Radiotherapy and Oncology 191 (2024) 110056

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Uncertainty-aware MR-based CT synthesis for robust proton therapy planning of brain tumour

Xia Li [a,b], Renato Bellotti [a,c], Gabriel Meier [a], Barbara Bachtiary [a], Damien Weber [a,d,e], Antony Lomax [a,c], Joachim Buhmann [b], Ye Zhang [a,*]

[a] Center for Proton Therapy, Paul Scherrer Institut, Switzerland
[b] Department of Computer Science, ETH Zurich, Switzerland
[c] Department of Physics, ETH Zurich, Switzerland
[d] Department of Radiation Oncology, University Hospital of Zurich, Switzerland
[e] Department of Radiation Oncology, Inselspital, Bern University Hospital, University of Bern, Switzerland

ARTICLE INFO

Keywords:
MR-based CT synthesis
Uncertainty estimation
Robust planning
Proton therapy
Brain tumours

ABSTRACT

Background and purpose: Deep learning techniques excel in MR-based CT synthesis, but missing uncertainty prediction limits its clinical use in proton therapy. We developed an uncertainty-aware framework and evaluated its efficiency in robust proton planning.

Materials and methods: A conditional generative-adversarial network was trained on 64 brain tumour patients with paired MR-CT images to generate synthetic CTs (sCT) from combined T1-T2 MRs of three orthogonal planes. A Bayesian neural network predicts Laplacian distributions for all voxels with parameters (μ, b). A robust proton plan was optimized using three sCTs of μ and $\mu \pm b$. The dosimetric differences between the plan from sCT (sPlan) and the recalculated plan (rPlan) on planning CT (pCT) were quantified for each patient. The uncertainty-aware robust plan was compared to conventional robust (global $\pm$ 3 %) and non-robust plans.

Results: In 8-fold cross-validation, sCT-pCT image differences (Mean-Absolute-Error) were 80.84 $\pm$ 9.84HU (body), 35.78 $\pm$ 6.07HU (soft tissues) and 221.88 $\pm$ 31.69HU (bones), with Dice scores of 90.33 $\pm$ 2.43 %, 95.13 $\pm$ 0.80 %, and 85.53 $\pm$ 4.16 %, respectively. The uncertainty distribution positively correlated with absolute prediction error (Correlation Coefficient: 0.62 $\pm$ 0.01). The uncertainty-conditioned robust optimisation improved the rPlan-sPlan agreement, e.g., D95 absolute difference (CTV) was 1.10 $\pm$ 1.24 % compared to conventional (1.64 $\pm$ 2.71 %) and non-robust (2.08 $\pm$ 2.96 %) optimisation. This trend was consistent across all target and organs-at-risk indexes.

Conclusion: The enhanced framework incorporates 3D uncertainty prediction and generates high-quality sCTs from MR images. The framework also facilitates conditioned robust optimisation, bolstering proton plan robustness against network prediction errors. The innovative feature of uncertainty visualisation and robust analyses contribute to evaluating sCT clinical utility for individual patients.

Introduction

Proton therapy, a state-of-the-art technology in radiotherapy, is garnering increasing attention [1]. Its exceptional characteristic of high-dose conformity underscores the critical importance of accurately representing the patient's geometry in proton therapy planning. Consequently, obtaining the most up-to-date geometric model of the patient through frequent daily imaging is essential to harness the full potential of proton physics. Meanwhile, MR-based radiotherapy has emerged as an auspicious approach, marked by the commercialization of the MR-Linac technology and its active clinical implementation. This innovative method not only eliminates the need for MR to CT registration [3,4,5], but also significantly reduces the additional radiation exposure associated with more repeated CT scans [6]. These advantages are particularly important for vulnerable populations like children and in the context of daily adaptive proton therapy [2].

Enabling MR-based treatment planning requires accurate CT-like data generation from MRI geometry [7,8]. This can be traditionally

* Corresponding author at: Center for Proton Therapy, Forschungsstrasse 111, WMSA C27, 5232 Villigen, PSI West.
E-mail address: ye.zhang@psi.ch (Y. Zhang).

https://doi.org/10.1016/j.radonc.2023.110056
Received 1 May 2023; Received in revised form 6 December 2023; Accepted 8 December 2023
Available online 15 December 2023

achieved through atlas-based methods [9,10], which initially segmented MRI voxels into distinct tissue regions and subsequently assigned predefined HU values to each region [10]. The atlas-based method [9] involved registering atlas-MRIs to new MR images and warping the atlas CTs with displacement vector field (DVF), which depends highly on the accuracy of the deformable registration result [11]. In the new era of artificial intelligence, deep learning (DL) has become the predominant approach in computer vision and pattern recognition [12]. Synthetic CT generation based on deep learning has also emerged as a popular research topic [13,14]. By leveraging their exceptional ability to extract informative features from input images, deep neural networks have achieved remarkable results in MR-based CT synthesis tasks [7]. Various network architectures have been proposed to learn voxel mapping from MR intensity to CT Hounsfield Units [15–21], and several works have also explored the incorporation of synthesised CT into the workflow of proton therapy [19–25] or carbon ion therapy [26]. Due to the large size of full-resolution CTs, it was often infeasible to feed the entire 3D image into a single neural network. Consequently, different strategies have been employed to split the volume by overlapping or non-overlapping 2D patches, 2D slices, 2.5D slices or 3D patches [27], which were then transformed by the networks individually and subsequently merged to achieve the final estimation.

For well-aligned MR-CT pairs, conditional generative adversarial networks (cGANs) [15] have achieved significantly lower image-level errors than conventional methods [9], but obtaining large-scale paired images remains to be challenging. Furthermore, registration methods exhibit limited performance, hindering the flawless registration of cross-modality images [28,29]. In such scenarios, CycleGAN-based CT synthesis was proposed [16], demonstrating superior image quality compared to Pix2pix methods when paired images are unavailable [17,18]. These approaches yielded satisfactory results at both the image-level and the dosimetric-level, highlighting the immense potential of MR-only treatment planning [8,30]. Nevertheless, as data-driven methods, DL-based approaches can experience substantial performance declines when applied to new data with significant distribution changes compared to the training dataset [31]. Therefore, the ability to automatically detect these failures and estimate the associated uncertainty of both data and network was considered compulsory for the clinical use of these methods [32], particularly for proton therapy applications.

In this paper, we present a novel uncertainty-aware MR-based proton therapy framework. As shown in Fig. 1, it consists of the MR-to-CT synthesis neural network with uncertainty estimation and the uncertainty-conditioned robust plan optimisation. The framework can not only predict a high-quality synthetic CT (sCT) but also estimate the voxel-wise uncertainty for the predicted Hounsfield Unit (HU). Subsequently, the predicted uncertainty can be incorporated into the proton robust optimisation process [33]. Furthermore, this framework offers both image-level and dosimetric-level robustness analysis, assisting physicians in making decisions on the clinical usefulness of generated sCT. The structure of this paper follows the guidelines for AI in medical physics [34].

