Automatic generation of three-dimensional dose reconstruction data for two-dimensional radiotherapy plans for historically treated patients

Ziyuan Wang
Marco Virgolin
Peter A. N. Bosman
Koen F. Crama
Brian V. Balgobind
Arjan Bel
Tanja Alderliesten

Automatic generation of three-dimensional dose reconstruction data for two-dimensional radiotherapy plans for historically treated patients

Ziyuan Wang,^{a,*} Marco Virgolin,^b Peter A. N. Bosman,^b Koen F. Crama,^a
Brian V. Balgobind,^a Arjan Bel,^a and Tanja Alderliesten^a

^aUniversity of Amsterdam, Amsterdam UMC, Department of Radiation Oncology,
Amsterdam, The Netherlands

^bCentrum Wiskunde and Informatica, Life Sciences and Health Group, Amsterdam,
The Netherlands

Abstract. Performing large-scale three-dimensional radiation dose reconstruction for patients requires a large amount of manual work. We present an image processing-based pipeline to automatically reconstruct radiation dose. The pipeline was designed for childhood cancer survivors that received abdominal radiotherapy with anterior-to-posterior and posterior-to-anterior field set-up. First, anatomical landmarks are automatically identified on two-dimensional radiographs. Second, these landmarks are used to derive parameters to emulate the geometry of the plan on a surrogate computed tomography. Finally, the plan is emulated and used as input for dose calculation. For qualitative evaluation, 100 cases of automatic and manual plan emulations were assessed by two experienced radiation dosimetrists in a blinded comparison. The two radiation dosimetrists approved 100%/100% and 92%/91% of the automatic/manual plan emulations, respectively. Similar approval rates of 100% and 94% hold when the automatic pipeline is applied on another 50 cases. Further, quantitative comparisons resulted in on average <5 mm difference in plan isocenter/borders, and <0.9 Gy in organ mean dose (prescribed dose: 14.4 Gy) calculated from the automatic and manual plan emulations. No statistically significant difference in terms of dose reconstruction accuracy was found for most organs at risk. Ultimately, our automatic pipeline results are of sufficient quality to enable effortless scaling of dose reconstruction data generation. © 2020 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.7.1.015001](https://doi.org/10.1117/1.JMI.7.1.015001)]

Keywords: digitally reconstructed radiograph; landmark detection; radiotherapy; dose reconstruction; plan emulation.

Paper 19269R received Oct. 8, 2019; accepted for publication Jan. 17, 2020; published online Feb. 3, 2020.

1 Introduction

Studies have shown that childhood cancer survivors who received radiotherapy (RT) are prone to develop late adverse effects, i.e., long-term morbidity and health problems that occur more than ten years after the treatment.¹⁻³ To improve the design of RT plans and reduce late adverse effects for future childhood cancer patients, it is important to understand how RT dose relates to late adverse effects. RT dose information (e.g., prescribed dose) can be inferred from retrospective data.⁴⁻⁶ However, detailed information on the dose in organs at risk (OARs) is often missing in these historical patients' RT records, and so is three-dimensional (3-D) imaging of the patient anatomy at the time of treatment, which is required to estimate a 3-D dose distribution.^{7,8} In fact, many cancer survivors for whom long-term follow-up is available were treated before computed tomography (CT) was widely used for RT planning (e.g., the late-90s in The Netherlands) after its first introduction in the 1970s.⁹ For these historically treated patients, solely two-dimensional (2-D) radiographs were used for RT planning.^{10,11} The lack of 3-D anatomy images (i.e., CTs) of patients precludes direct 3-D dose calculation on CT, as is routinely performed nowadays.

*Address all correspondence to Ziyuan Wang, E-mail: z.wang@amc.uva.nl

To deal with the absence of 3-D imaging of historically treated patients, the dose reconstruction methods have been developed.^{8,12–15} These methods provide 3-D dose estimations by emulating the historical treatment on a 3-D patient-resembling surrogate anatomy, e.g., physical phantoms,¹² computational phantoms,^{8,12,15} or surrogate CTs.^{12,14,16}

