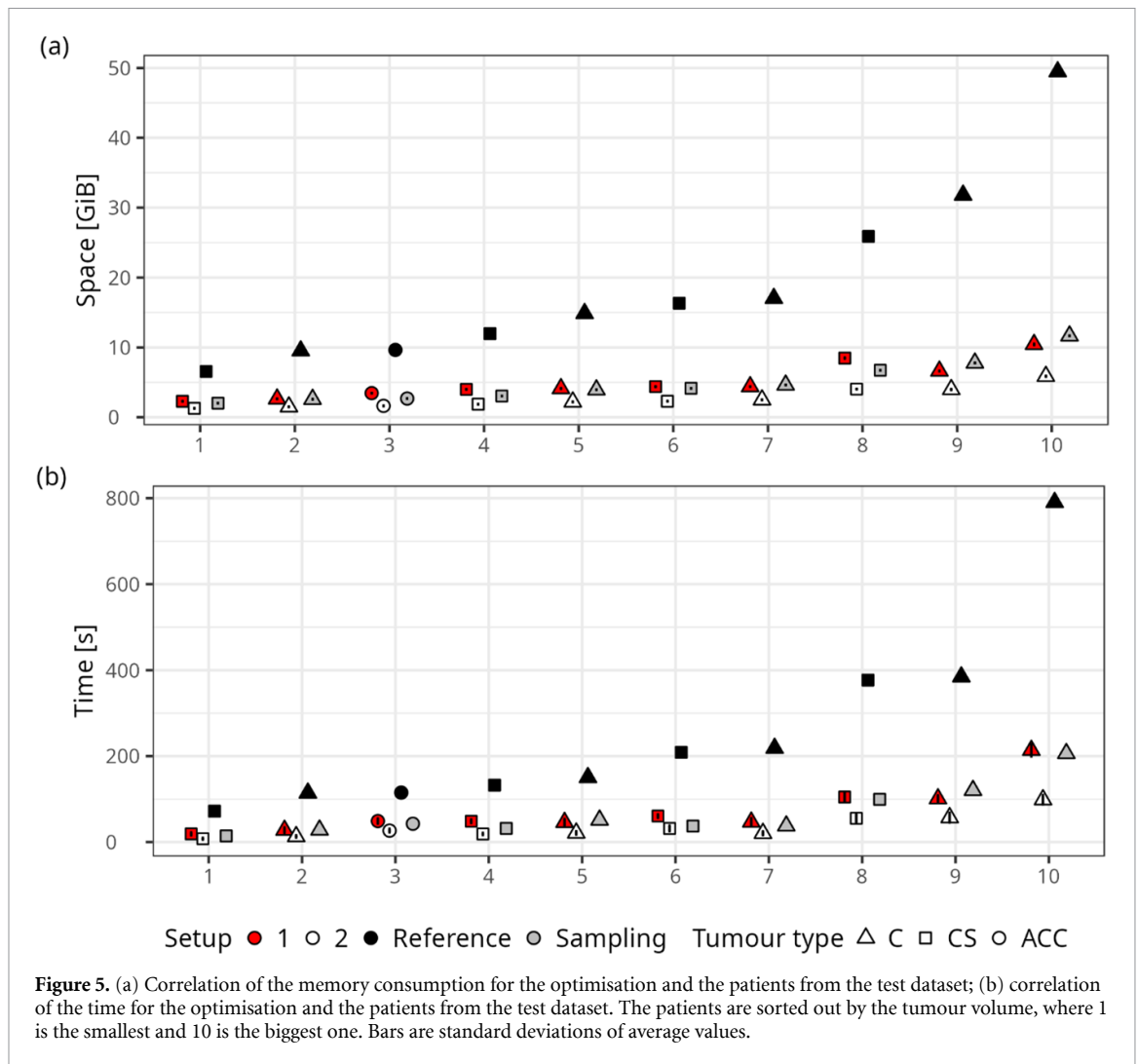
To overcome this issue, we developed a bespoke deep-learning algorithm. Our deep learning voxel sampling algorithm gave excellent results in selecting the voxels important for plan optimization. The basic underlying concept is that the dose of a single pencil beam to neighboring voxels is not independent but correlated with the physical and biological beam characteristics. Hence, it is sufficient to rely on only a subset of the voxels for optimization. This is particularly the case for robust optimization, where the consideration of setup uncertainties results in an oversampling of the voxels anyways.