Materials and methods

Image dataset

The dataset utilised in this paper comprises images from a population of patients with brain tumours previously treated at PSI from 2017 to 2020. Paired scans from 74 adult (n = 29) and pediatric (<18 years; n = 35) patients were collected, among which 64 cases were used for cross-validation, and 10 were reserved as hold-out data for testing. Before treatment planning, each patient underwent at least one MRI and CT scan. Of note, there were pronounced anatomic differences between the image acquisition processes, as patients were normally fixated using a bite block for CT imaging but not for MR imaging. CT acquisitions were carried out with a Siemens Sensation Open CT scanner (with tube voltage 120 kV) with resolutions of $1 \times 1 \times 2$ mm or $1 \times 1 \times 3$ mm. MR images were acquired using 1.5 T Siemens Aera and Siemens Skyra MR scanners, with voxel size $1 \times 1 \times 1$ mm. T1-weighted MP-RAGE and T2-weighted FLAIR sequences were used for network training. As preprocessing, MR images were registered to the corresponding CT

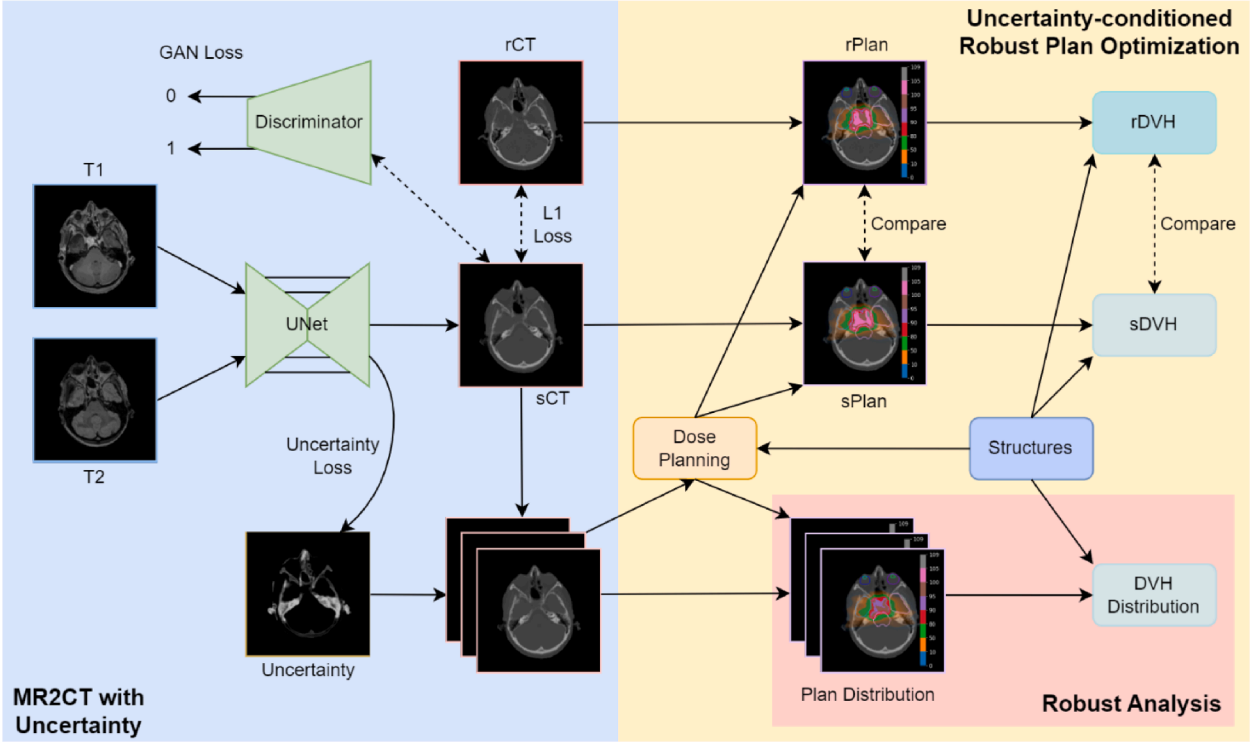

Fig. 1. Overview of the proposed uncertainty-aware MR-based proton therapy framework.

images using mutual information-based rigid registration provided by the open-source software ITK SNAP [35]. The MR images were subsequently resampled to match the resolution of the corresponding CT. This study adhered to ethical standards for research involving human data. Informed consent was obtained from all patients for the use of their anonymized data in scientific research. The anonymization process was conducted prior to the analysis to ensure confidentiality and compliance with ethical guidelines.

MR-based CT synthesis with uncertainty estimation

We employed Pix2pix as the image generation approach, as it can outperform CycleGAN on brain images when images are roughly aligned [36]. CT synthesis was achieved through a UNet-shape deep neural network [37], which took paired T1 and T2 images as inputs and generated synthetic CT (sCT) as output. Additionally, an uncertainty map with the same shape as sCT was predicted, measuring the voxel-wise uncertainty of the predicted sCT in the HU unit. Consequently, we predicted the distribution of the HU value of each voxel in the sCT instead of inferring a fixed value by the traditional methods [11,17,27,30,38,39]. In this case, the distribution was assumed to be a Laplacian distribution, as maximising its log-likelihood was consistent with minimising the mean absolute error (MAE) loss:

$$f(y|\mu,b)=\frac{1}{2b}\exp\left(-\frac{|y-\mu|}{b}\right)$$

Given the sCT prediction μ and the uncertainty prediction b, we aimed to maximise the log-likelihood of the Laplacian distribution given the observed CT data y with the paired MR data x. The log-likelihood can be formulated as follows:

$$LE=\log\prod_i f(y_i|\mu_i,b_i)$$

$$=\sum_i \log\frac{1}{2b_i}\exp\left(-\frac{|y_i-\mu_i|}{b_i}\right)$$

where i is the index of each voxel. Therefore, by maximising the log-likelihood, the loss can be calculated as:

$$L_{Laplacian}=\sum_i \log(b_i)+\frac{|y_i-\mu_i|}{b_i}.$$

To increase the numerical stability during optimisation, we replaced $\log(b_i)$ with c_i [40], then the final loss is:

$$L_{Laplacian}=\sum_i c_i+|y_i-\mu_i|\times\exp(-c_i).$$

The network illustrated in Fig. 1 follows the design in nnUNet[41], which comprises two parts: an encoder and a decoder. The encoder comprised a series of downsampling blocks, each consisting of a sequence of convolutional layers, a normalisation layer, and a nonlinear activation layer. The decoder has an equal number of blocks as the encoder, with each block replacing the convolution layer in the downsampling with a de-convolutional layer that functions as an upsampling operator. Following the decoder, two 1×1 convolutional layers output the estimated sCT and uncertainty map.

As for the network inputs, the 2D slices were used, which conserved computational resources while preserving the 2D geometry for consistent predictions. However, our approach differs from others in that instead of strictly sampling along a fixed axis, we sampled slices with random angles using bilinear interpolation. This strategy enlarges the potential sampling space, thus implicitly expanding the training cases. For implementation, rather than rotating the entire volume and sampling a slice, we first sample the coordinates of the slice and then rotate the coordinates of each voxel, reducing the computational load by order of magnitude. To train the network, we divided the 64 cases into eight folds and adopted a full cross-validation technique. For each of the eight-time validations, seven folds were used for training and the remaining one for validation. The networks were trained with the Adam optimiser employing an initial learning rate of 1e-3. The training lasted for 20 epochs; during the first half, the learning rate remained unchanged, while for the latter half, it decayed exponentially. The batch size was set to 16, and the network was trained on two V100 GPUs. In total, eight different network weights were acquired individually to fully utilise the limited patient dataset for a more thorough dosimetric evaluation below.