For a dose reconstruction approach, the discrepancy between the surrogate anatomy and the true anatomy of the patient for whom the plan was originally designed is one of the main sources of uncertainties, making it an important topic to research.^{7,8,17} As variations in anatomy exist in the population, it has been found that an approach that uses a representative surrogate patient's anatomy to reconstruct RT dose for an individual patient can introduce large dose uncertainties.^{8,18} In recent years, multiple computational phantom libraries have been developed that provide representative anatomies of patients categorized by gender, age, weight, and height, or other types of features.^{19–23} Also, studies have been published on how to simulate 3-D organ shapes given patient features extracted from the clinical record (such as patient diameter at isocenter) and 2-D radiographs (such as vertebral column length), in an attempt to adapt phantoms to be more individualized.^{13,24} However, validation of individualized RT dose reconstruction approaches is rarely reported, and the few studies reporting validation results are based on a small data set.^{13,15}

Instead of selecting the best-matching surrogate anatomy based on some reasonable but perhaps too simplistic criteria (e.g., age and gender), researchers have been exploring new criteria and have validated the use of these in terms of dose reconstruction accuracy, which is what ultimately matters. For example, a study on organ dose reconstruction for CT dosimetry assessed several matching criteria, such as age and gender, height and weight, and water equivalent diameter, and concluded that the latter one is superior in terms of reconstructed dose quality.²⁵ In our recent study, we investigated the possible correlations between dose reconstruction accuracy and deviations in patient age, height, weight, and several anatomical features, with the goal of understanding which patient features should be used for surrogate selection.¹⁷ Organ location was found to be a dominant feature, but it was also observed that these findings depend on the type of RT plan since the plan geometry has a large influence on the dose reconstruction.

We identify the need to generate a large number of dose reconstruction data for the following purposes:

- for the development of an individualized dose reconstruction approach. It is expected that more powerful data analysis techniques such as machine learning can be utilized if dose reconstruction is performed on big data sets, including many different patients' anatomies and many different plan geometries;^{17,18,24}
- for the validation of a developed dose reconstruction approach;^{7,26}
- for the use of a dose reconstruction approach to enable effective modeling of the relationship between the treatment (including radiation treatment) and onset of late adverse effects.^{12,27}

Currently, generating large numbers of dose reconstruction data for historically treated patients (i.e., patients treated with conventional RT with treatment plans designed according to specific anatomical features as visible on 2-D radiographs) in a short period of time is not possible. Following the challenge of collecting big amounts of patient data, time-consuming manual steps are needed to reconstruct the dose. In general, dose reconstruction consists of three fundamental steps:²⁴ (1) preparation/selection of the surrogate anatomy; (2) emulation of the RT plan of the historical patient on the surrogate anatomy; and (3) calculation/measurement of the 3-D dose distribution.

The last step, i.e., the dose calculation, can be automated, by scripting the dose calculation algorithms in a commercial treatment planning system (TPS), or using other standalone analytical or Monte-Carlo dose calculation algorithms.^{15,16,28} The bottleneck that prevents performing dose reconstruction on a large scale is the intermediate step of RT plan emulation (step 2).

Plan emulation is currently a manual task where an experienced RT planner makes a plan on the surrogate anatomy that resembles as much as possible the plan made for the historical patient. Reproducing the geometry of the historical treatment field is herein the key difficulty. In previous studies, geometrical features (e.g., positions of field borders) of the treatment field with respect to anatomical landmarks based on the clinical record or historical 2-D radiographs of the patient are

collected and used to emulate the plan on a surrogate phantom/CT.^{12,15,17,18,26} For each reconstruction case, RT planning experts need to visually check how the treatment field shape is configured with respect to the visible anatomy, to then reproduce the field geometry on the surrogate anatomy, which makes the process subjective and time-consuming.

To enable quick and effortless generation of large-scale dose reconstruction data, we recently proposed a pipeline based on image processing techniques for automatic plan emulation on surrogate CT scans and reported promising preliminary results.²⁹ The focus of that work was on one of the most common types of childhood abdominal cancer (i.e., Wilms' tumor, a type of kidney tumor), which is treated with two opposing RT beams [i.e., anterior-to-posterior (AP) and posterior-to-anterior (PA)]. In this article, we extend our previous pilot study by including: (1) an improved methodology for landmark detection and plan emulation; (2) addition of automatic digitally reconstructed radiograph (DRR) generation, dose calculation, and a plan visualization module to the pipeline; and (3) a more informative and comprehensive validation of the pipeline on a larger data set.