This work has demonstrated how much the number of voxels can be reduced. With a median of only approximately 15% of the voxels used in the optimization (Setup 2), a median loss of less than 1% in target coverage was observed. Even in this low sampling case, the dose homogeneity suffered by less than 1%. With

Figure 4. (a) Difference of dose coverage between the fully optimized mask and produced by AI for patients sorted by the tumour size. (b) Percentage of the selected voxels (SV) in the target; (c) conformation number (CN); (d) homogeneity index (HI) for patients from test dataset sorted by the tumour site (1.the smallest, 10 the biggest). ACC—Adenoid cystic carcinoma, CS—Chondrosarcoma, C—Chordoma. Bars are standard deviations of average values.

Setup 1, i.e. when the model was trained to include all relevant OAR voxels, the median number of SV was approximately 28% at little to no loss in plan quality with respect to the chosen dose objectives. The method worked best for larger targets, where it also yielded the most computational benefit. For smaller targets (patient 1 and 2) plan quality suffered, at small absolute gain in computation times and problem size. While a low gamma passing rate was present in some cases, with a minimum of 43.6% for the heuristic sampling, indicating meaningful dose distributions differences between Reference and sampling plans, the sampling plans were still acceptable with regards to the investigated DVH metrics. Overall, the deep learning-based approach demonstrated high potential for reducing the computational demand at comparable plan quality to the Reference plans.

Contrary to increasing the dose voxel size, which can lead to issues in the treatment plan (Espãa and Paganetti 2011), the deep learning voxel sampling keeps the original resolution of the CT for planning. Large planning voxels also counteract the effort to reduce margins to a minimum. Moreover, no consensus on the best voxel size for different target sites exists. Both the deep learning voxel sampling and voxel size increase can be used together in principle, but the model would need to be retrained with the desired voxel size.

Figure 5. (a) Correlation of the memory consumption for the optimisation and the patients from the test dataset; (b) correlation of the time for the optimisation and the patients from the test dataset. The patients are sorted out by the tumour volume, where 1 is the smallest and 10 is the biggest one. Bars are standard deviations of average values.

The efficiency increase opens possibilities for advanced carbon ion therapy treatment planning strategies. Especially for 4D treatment planning used to mitigate target motion related dose uncertainties in lung cancer, and robust optimization, the size of the dose influence matrix becomes an issue, exceeding 100 GB in some cases. This is also limiting for investigating new planning strategies, like the derivation of Pareto-optimal plans, which has long been possible for proton (Hong *et al* 2008) and photon (Chen *et al* 2010) therapy. However, these advanced planning strategies also bring new challenges with them to be included in the deep-learning model training, such as, for example, the inclusion of not only multiple fields, but also multiple breathing phases between which the target and OAR geometry changes in the case of 4D optimization strategies. This will be subject for future studies within our group.

Voxel sampling is the key to bringing these advanced treatment planning techniques into fruition. Furthermore, it will aid in enabling fast online treatment adaptation for carbon ion therapy. Weekly or possibly even daily treatment adaptation is often needed to ensure the best therapy, but workflow efficiency is still a major bottleneck (Jia *et al* 2022). Moreover, the efficiency increase achieved with deep learning voxel sampling would enable to run many planning processes in parallel, providing an improved platform for developing machine learning methods. For complex models, treatment planning needs to be repeated many thousand times, like in this work, such that any time efficiency improvement can make a major difference.