Uncertainty-conditioned robust treatment planning

Robust optimisation is becoming increasingly common in proton therapy [42] for dealing with range and setup uncertainty during delivery. With MR-based planning, the proposed framework can avoid the registration uncertainty, but network prediction uncertainties remain. To mitigate this, prediction uncertainty can be considered during the plan optimisation process to achieve robust plans. Usually, range robust optimisation assumes a uniform range error distribution applied to each voxel in the CT, typically of 3 %. However, such error assumptions do not apply to sCT imagery, as uncertainty is not expected to be homogenous across an image (see Fig. 2 below). As such, our proposed approach enables the derivation of voxel-specific uncertainty estimates. As described in the previous section, the prediction per-voxel is a Laplacian distribution $f(y_i|\mu_i,b_i)$. From this joint distribution, a batch of possible sCTs can be generated as error scenarios for robust plan optimisation. Due to memory limitations, only three sCTs were considered for each case, sampled by $\{\mu_i, \mu_i\text{-}b_i, \mu_i\text{+}b_i\}$. This sampling strategy is consistent with the general robust optimisation using $\{\mu_i, 0.97\mu_i, 1.03\mu_i\}$ when the error scenario follows a Gaussian distribution. For each case, the standard prescription was applied according to [43], and the fields were selected using the automatic algorithm developed at PSI [44]. Validation and comparison were performed for all 64 cases in the same manner of 8-fold cross-validation. To accommodate such large-scale dose planning, automatic data processing was adopted to reduce the manual burden.

Evaluation and statistics

The effectiveness of the proposed uncertain-conditioned framework was evaluated using both image-level and dosimetric-level metrics. All results were reported with the average and standard deviation scores of all 64 patients. For the image-level evaluation, we measured the conformality between synthetic CTs (sCTs) and real CTs (pCTs). The mean absolute error (MAE) for the body, bone, and soft tissue regions were evaluated, respectively. Additionally, the Dice score for bone (HU $>$ 200) and soft tissue regions (-200 $<$ HU $<$ 200) was calculated to evaluate the implicit classification ability of the network, which can be formulated as follows:

$$DSC=\frac{2|X\cap Y|}{|X|+|Y|}$$

Where X is the predicted region and Y is the ground-truth region, $\cap$ means the intersection operator. Moreover, the Pearson Correlation Coefficient (PCC) was calculated between the uncertainty and the absolute error $|y_i-\mu_i|$, to quantify the estimated uncertainty map b. Besides cross-validation, image-level evaluations were also adopted to the hold-out testing set to verify the performance of the trained networks.

For the dosimetric evaluation, for each of the 64 patients, the proton plan was generated and optimised on sCT (annotated as sPlan) using the proposed uncertainty-conditioned optimisation, the conventional 3 %/3mm global robust optimisation, and non-robust optimisation approaches. All types of sPlans were then recalculated on the corresponding pCT (as rPlan), and typical dose indexes were extracted from

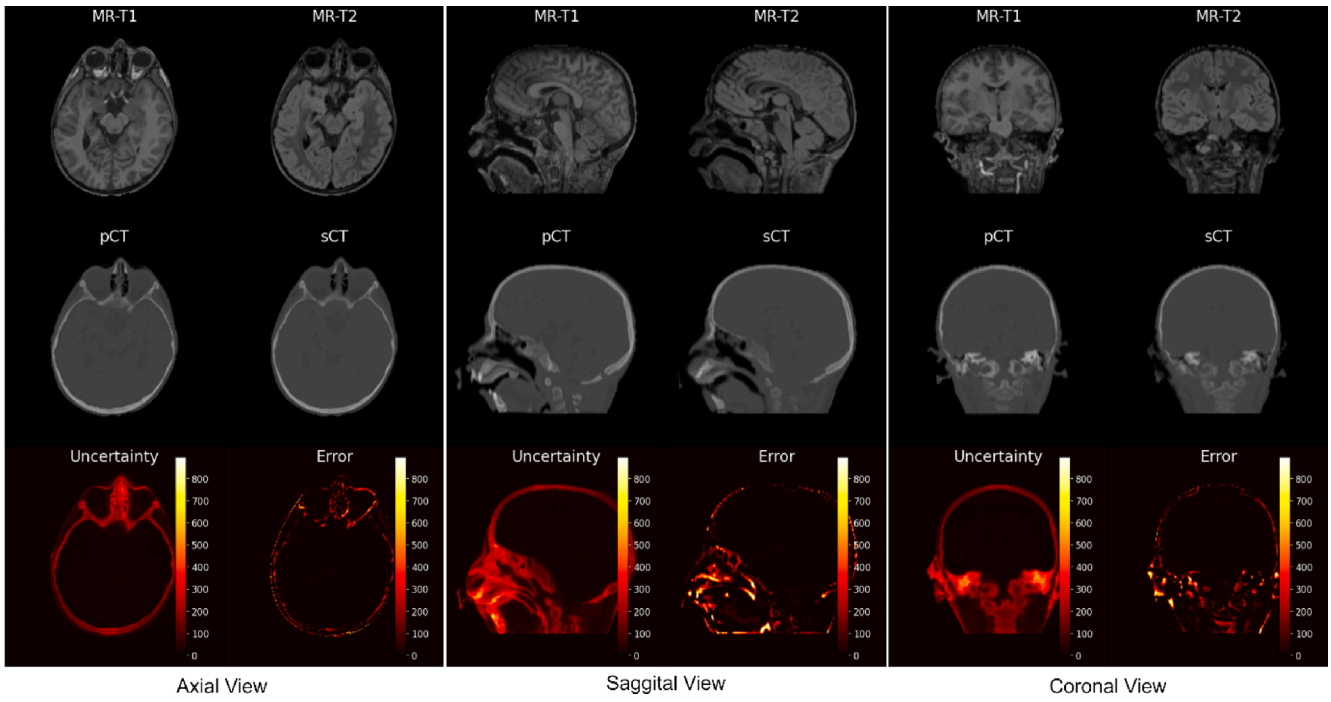

Fig. 2. Visualisations of MRIs (T1s and T2s), pCTs, sCTs, uncertainty maps, and error maps.

dose-volume histograms (DVHs) of CTV and affected OARs (e.g., chiasm, brain stem et al). Moreover, plan differences for each case were derived by subtracting point-to-point doses of the rPlan from those of each sPlan.