2 Methods

The pipeline proposed in this article is shown in Fig. 1. The input of our pipeline is the 2-D radiograph (in coronal view) of the patient for whom dose reconstruction needs to be performed (reference patient), the RT record, and the surrogate CT scan for which the plan must be emulated. The outputs are the emulated plan and the calculated 3-D dose distribution. Since no ground-truth dose distributions exist for historically treated patients, we evaluate our approach by creating RT records of recently treated patients' CTs, that are in line with the data available in the past, as reference data for which dose need to be reconstructed. In detail, DRRs are derived from the reference patient's CT scans (reference CTs), as substitutes for historical 2-D radiographs. DRRs of the surrogate CTs are also generated (surrogate DRRs), as the plan will be emulated based on information extracted from DRRs.

In the following sections, we present step by step the preprocessing, landmarks detection, plan emulation, and dose calculation and plan visualization parts of our pipeline. The code used in the dose reconstruction pipeline is available in a GitHub repository: https://github.com/ZiyuanMadoka/code_dose_reconstruction_pipeline.

Fig. 1 The proposed pipeline for automated plan emulation. Note: the term “surrogate DRR” denotes the DRR derived from the surrogate CT.

2.1 Preprocessing: Parameters of Reference RT Plans

For historically treated patients, the dose [and monitor units (MUs)] was calculated manually using percental depth dose tables and plan parameters that define the geometrical setup, and prescribed doses of a particular plan were documented in the RT records. In modern TPSs, these plan parameters are saved in a standard format, i.e., digital imaging and communications in medicine (DICOM).³⁰ Below, we describe the plan parameters that we consider to be of importance in the plan emulation process.

The isocenter of a plan defines the point about which the collimator system and gantry of the linear accelerator (i.e., the device used for external beam RT) rotates³¹ (Fig. 2). The RT field is an area treated by the radiation beam at a particular gantry angle. In the case of an AP-PA beam setup, the gantry angle (θ_G) is set to 0 deg and 180 deg, for the AP and PA beams, respectively, and correspondingly two RT fields are generated. The collimator angle (θ_C) defines the orientation of the collimator coordinate system with respect to the gantry coordinate system [Fig. 2(b)]. The collimator system consists of asymmetric jaw pairs and multileaf collimators (MLCs). Per beam, two asymmetric jaw pairs in the collimator system are used to define the RT field size in a rectangular shape. The jaw consists of two pairs of opposing collimator blocks and their positions are defined as a list of field boundary distances to the isocenter, in left/right and top/bottom directions (when the collimator angle is 0 deg). In the past, the standard or irregular blocks of certain shapes were made, when there was a need to spare certain organs from receiving radiation. Nowadays, modern machines are equipped with MLCs, which consist of a large number of narrow, closely abutting individual leaves (usually 40 to 120, arranged in pairs). By moving and controlling the individual leaves of an MLC, one can generate almost any desired field shape. In our pipeline, MLCs are used to simulate the historical shielding.³²

2.2 Preprocessing: Generation of DRRs

An in-house-developed DRR calculation algorithm is utilized to automatically generate DRRs of both the reference and surrogate CTs in AP view. The algorithm mimics the historical kilovoltage radiograph by simulating a divergent beam through the CT volume using a ray-tracing algorithm.

Each reference CT is associated with a Wilms' tumor plan (reference plan), which was originally planned on the reference CT and is considered similar to a typical historical Wilms' tumor

Fig. 2 (a) An example of an RT field shaped by components of the collimator system is illustrated. In this case, an AP beam [in which case the gantry angle (θ_G) is 0 deg] is simulated and the geometry of the collimator system is projected on the patient's DRR in AP view. (b) Components of a linear accelerator (Elekta AB, Stockholm, Sweden) together with the patient coordinate system (x, y, z)_p, the collimator coordinate system (X, Y, Z)_c, a sketched RT field and an isocenter associated with a patient.

plan (as the RT field set-up for Wilms' tumor has not changed much over the years^{33,34}). The historical radiograph-like DRR is generated using the coronal plane which contains the plan isocenter (Fig. 2). To simulate the contrast of historical radiographs, which is the truly available data in dose reconstruction, the window level and window width are set to 1441 and 3200 Hounsfield Unit (HU), respectively, and no bone enhancement is applied.