The deep learning voxel sampling yielded denser sampling close to the target boundary compared to the interior, which was also expected from empirical considerations. In the center of the target, where the dose should be homogeneous, the correlation between neighboring voxels in the optimization objective function is comparably larger than at the boundary of the target, where strong dose gradients occur. This should enable to remove more voxels in the interior compared to the boundary, as confirmed by the deep learning inference in this work. This work therefore also validates heuristic approaches that assign different sampling probabilities to the interior and shell. Nevertheless, these approaches would still require manual input, limiting their use in clinical practice. While in principle, the deep-learning results could be used to find such

a heuristic, it is simpler to directly apply the model itself. A further aspect that can influence the sampling-rate of the deep-learning voxel sampling method is given by RBE gradients within the target region. Heterogeneities in the RBE (or the underlying dose-averaged linear energy transfer, LETd) are likely require denser sampling for better results compared to regions with homogeneous RBE distribution. This may be particularly important for LETd optimization or particle arc techniques, where stronger RBE/LETd gradients in the target are expected, compared to the two-field approach assumed in this work.

The model generality presents a central aspect of this work. While developed within a research treatment planning platform, the deep-learning voxel sampling can easily be incorporated into other treatment planning platforms. The model can also accept different input sizes for retraining, unlike *U*-Net (Ronneberger *et al* 2015) or *V*-Net (Milletari *et al* 2016) where the input size depends on the number of blocks because of the application of max-pooling.

Still the training of the model needs to be site-specific and is time consuming due to the generation of many treatment plans. The voxel reduction problem is highly complex, especially due to the lack of suitable training labels. The classical reinforcement learning approach requires many combinations before learning the main principles of the reward system. That was not possible due to the time needed for optimization and evaluation of each mask, such that we opted for a learning approach with a specifically designed loss mechanism with pseudo-labels.

The loss is mainly regulated by the chosen dose threshold. As the loss automatically enforces to increase the voxel selection probability in regions that were underdosed compared to the threshold, the threshold steers the trade-off between plan quality and voxel reduction. The second regulator is the model independent variable (in our case the fraction of selected voxels in the target), which slows down the process of voxel deselection proportionally to the fraction of SV in the target. In other words, the reduction is strong when the fraction of voxel is high, but when the residual fraction is small, the reduction becomes slower. By choosing another regulator, the user can define the speed of voxel exclusion. Without that step, the system tends to get lower dose coverage in the target than the current model adopted.

The bottleneck of the training is the need for continuous recalculation of the dose maps for each produced mask in combination with contradictory objectives: voxel reduction and plan quality. At the same time, it allows to include multiple criteria into the loss function to achieve wished result. Here we decided to use only the target dose coverage.

This work presents the first demonstration of applying deep learning for problem size reduction in particle therapy. Further work should consider including a regulator of overdosed regions, especially those located close to OARs. The OARs' constraints should be included as well, and they should be either predefined for each organ or be set as additional input. In addition, removing voxels from any VOI results in the reduction of this VOIs relative weight in the optimization cost function. This can impact the treatment plan quality as seen for the smaller tumor sites (patient 1 and 2). The close vicinity to the optical nerve, and the optical nerve's increased relative importance for optimizing beam spots in that region results in a compromise in target coverage and better sparing of the optical nerve. In future works, adaptive weight strategies may yield even better results than those presented in this work.

The treatment plans presented in this work have not been experimentally verified. However, all forward dose calculations reported here were computed with the unchanged dose engine of TRiP98, which has been extensively validated since 1997, both experimentally, for the last time here (Steinsberger *et al* 2023), and against clinical TPS (Sheng *et al* 2023). All plans fulfilled clinical delivery criteria, such as minimum particles per spot.