Results

For the image-level evaluation on the cross-validation set, image differences between sCT and pCT, in terms of the average Mean-Absolute-Error (MAE), were 80.84 ± 9.84 HU for the whole-body area, 35.78 ± 6.07 HU for soft tissues and 221.88 ± 31.69 HU for bone, with Dice scores of 90.33 ± 2.43 %, 95.13 ± 0.80 %, and 85.53 ± 4.16 %, respectively. The differences align with the MAE spectra in [25] that soft tissue areas (–200 < HU < 200) have a much lower MAE than bones (>200). Moreover, the predicted uncertainty distribution positively correlates with the absolute prediction error, with a Correlation Coefficient of 0.62 ± 0.01. Besides, on the testing set, the MAE is 78.23 ± 15.44 HU for the whole-body area, 39.89 ± 9.51 HU for soft tissues and 193.08 ± 32.99 HU for bone, while the Dice score is 91.69 ± 2.13. The full results of the testing set can be found in Supplement A. An example case is shown in Fig. 2 (more cases can be found in Supplement B), where the positive correlation between error maps and uncertainty maps can be clearly observed. Most of the above results show no significant differences among different populations. For example, the MAE is 82.55 ± 8.66 HU for adult patients, while it is 79.43 ± 10.51 HU for children. The only exception is the Dice scores of the bone region, which shows obvious differences between adults (88.01 ± 2.45 %) and children (83.47 ± 4.16 %). The best and worst cases for MAE are 59.00 and 108.82 HU, and 94.05 and 84.34 % for Dice. Pix2pix achieved a mean absolute error (MAE) of 80.29 ± 4.49 HU, significantly outperforming CycleGAN, which had a MAE of 153.36 ± 8.32 HU. The mean differences in stopping power relative to water were 5.25 ± 0.94 % for the cross-validation set and 5.54 ± 0.62 % for the hold-out set. Note that the above results are all from comparing the registered pCT with sCT. With the original pCT, the MAE of the whole body become 117.79 ± 16.60 HU.

In Fig. 3 (and supplement C), we show calculated rPlans, sPlans and the corresponding plan differences for the different plan optimisation algorithms. Fig. 4 presents and compares the uncertainty effect on DVHs of CTV and affected OARs. It is evident that the proposed uncertainty-conditioned approach can achieve much better agreement between sPlans and rPlans. For example, the absolute difference of CTV-D95 from the proposed optimisation algorithm is only 0.66 ± 1.11 %, in contrast to 1.32 ± 1.93 % and 1.65 ± 2.22 % from the 3 %/3mm global robust and non-robust optimisation, respectively. Curves of rPlans are generally inferior to sPlans because we optimize the dose planning on sCT to mimic the real clinical application of MR-based radiotherapy, where only MR and sCT can be acquired. Besides, for Spinal cord-D2, the numbers are 1.15 ± 2.94 %, 1.45 ± 3.58 % and 1.85 ± 4.10 %, respectively. This trend was statistically consistent across most indexes of both target and OARs for all patients in this dataset, as shown in Fig. 5.

As summarized by [50], the Deep Learning-based method can be used for Quality assurance (QA). With the proposed framework, physicians should be able to visualise the over-shoot and under-shoot cases during the pre-treatment discussion, as illustrated in Fig. 6 *(a-d)*. Moreover, uncertainty analysis using DVH can serve as a valuable tool for assessing the generation quality of sCT, where superior generation with low uncertainty is associated with a tight band of curves in the last sub-figure of Fig. 6. Possible usages of the uncertainty map: (upper) plan robust analysis by considering example extreme sCT estimation. (e.g., maximum and minimum cases, under-shoot and -over-shoot cases); (lower) uncertainty-aware DVH. (Fig. 6e and supplement D). These metrics and visualisation are useful tools when implementing MR-based treatment planning, as clinicians can decide, based on analytic data evidence, if new CT acquisition needs to be conducted or if it suffices to utilise the MR-based sCT directly.

Discussion

We developed a deep learning framework that produces high-quality synthetic CT from MR images while also generating a corresponding uncertainty map to assist in the creation of a robust treatment plan. This framework enables uncertainty-aware proton planning, bringing DL-based CT synthesis closer to the proton clinical practice. The derived uncertainty-conditioned optimisation algorithm outperforms the default robust optimisation in dosimetric-level conformity. The novelty arises from considering the voxel-wise uncertainty estimates for proton

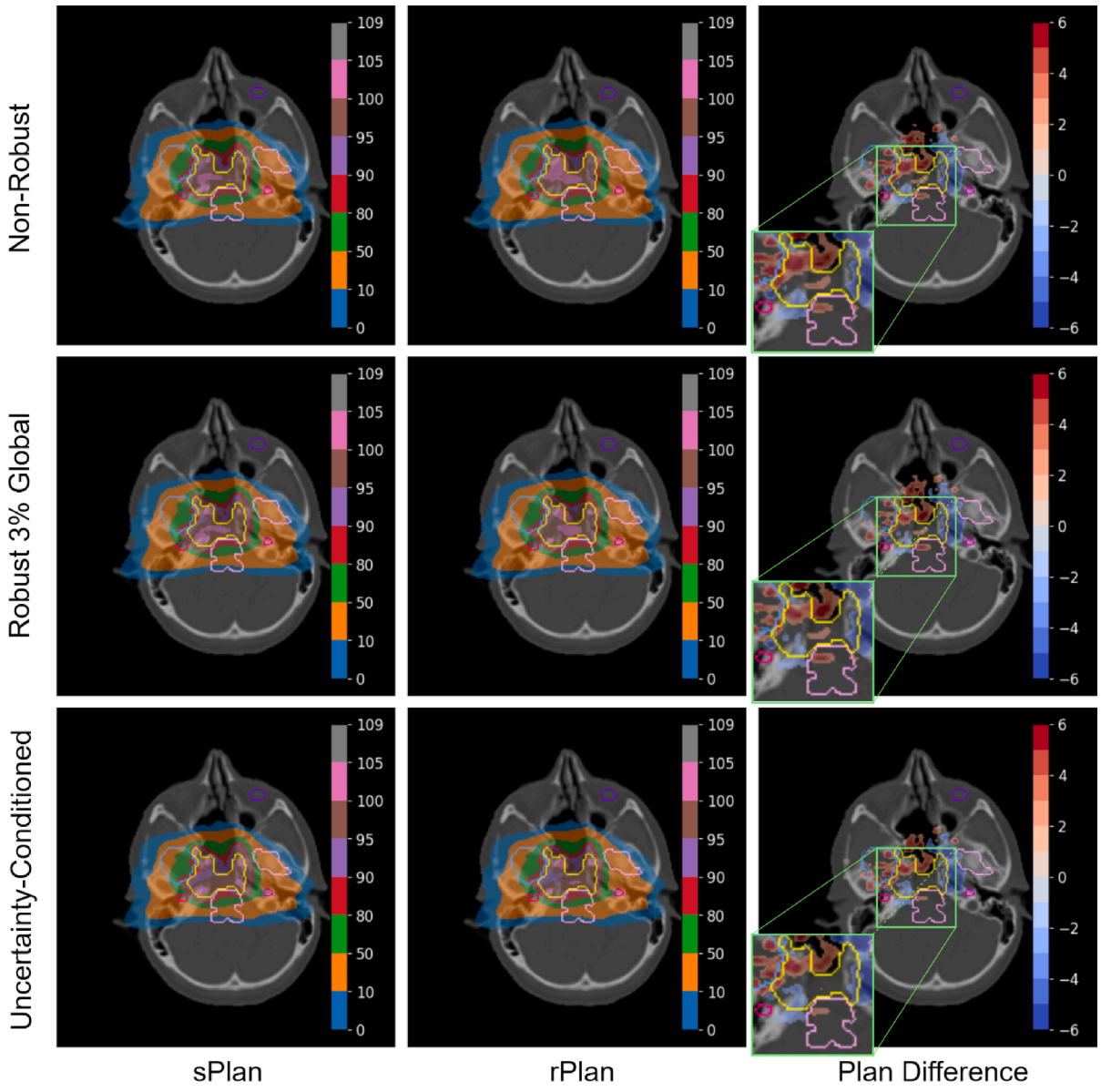

Fig. 3. Visualisations of optimized plan on pCT, the recalculated plan sCT and their difference across the three planning strategies of (upper) non-robust, (middle) conventional robust optimized and (bottom) uncertainty-aware.

treatment planning, including its principle, estimation, and consistent application for proton plan optimisation.