To generate DRRs for surrogate CTs (the CTs on which the reference plan will be emulated), an isocenter is needed to direct the ray-tracing algorithm. Yet, a reference plan still needs to be emulated (the DRRs are needed first), and thus no isocenter is available. To overcome this, we automatically determine a "pseudo" isocenter, based on delineations of OARs. We choose the center point of the body contour between the 10th thoracic vertebra (T10) and the first sacral vertebra (S1) vertebral region. We found this heuristic to be sufficient in early experiments, as changing the isocenter with typical positions for abdominal RT results in similarly looking DRRs, and negligible deviations in landmark detection. Surrogate DRRs do not need to resemble historical radiographs, as they are ultimately used for plan emulation on the respective surrogate CTs. Therefore, to generate better bone-to-tissue contrast to ease landmark detection, the following settings are used to generate the surrogate DRRs (in HU): window level: 1441, window width: 3200, bone threshold: 200, and bone enhancement factor: 2.5.

All generated DRRs are then automatically cropped to a common region of interest from T10 to S1, using existing delineations of the spinal cord (a subregion of the spinal cord from vertebral bone T10 to S1) from the CT scans.

2.3 Automatic Landmark Detection in DRRs

Several image processing techniques are applied on the DRRs for automatic landmark detection. These landmarks will later be used to determine how to change the RT plan parameters, i.e., by comparison between the reference and surrogate DRR. An illustration of landmark detection is shown in Fig. 3.

To be consistent with the CT/patient coordinate system $(x, y, z)_p$ as shown in Fig. 2, we denote the axis along right-to-left (RL) of the patient as x , the axis along AP direction as y , the axis along the feet-to-head (FH) direction as z , and the image signal at point (x, z) on the coronal DRR as $I(x, z)$. Since y is the constant on a coronal plane, it is ignored in this section. We denote $V(x)$ as the sum of the signal along the z -direction (vertical), and $H(z)$ as the sum of the signal along the x -direction (horizontal)

$$V(x) = \sum_z I(x, z), \quad H(z) = \sum_x I(x, z).$$

Note that $V(x)$ can be thought of as a row-vector, and $H(z)$ can be thought of as a column-vector.

2.3.1 Horizontal cropping

Since patient arm positioning is inconsistent and may complicate landmark detection, we design a horizontal cropping step to eliminate the arms (if present) from the image [Fig. 3(a)]. The minimal $V(x)$ (corresponding to air outside the patient body) is applied to $V(x)$ as a threshold. Because a gap of air exists between the arms and the thorax and abdomen, we identify the latter two body parts as the set of x -values that is the largest connected component of x -values satisfying $V(x) > \min[V(x)]$. Next, we crop the image to this largest component, thereby excluding the arms, if present. We then apply a 3×3 pixel averaging filter to the newly cropped image, to suppress noise. This results in a cropped image that is the starting point for the next computations [Fig. 3(a)].

2.3.2 Initial estimation of the vertebral column borders

We want to make an initial estimation of the location of the vertebral column. For this, we want to use the high-intensity signals of the vertebrae [Fig. 3(b)]. However, also the iliac crests and the ribs have high-intensity signals. To make the identification of the vertebral column easier to

Fig. 3 An example case of the landmark detection steps. (a) The vertical orange lines represent the initial estimation of the vertebral column boundaries based on the peaks in (b) (computed for the center middle part of the top half of the radiograph). The blue and fuchsia horizontal lines in (a) indicate the results of the two rounds of estimating the intervertebral disc locations (segments between the vertebral bones) [see (c) and (f)]. The green solid circles indicate the middle-right and middle-left boundaries of the vertebral bones as estimated in (d) and (e). The black line fits the centers of the vertebrae, based on these identified boundaries. The solid red points in (a) were transformed from the estimated rib cage boundaries on the rotated DRR illustrated in (g). The T11, T12, L2, and L4 vertebrae are pointed out with arrows in (a).

obtain, only the center middle part [i.e., the two inner horizontal quarters of the image (to exclude the ribs nearby the body border)] of the top-half of the image (to exclude the iliac crests) is considered. This leads to $V_{\text{middle_top_half}}(x)$ as the vertically summed signals of the middle and top half of the image. We calculate its smoothed derivative as $\text{smooth}[V'_{\text{middle_top_half}}(x)]$, where $\text{smooth}()$ applies a five-point moving average filter to the signal series.³⁵ Peak detection,³⁶ an algorithm to find the positions corresponding to local maxima in a series of signals, is then applied to $\text{smooth}[V'_{\text{middle_top_half}}(x)]$. The two detected peak locations $\{x_{\text{init_vc_R}}, x_{\text{init_vc_L}}\}$ give an initial estimation of the right and left borders of the vertebral column.