5. Conclusions

We have developed a new AI algorithm for reducing the number of voxels used in the calculation of the optimization cost function for particle therapy. The model is based on *P*-Net and exploits quasi-labels and a tailored loss function. Depending on the fraction of voxels in OARs, the model reduced the number of voxels by up to 82%. This correlates to a considerable advantage in memory and time consumption. Deep learning-based voxel sampling presents itself as a prime candidate to bring advanced particle therapy treatment planning procedures into clinical practice, especially also in the context of adaptive therapy. Future work will further improve the model and investigate different treatment sites that expand the model to 4D-applications, like lung cancer.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://git.gsi.de/medphys/public/ai-sampling .

Acknowledgments

Code availability

The code is publicly available at https://git.gsi.de/medphys/public/ai-sampling

ORCID iDs

A Quarz ⬤ https://orcid.org/0009-0001-9994-4348
L Volz ⬤ https://orcid.org/0000-0003-0441-4350
C Hoog Antink ⬤ https://orcid.org/0000-0001-7948-8181
M Durante ⬤ https://orcid.org/0000-0002-4615-553X
C Graeff ⬤ https://orcid.org/0000-0002-5296-7649

References

Abadi M, Barham P, Chen J, Chen Z, Davis A and Dean J 2016 TensorFlow: a system for large-scale machine learning *Proc. 12th USENIX Symp. on Operating Systems Design and Implementation (OSDI 16)*

Bobic M, Lalonde A, Sharp G C, Grassberger C, Verburg J M, Winey B A, Lomax A J and Paganetti H 2021 Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy *Phys. Med. Biol.* **66** 055023

Chen W, Craft D, Madden T M, Zhang K, Kooy H M and Herman G T 2010 A fast optimization algorithm for multicriteria intensity modulated proton therapy planning *Med. Phys.* **37** 4938–45

Dillon J V, Langmore I, Tran D, Brevdo E, Vasudevan S and Moore D 2017 TensorFlow distributions (arXiv:1711.10604)

Durante M and Paganetti H 2016 Nuclear physics in particle therapy: a review *Rep. Prog. Phys.* **79** 096702

Espãa S and Paganetti H 2011 Uncertainties in planned dose due to the limited voxel size of the planning CT when treating lung tumors with proton therapy *Phys. Med. Biol.* **56** 3843

Furuta R, Inoue N and Yamasaki T 2020 PixelRL: fully convolutional network with reinforcement learning for image processing *IEEE Trans. Multimedia* **22** 1704–19

Graeff C 2014 Motion mitigation in scanned ion beam therapy through 4D-optimization *Phys. Medica* **30** 570–7

Graeff C, Volz L and Durante M 2023 Emerging technologies for cancer therapy using accelerated particles *Prog. Part. Nucl. Phys.* **131** 104046

Hild S, Graeff C, Trautmann J, Kraemer M, Zink K, Durante M and Bert C 2014 Fast optimization and dose calculation in scanned ion beam therapy *Med. Phys.* **41** 071703

Hong T S, Craft D L, Carlsson F and Bortfeld T R 2008 Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head *Int. J. Radiat. Oncol.* **72** 1208–14

Huynh E, Hosny A, Guthier C, Bitterman D S, Petit S F, Haas-Kogan D A, Kann B, Aerts H J W L and Mak R H 2020 Artificial intelligence in radiation oncology *Nat. Rev. Clin. Oncol.* **17** 771–81

Jia S, Chen J, Ma N, Zhao J, Mao J, Jiang G, Lu J and Wu K 2022 Adaptive carbon ion radiotherapy for locally advanced non-small cell lung cancer: organ-sparing potential and target coverage *Med. Phys.* **49** 3980–9

Kingma D P and Ba J L. 2015 Adam: a method for stochastic optimization *3rd Int. Conf. on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conf. Track Proc.*

Krämer M and Scholz M 2006 Rapid calculation of biological effects in ion radiotherapy *Phys. Med. Biol.* **51** 1959–70

Liao X, Li W, Xu Q, Wang X, Jin B and Zhang X 2020 Iteratively-refined interactive 3D medical image segmentation with multi-agent reinforcement learning *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (https://doi.org/10.1109/CVPR42600.2020.00941)