Along with presenting experimental results and visualisations, we elaborate on our findings in the subsequent paragraphs. The first issue to be addressed pertains to alignment. In previously published studies [17,38] on MR-based CT synthesis, two primary frameworks, Pix2pix and CycleGAN, have been employed for image generation. However, different papers yielded contradictory conclusions regarding their comparative performance [14]. Typically, Pix2pix requires well-aligned image pairs, while CycleGAN does not depend on such a critical pre-processing. Within our experiments, Pix2pix outperforms CycleGAN by a large margin, showing rigid registration can generate satisfactory alignments for Pix2pix in brain regions (see results in Supplement E). It is worth noting that we deliberately computed sCT-pCT differences in the original image resolution of the pCT (the planning CT) without any down-sampling, as it would be used in clinical practice. When such evaluation is conducted in the down-sampled image, the error reduction of approximately 20HU would be additionally achieved. Indeed, the lack of homogeneity in the dataset and evaluation metrics makes it difficult to draw a quantitative conclusion about the superiority of any single method over the others. We expect that the current ongoing grand challenge (SynthRAD2023 [45]) will assist in fairly understanding the performance discrepancies among the different proposed methods for the task of MR-based synthetic CT generation.

Nonetheless, this observation does not imply that the achieved registration was sufficiently aligned. Imperfect alignment can inheritably bias the evaluation since a portion of MAE stems from spatial misalignment rather than range errors. To address this issue, we employed the correlation-based deformable registration algorithm by ITK SNAP to register the pCT to sCT, annotated as regCT. All the above-reported values were results from the comparison of regCT and sCT. In contrast, the MAE between pCT and sCT without this extra deformable registration is significantly higher, indicating that the post-deformable registration can further reduce the misalignment error by a large percentage. These findings demonstrated that a substantial portion of errors originates from input image misalignment, highlighting the need for more advanced registration methods to achieve high-quality MR-based CT generation.

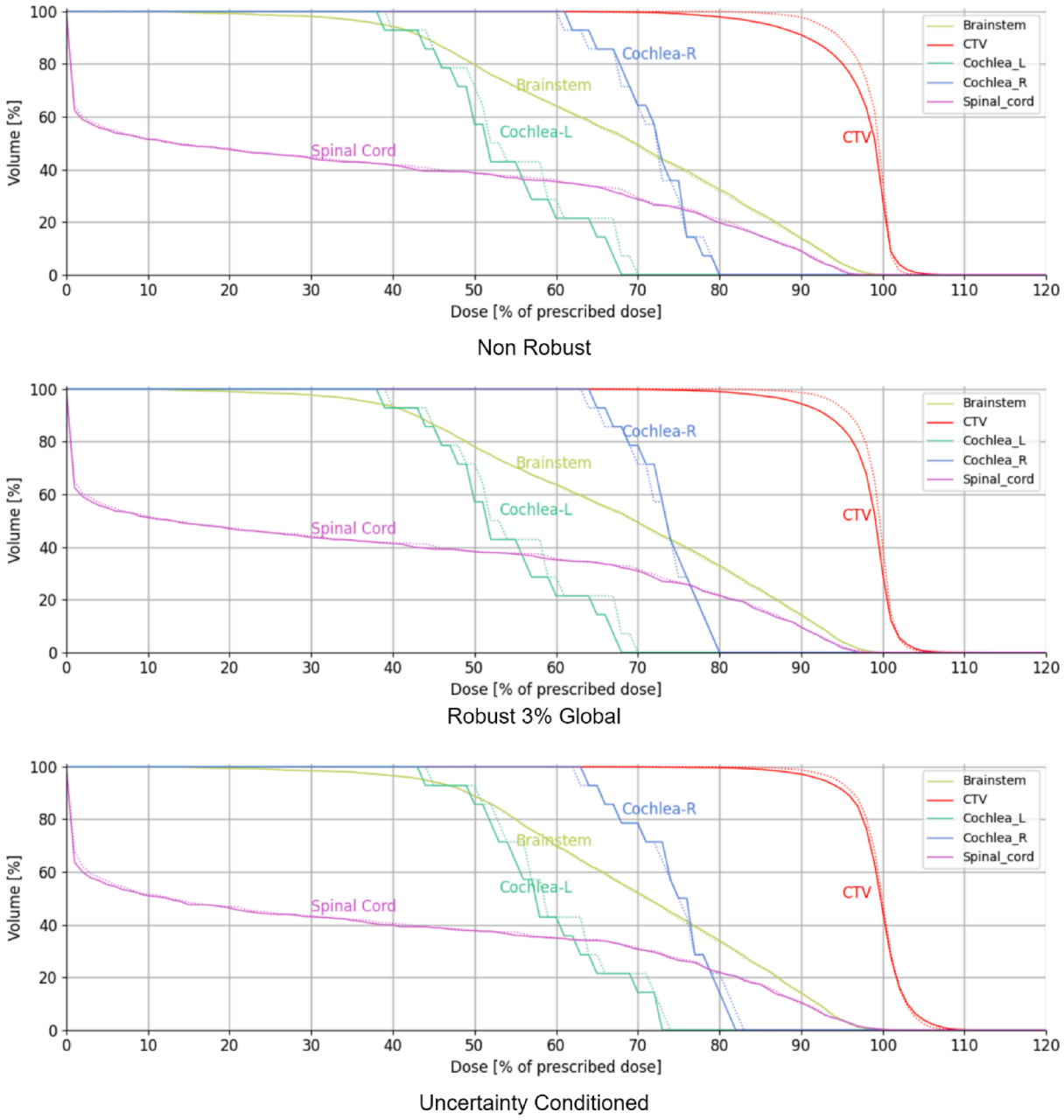

Fig. 4. Comparisons of DVHs of one example case between sPlan and rPlan among the three planning strategies (rPlans: solid lines, sPlan: dashed lines).

In addition to the alignment issues, deep neural networks often exhibit over-confidence in their prediction, even when encountering out-of-distribution data. In radiotherapy, corner cases (e.g., patient-specific outliers) are inevitable, posing a challenge due to the limited generalizability of the trained network. In this study, the output uncertainty map is a proxy measure of network confidence. The uncertainty quantification not only offers an overall metric but also provides a detailed spatial distribution to assist the treatment planning procedure, such as the selection of optimal field direction. Given its high Correlation Coefficient with the absolute error values, the estimated uncertainty map can also act as a surrogate for the patient-specific error map that is inaccessible in clinical practice. We see further research into considering the uncertainty map during radiation field selection as desirable.