2.3.3 Initial estimation of the intervertebral disc locations

We denote the horizontally summed signals within the interval $[x_{\text{init_vc_R}} - 6, x_{\text{init_vc_L}} + 6]$ as $H_{\text{init_vc}}(z)$. The 6-mm expansion of the initial vertebral column estimations on both sides is applied to account for vertebrae tilting and uncertainty of the initial estimation. We calculate the smoothed negation of the signals as $-\text{smooth}[H_{\text{init_vc}}(z)]$, on which peak detection is applied to identify the eight intervertebral discs [Fig. 3(c)]. The estimations of intervertebral disc locations are denoted by $\{z_{\text{init_id_i}} | i = 1, \dots, 8\}$.

2.3.4 Right and left borders estimation for each vertebral bone

The right and left borders of vertebral bones can differ due to vertebrae tilting and difference in vertebrae size [Figs. 3(d) and 3(e)]. To accurately detect each vertebral border, we define $V_{\text{vb_i}}(x)$ ($i \in \{1, \dots, 7\}$) as the vertically summed signals within each segment defined by the estimated vertebral column $\{x_{\text{init_vc_R}}, x_{\text{init_vc_L}}\}$ with 20 to 25 mm expansion on both sides and

the initially estimated intervertebral disc locations $\{z_{\text{init_id_}i}|i = 1, \dots, 7\}$. For each segment, we apply peak detection on the smoothed derivative of $V_{vb_i}(x)$, i.e., $\text{smooth}[V'_{vb_i}(x)]$, which results in $\{x_{\text{init_vb_}R_i}, x_{\text{init_vb_}L_i}\}$. We further apply $\text{smooth}()$ to the detected right and left borders, respectively, to suppress noise. The ultimate estimation of the right and left borders of the vertebral bones is denoted by: $\{x_{vb_R_i}, x_{vb_L_i}|i = 1, \dots, 7\}$.

2.3.5 Second estimation of the intervertebral disc locations

A second round of detection of intervertebral disc locations is performed on the piecewise function $H_{vc}(z)$, that accumulates the summed row signals from each segment defined by the estimated right and left vertebral bone borders and the initial estimations of the intervertebral disks [Fig. 3(f)]. Peak detection is applied to the negation of the smoothed $H_{vc}(z)$ [i.e., $-\text{smooth}(H_{vc}(z))$]. The final result is denoted by $\{z_{id_i}|i = 1, \dots, 8\}$.

2.3.6 Vertebral column tilt angle estimation

We estimate the vertebral column tilt angle on the coronal plane by fitting a line to the centers of the vertebrae defined by the right and left vertebral bone borders and intervertebral discs (i.e., $\{(x_i, z_i) = (\frac{x_{vc_R_i} + x_{vc_L_i}}{2}, \frac{z_{id_i} + z_{id_i+1}}{2})|i = 1, \dots, 7\}$), black line in Fig. 3(a)). The slope of the linear fitting function $\hat{x} = az + b$ is used to calculate the tilt angle of the vertebral column: $\text{angle} = \arctan(a)$.

The vertebral column tilt angle is used to rotate the right and left vertebral bone borders of T10 to T12 and L1 (first lumbar vertebra) to L4 (fourth lumbar vertebra) to estimate these borders accurately [green dots in Fig. 3(b)].

2.3.7 Estimation of right-most/left-most rib boundaries

When the patient is not lying straight, the estimation of the rib boundary based on vertically summed signals can be too coarse and imprecise [Fig. 3(g)]. To deal with this situation, we apply a rotation to the whole image $I(x, z)$ by the estimated tilt angle such that the vertebral column appears straight on the rotated image $I^R(x^R, z^R)$. The summed column signals of the top half image I^R are calculated as $V_{\text{top_half}}^R(x^R)$. To detect the horizontal outer boundaries of the ribs, a threshold of $\min[V_{\text{top_half}}^R(x^R)] + 0.3\{\max[V_{\text{top_half}}^R(x^R)] - \min[V_{\text{top_half}}^R(x^R)]\}$ is then applied on $V_{\text{top_half}}^R(x^R)$. The z -value of the intervertebral disc located below T12 (the third intervertebral disc in our cropped image) is used for the rib boundaries such that the two landmarks representing the right-most/left-most rib boundaries ($lm_{\text{rib_}R}^R$ and $lm_{\text{rib_}L}^R$) are estimated as: $lm_{\text{rib_}R}^R = (x_{\text{rib_}R}^R, z_{id_3}^R)$, $lm_{\text{rib_}L}^R = (x_{\text{rib_}L}^R, z_{id_3}^R)$.