Lis M, Newhauser W, Donetti M, Durante M, Weber U, Zipfel B, Hartmann-Sauter C, Wolf M and Graeff C 2021 A facility for the research, development, and translation of advanced technologies for ion-beam therapies *J. Instrum.* **16** T03004

Lu X Q and Chin L M 1993 Ampling techniques for the evaluation of treatment plans *Med. Phys.* **20** 151–61

Ma C, Xu Q, Wang X, Jin B, Zhang X, Wang Y and Zhang Y 2021 Boundary-aware supervoxel-level iteratively refined interactive 3D image segmentation with multi-agent reinforcement learning *IEEE Trans. Med. Imaging* **40** 2563–74

Martin B C 2007 Accelerating IMRT optimization by voxel sampling *Phys. Med. Biol.* **52** 7211

Maurer C R, Qi R and Raghavan V 2003 A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions *IEEE Trans. Pattern Anal. Mach. Intell.* **25** 265–70

Mein S, Tessonnier T, Kopp B, Schömers C, Harrabi S, Abdollahi A, Debus J, Haberer T and Mairani A 2022 Biological dose optimization for particle arc therapy using helium and carbon ions *Int. J. Radiat. Oncol. Biol. Phys.* **114** 334–48

Milletari F, Navab N and Ahmadi S A 2016 V-Net: fully convolutional neural networks for volumetric medical image segmentation *Proc.—2016 4th Int. Conf. on 3D Vision (3DV)* (https://doi.org/10.1109/3DV.2016.79)

Paganetti H, Botas P, Sharp G C and Winey B 2021 Adaptive proton therapy *Phys. Med. Biol.* **66** 22TR01

Pompos A, Foote R L, Koong A C, Le Q T, Mohan R, Paganetti H and Choy H 2022 National effort to re-establish heavy ion cancer therapy in the United States *Front. Oncol.* **12** 880712

(PTCOG) PTC-OG 2023 Particle therapy facilities in operation (available at: www.ptcog.site/) (Accessed 04 May 2023)

Riet A V T, Mak A C A, Moerland M A, Elders L H and Van Der Zee W 1997 A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate *Int. J. Radiat. Oncol. Biol. Phys.* **37** 731–6

Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Lecture Notes in Computer Science* vol 9351 pp 234–41

Sheng Y, Volz L, Wang W, Durante M and Graeff C 2023 Evaluation of proton and carbon ion beam models in TReatment planning for particles 4D (TRiP4D) referring to a commercial treatment planning system *Z. Med. Phys.* (https://doi.org/10.1016/j.zemedi.2023.06.002)

Steinsberger T, Donetti M, Lis M, Volz L, Wolf M, Durante M and Graeff C 2023 Experimental validation of a real-time adaptive 4D-Optimized particle radiotherapy approach to treat irregularly moving tumors *Int. J. Radiat. Oncol. Biol. Phys.* **115** 1257–68

Tinganelli W and Durante M 2020 Carbon ion radiobiology *Cancers* **12** 3022

Wang G *et al* 2019 DeepIGeoS: a deep interactive geodesic framework for medical image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 1559–72

Wolf M E 2019 Robust optimization in 4D treatment planning for carbon ion therapy of lung tumors *Dissertation* TU Darmstadt

Wolf M, Anderle K, Durante M and Graeff C 2020 Robust treatment planning with 4D intensity modulated carbon ion therapy for multiple targets in stage IV non-small cell lung cancer *Phys. Med. Biol.* **65** 215012

Wolf M, Anderle K, Prall M and Graeff C 2019 Random voxel subsampling enables large scale robust optimization in heavy ion therapy *Annual Meeting of the German Medical Physics Society* (*Stuttgart, 2019*) pp 252–3 (available at: www.dgmp.de/de-DE/131/dgmp-tagungsbaende/)

Yu F and Koltun V 2016 Multi-scale context aggregation by dilated convolutions *4th Int. Conf. on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conf. Track Proc.*