In natural images, for example Pix2pix tasks, mean squared error (MSE) is usually adopted as the evaluation metric. Therefore, Gaussian distribution per-voxel is assumed for the uncertainty estimation because maximising the log-likelihood of Gaussian distributions corresponds to minimising the MSE loss. However, for medical images, MAE is preferable, as it directly correlated to the meaningful unit (such as HU), so that Laplacian distribution was chosen.

Typically, robust plan optimisation was used to address setup and range uncertainty [46,47], e.g., 3 mm and 3 % (or other uniform values), often used as experience values. However, with our proposed framework, since per-voxel distribution can already be obtained by uncertainty estimation, it is feasible to have a finer model of the uncertainty in 3D with high spatial resolution. Sampling with per-voxel distribution rather than global distribution can more effectively cover the possible sCT. In this study, the pCT was treated as the ultimate reference against which the sCT was compared. However, we have not yet considered the additional uncertainty of pCT calibration for proton treatment. Nevertheless, the voxel-specific uncertainty approach can also be employed in the conventional range robust evaluation and optimisation to mitigate the CT calibration uncertainty [48,49]. For instance, although the uncertainty in CT due to calibration from HU to proton stopping power is as

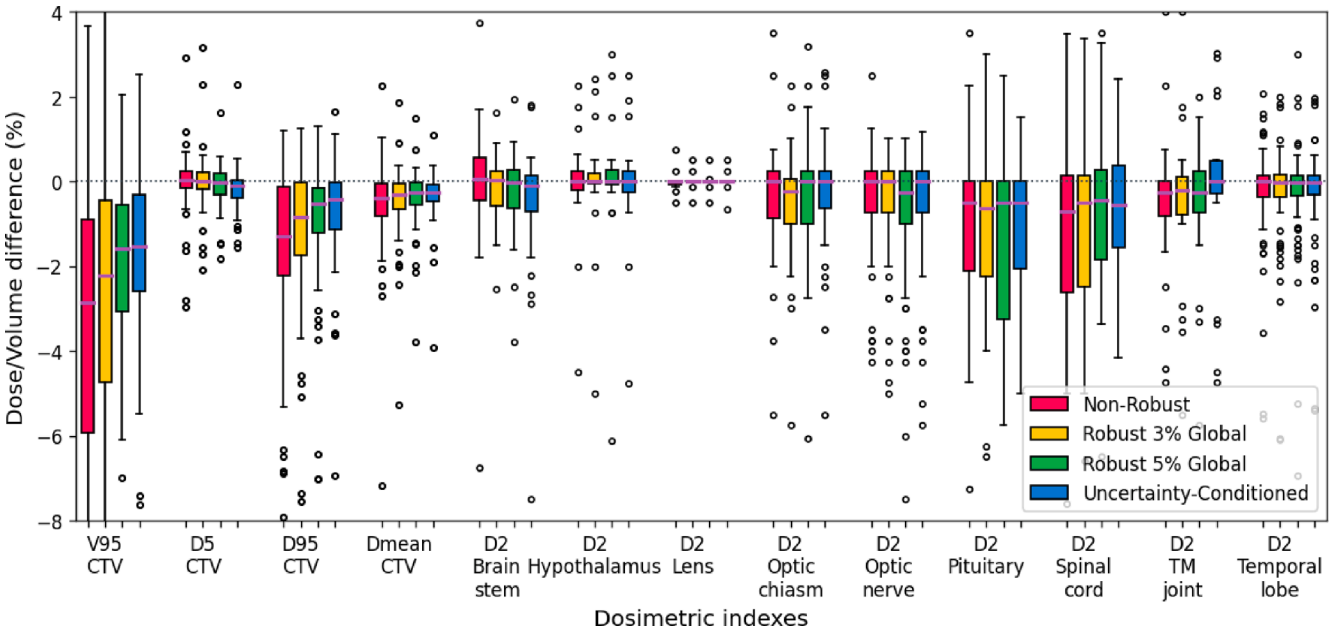

Fig. 5. Distributions of dose index differences, compared across three dose planning algorithms. (Boxplot derived from the evaluations of all 64 patient cases).

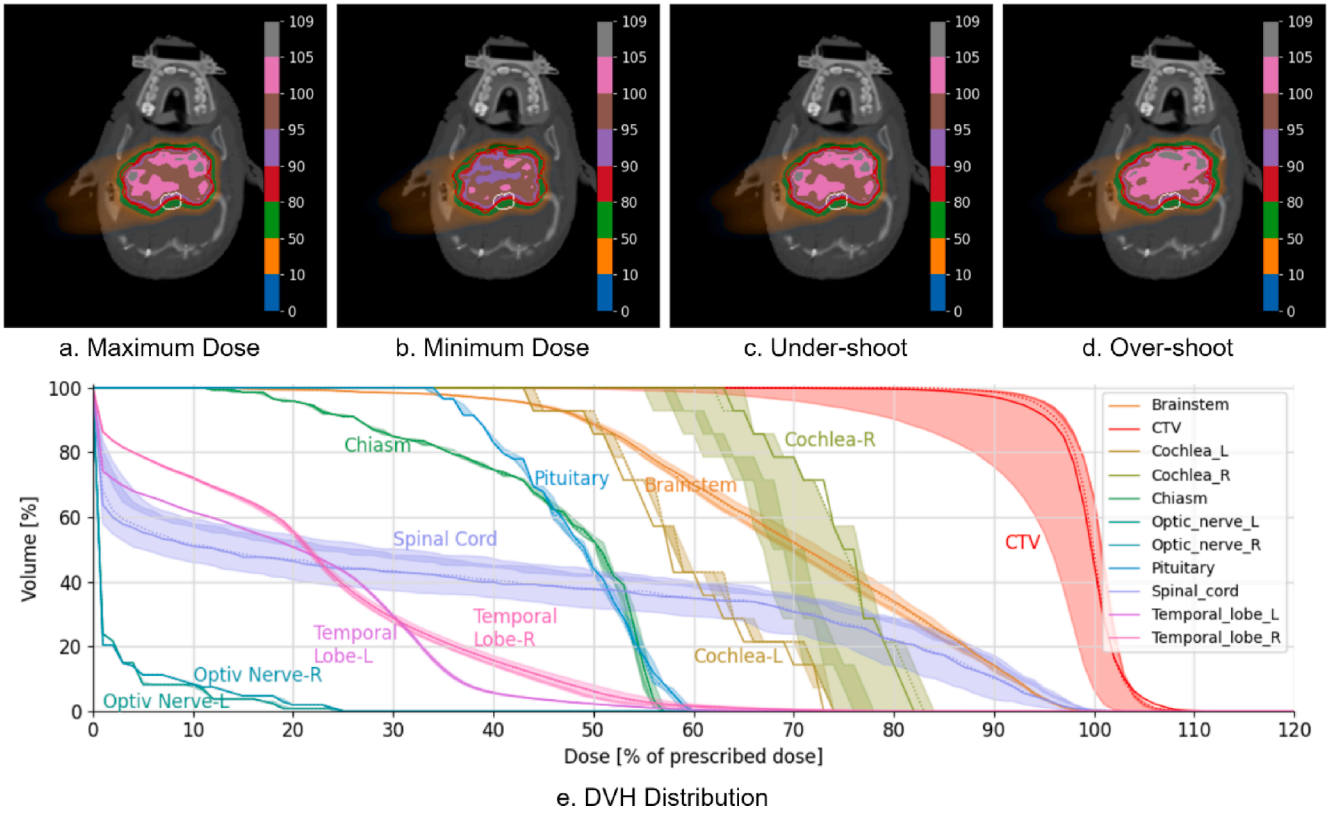

Fig. 6. Possible usages of the uncertainty map: (upper) plan robust analysis by considering example extremed sCT estimation. (e.g., maximum and minimum cases, under-shoot and -over-shoot cases); (lower) uncertainty-aware DVH.

high as 3 % in bone, it is much lower in soft tissue (more likely 1 % or lower). Instead of using a systematically conservative value of 3 % or 5 % globally, a range uncertainty distribution specific to the region can also be considered.

Although the discrepancy at the image-level remained substantial (more than 80 HU), we observed quite satisfactory results at the dosimetric-level. Despite the observed sCT-rCT differences, the dose recalculation error remains below 2 %, with V95 and D95 discrepancies under 0.2 %, as depicted in Fig. 5. As advancements in network architecture and dataset size continue, the error is expected to drop, rendering MR-based proton therapy increasingly feasible.

Our work was limited by the number of sampled sCTs used for robust plan optimisation. As we employed a full-batch optimisation algorithm in our treatment planning system, considering numerous possible

sampling cases simultaneously was challenging. In this study, only 3 sampled sCTs were used. To enable better coverage of sampling spaces, mini-batch optimisation [51] may provide an effective solution, as it could reduce memory requirements. In this study, patients with both T1 and T2 images were evaluated. Utilizing two distinct sequences enhances the robustness of predictions due to the complementary information they provide. However, this approach limits the inclusion of additional cases in the dataset. Such a trade-off warrants consideration in subsequent research.

In summary, we have demonstrated the high-quality sCT generation from MR images using a modified Pix2pix-based framework. Besides the improved accuracy for proton therapy planning, the associated 3D uncertainty distribution enables voxel-specific robust optimisation for improving plan robustness against prediction errors for proton treatment of brain tumours. We believe the direct visualisation of the network prediction uncertainty and the subsequential robust analyses are powerful tools for determining the clinical usefulness of synthetic CTs for individual patients.

CRediT authorship contribution statement

Xia Li: Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Renato Bellotti:** Software, Writing – review & editing. **Gabriel Meier:** Software, Writing – review & editing. **Barbara Bachtiary:** Resources, Writing – review & editing. **Damien Weber:** Conceptualization, Resources, Writing – review & editing, Funding acquisition. **Antony Lomax:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Joachim Buhmann:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Ye Zhang:** Conceptualization, Methodology, Investigation, Validation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This project is supported by the interdisciplinary doctoral grants (iDoc 2021-360) from the Personalized Health and Related Technologies (PHRT) of the ETH domain, Switzerland. The automatic planning system used in this work was developed as part of the EU-H2020 project 'INSPIRE' (INfraStructure in Proton International REsearch; grant ID: 730983)

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2023.110056.

References

[1] Smith AR. Proton therapy. Phys Med Biol 2006:51. https://doi.org/10.1088/0031-9155/51/13/R26.
[2] Albertini F, Matter M, Nenoff L, Zhang Y, Lomax A. Online daily adaptive proton therapy. Br J Radiol 2020;93. https://doi.org/10.1259/bjr.20190594.
[3] Khoo VS, Dearnaley DP, Finnigan DJ, Padhani A, Tanner SF, Leach MO. Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. Radiother Oncol 1997;42. https://doi.org/10.1016/S0167-8140(96)01866-X.
[4] Akbarzadeh A, Gutierrez D, Baskin A, Ay MR, Ahmadian A, Riahi Alam N, et al. Evaluation of whole-body MR to CT deformable image registration. J Appl Clin Med Phys 2013;14. https://doi.org/10.1120/jacmp.v14i4.4163.
[5] Ribeiro CO, Knopf A, Langendijk JA, Weber DC, Lomax AJ, Zhang Y. Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. Radiother Oncol 2018;128. https://doi.org/10.1016/j.radonc.2018.03.001.
[6] Hoffmann A, Oborn B, Moteabbed M, Yan S, Bortfeld T, Knopf A, et al. MR-guided proton therapy: a review and a preview. Radiat Oncol 2020;15. https://doi.org/10.1186/s13014-020-01571-x.
[7] Thummerer A, Seller Oria C, Zaffino P, Meijers A, Guterres Marmitt G, Wijsman R, et al. Clinical suitability of deep learning based synthetic CTs for adaptive proton therapy of lung cancer. Med Phys 2021;48. https://doi.org/10.1002/mp.15333.
[8] Thummerer A, de Jong BA, Zaffino P, Meijers A, Marmitt GG, Seco J, et al. Comparison of the suitability of CBCT- And MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients. Phys Med Biol 2020:65. https://doi.org/10.1088/1361-6560/abb1d6.
[9] Guerreiro F, Burgos N, Dunlop A, Wong K, Petkar I, Nutting C, et al. Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning. Phys Med 2017;35. https://doi.org/10.1016/j.ejmp.2017.02.017.
[10] Burgos N, Guerreiro F, McClelland J, Presles B, Modat M, Nill S, et al. Iterative framework for the joint segmentation and CT synthesis of MR images: Application to MRI-only radiotherapy treatment planning. Phys Med Biol 2017;62. https://doi.org/10.1088/1361-6560/aa66bf.
[11] Korsager AS, Carl J, Østergaard LR. Comparison of manual and automatic MR-CT registration for radiotherapy of prostate cancer. J Appl Clin Med Phys 2016;17. https://doi.org/10.1120/jacmp.v17i3.6088.
[12] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. Comput Intell Neurosci 2018;2018. https://doi.org/10.1155/2018/7068349.
[13] Massa HA, Johnson JM, McMillan AB. Comparison of deep learning synthesis of synthetic CTs using clinical MRI inputs. Phys Med Biol 2020;65. https://doi.org/10.1088/1361-6560/ABC5CB.
[14] Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. Med Phys 2021;48:6537–66. https://doi.org/10.1002/MP.15150.
[15] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017- January, 2017. https://doi.org/10.1109/CVPR.2017.632.
[16] Zhu JY, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2017- October, 2017. https://doi.org/10.1109/ICCV.2017.244.
[17] Yang H, Sun J, Carass A, Zhao C, Lee J, Prince JL, et al. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. IEEE Trans Med Imaging 2020;39:4249–61. https://doi.org/10.1109/TMI.2020.3015379.
[18] Lei Y, Harms J, Wang T, Liu Y, Shu HK, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med Phys 2019;46:3565–81. https://doi.org/10.1002/MP.13617.
[19] Thummerer A, Zaffino P, Meijers A, Marmitt GG, Seco J, Steenbakkers RJHM, et al. Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy. Phys Med Biol 2020;65:095002. https://doi.org/10.1088/1361-6560/AB7D54.
[20] Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep Convolution Neural Network (DCNN) multiplane approach to synthetic CT generation from MR images—Application in brain proton therapy. Int J Radiat Oncol*Biol*Phys 2019;105:495–503. https://doi.org/10.1016/J.IJROBP.2019.06.2535.
[21] Liu Y, Lei Y, Wang Y, Wang T, Ren L, Lin L, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method HHS public access. Phys Med Biol 2019;64:145015. https://doi.org/10.1088/1361-6560/ab25bc.
[22] Maspero M, Bentvelzen LG, Savenije MHF, Guerreiro F, Seravalli E, Janssens GO, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. Radiother Oncol 2020. https://doi.org/10.1016/j.radonc.2020.09.029.
[23] Liu Y, Lei Y, Wang Y, Shafai-Erfani G, Wang T, Tian S, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. Phys Med Biol 2019;64:205022. https://doi.org/10.1088/1361-6560/AB41AF.
[24] Landry G, Parodi K, Wildberger JE, et al. Synthetic dual-energy CT for MRI-only based proton therapy treatment planning using label-GAN You may also like Deriving concentrations of oxygen and carbon in human tissues using single-and dual-energy CT for ion therapy applications. Phys Med Biol 2021;66:65014. https://doi.org/10.1088/1361-6560/abe736.
[25] Zimmermann L, Knäusl B, Stock M, Lütgendorf-Caucig C, Georg D, Kuess P. An MRI sequence independent convolutional neural network for synthetic head CT generation in proton therapy. Z Med Phys 2022;32:218–27. https://doi.org/10.1016/J.ZEMEDI.2021.10.003.
[26] Parrella G, Vai A, Nakas A, Garau N, Meschini G, Camagni F, et al. Synthetic CT in carbon ion radiotherapy of the abdominal site. Bioengineering 2023;10:250. https://doi.org/10.3390/BIOENGINEERING10020250/S1.
[27] Zeng G, Zheng G. Hybrid generative adversarial networks for deep MR to CT synthesis using unpaired data. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11767 LNCS, 2019. https://doi.org/10.1007/978-3-030-32251-9_83.
[28] Mohanty S, Dakua SP. Toward computing cross-modality symmetric non-rigid medical image registration. IEEE Access 2022;10. https://doi.org/10.1109/ACCESS.2022.3154771.

[29] Wu J, Zhou S. A Disentangled Representations based Unsupervised Deformable Framework for Cross-modality Image Registration. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2021. https://doi.org/10.1109/EMBC46164.2021.9630778.
[30] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44:1408–19. https://doi.org/10.1002/MP.12155.
[31] Wilson G, Cook DJ. A survey of unsupervised deep domain adaptation. ACM Trans Intell Syst Technol 2020;11. https://doi.org/10.1145/3400066.
[32] Bonechi S, Andreini P, Bianchini M, Pai A, Scarselli F. Confidence measures for deep learning in domain adaptation. Appl Sci (Switzerland) 2019;9. https://doi.org/10.3390/app9112192.
[33] Unkelbach J, Paganetti H. Robust proton treatment planning: physical and biological optimization. Semin Radiat Oncol 2018;28. https://doi.org/10.1016/j.semradonc.2017.11.005.
[34] El Naqa I, Boone JM, Benedict SH, Goodsitt MM, Chan HP, Drukker K, et al. AI in medical physics: guidelines for publication. Med Phys 2021;48:4711–4. https://doi.org/10.1002/MP.15170.
[35] Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2016;2016-October:3342–5. https://doi.org/10.1109/EMBC.2016.7591443.
[36] Jin C-B, Kim H, Liu M, Jung W, Joo S, Park E, et al. Deep CT to MR synthesis using paired and unpaired data. Sensors 2019. https://doi.org/10.3390/s19102361.
[37] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351, 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
[38] Bin Jin C, Kim H, Liu M, Jung W, Joo S, Park E, et al. Deep CT to MR synthesis using paired and unpaired data. Sensors (Switzerland) 2019;19. https://doi.org/10.3390/S19102361.
[39] Guo Y, Wu X, Wang Z, Pei X, Xu XG. End-to-end unsupervised cycle-consistent fully convolutional network for 3D pelvic CT-MR deformable registration. J Appl Clin Med Phys 2020;21. https://doi.org/10.1002/acm2.12968.
[40] Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. Inf Fusion 2021;76. https://doi.org/10.1016/j.inffus.2021.05.008.
[41] Isensee F, Jaeger PF. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021. https://doi.org/10.1038/s41592-020-01008-z.
[42] Liu W, Zhang X, Li Y, Mohan R. Robust optimization of intensity modulated proton therapy. Med Phys 2012;39:1079–91. https://doi.org/10.1118/1.3679340.
[43] Lee AW, Ng WT, Pan JJ, Chiang CL, Poh SS, Choi HC, et al. International guideline on dose prioritization and acceptance criteria in radiation therapy planning for nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2019;105. https://doi.org/10.1016/j.ijrobp.2019.06.2540.
[44] Bellotti R, Willmann J, Weber D, Lomax T, Adelmann A, Hrbacek J. Fully automatic treatment planning for intensity-modulated proton therapy using machine learning and heuristic optimization. Madrid: PTCOG; 2023.
[45] Thummerer A, van der Bijl E, Jr Galapon A, Verhoeff JJ, Langendijk JA, Both S, et al. SynthRAD2023 Grand Challenge dataset: generating synthetic CT for radiotherapy. Med Phys 2023. https://doi.org/10.5281/zenodo.7260705.
[46] Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Med Phys 2011;38. https://doi.org/10.1118/1.3556559.
[47] Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. Phys Med Biol 2007;52. https://doi.org/10.1088/0031-9155/52/10/009.
[48] Wohlfahrt P, Möhler C, Hietschold V, Menkel S, Greilich S, Krause M, et al. Clinical Implementation of Dual-energy CT for Proton Treatment Planning on Pseudo-monoenergetic CT scans. Int J Radiat Oncol*Biol*Phys 2017;97:427–34. https://doi.org/10.1016/J.IJROBP.2016.10.022.
[49] Wohlfahrt P, Möhler C, Richter C, Greilich S. Evaluation of stopping-power prediction by dual- and single-energy computed tomography in an anthropomorphic ground-truth phantom. Int J Radiat Oncol*Biol*Phys 2018;100:244–53. https://doi.org/10.1016/J.IJROBP.2017.09.025.
[50] El Naqa I, Das S. The role of machine and deep learning in modern medical physics. Med Phys 2020;47:e125–6. https://doi.org/10.1002/MP.14088.
[51] Fabiano S, Bangert M, Guckenberger M, Unkelbach J. Accounting for range uncertainties in the optimization of combined proton-photon treatments via stochastic optimization. Int J Radiat Oncol Biol Phys 2020;108:792–801. https://doi.org/10.1016/J.IJROBP.2020.04.029.