The two points are then mapped back to the original image I [red points in Fig. 3(a)]. The landmark detection described in this section is implemented in MATLAB, by using functions from the Signal Processing Toolbox.

We remark that parameters for each peak detections (e.g., minimal peak vertical drop on both sides and minimal distance between two peaks) are manually tuned on 32 DRRs, to enable reliable results accounting for anatomical variations on DRRs. Constraints were also added based on the initial results of some steps to filter out unrealistic locations. Detailed information about the input signal, the tuned parameters, and constraints used for peak detection to each specific step is summarized in Table 5.

2.4 Plan Emulation from Landmarks

The plan emulation is achieved by a step-by-step estimation of the geometrical plan parameters. Geometrical transformation/modeling techniques are used to adapt the reference plan based on size and orientation differences of the reference and surrogate anatomies derived from the corresponding measurements and landmarks on DRRs. In addition to the running texts that describe the plan emulation steps, one example case of plan emulation is shown in Fig. 4.

Fig. 4. An example case of the plan emulation steps. The green dots are the landmarks used for this plan. The red dot indicates the isocenter of the plan. The circles with equations refer to the equations described in Sec. 2.4.

2.4.1 Measurements and landmarks

Several measures and landmarks are considered in the plan emulation. The vertebral column length from T11 to L4 is considered to scale the field size of the plan along the vertical dimension of the field. To account for possible differences in the dimensions of the patient's right and left sides as visible in the DRR (mainly caused by a slight rotation of the patient around the FH direction during scanning), we measure the distances from the right-most and the left-most rib boundaries to the middle bottom point of L1 (on the centerline of the vertebral column) separately, to derive lateral scaling factors (which will be described in the following sections). The vertebral column tilt angle is considered for collimator orientation. The middle-right and

middle-left points on the boundaries of T12 and L2 and the right-most and left-most rib boundaries are selected as landmarks, to estimate the position of the isocenter. Note that to this end, the coordinates of these landmarks need to be mapped from the DRR to the CT coordinate system (in which the plan properties are defined).

2.4.2 Selection of landmarks

If the reference plan concerns treatment of a Wilms' tumor in the left kidney (i.e., is a left-sided plan), we select the middle-right boundaries of T12 and L2, the middle-bottom boundary of L4 (which is the intersection point of the fitted line and the L4/L5 intervertebral disc described in Sec. 2.3), and the left rib boundary as landmarks. If the reference plan is a right-sided plan, we select the middle-left boundaries of T12 and L2, the middle-bottom boundaries of L4, and the right rib boundary as landmarks.

2.4.3 Scaling factors

We account for dissimilarities between the reference and the surrogate DRRs in terms of tilt angle, RL and FH scaling factors, with

$$\Delta\theta = (\theta_{\text{sur}} - \theta_{\text{ref}}), \quad (1)$$

$$S_{\text{WR}} = \frac{W\text{Rib}_{\text{sur}}}{W\text{Rib}_{\text{ref}}}, \quad S_{\text{WL}} = \frac{W\text{Lrib}_{\text{sur}}}{W\text{Lrib}_{\text{ref}}}, \quad (2)$$

$$S_L = \frac{Lvc_{\text{sur}}}{Lvc_{\text{ref}}}, \quad (3)$$

where θ is the tilt angle of the vertebral column, and $\Delta\theta$ is the tilt angle difference between the surrogate (sur) and reference (ref) DRR. $W\text{Rib}$ and $W\text{Lrib}$ are the right and left rib cage widths, respectively. Lvc denotes the vertebral column length from T11 to L4.

2.4.4 Collimator angle

The collimator angle of the AP beam in the emulated plan can now be set to the collimator angle of the AP beam in the reference plan plus $\Delta\theta$. Similarly, $-\Delta\theta$ is used for the PA beam.

2.4.5 Isocenter (x, z)

Given a landmark point lm , an estimation of the isocenter coordinates (x, z) in the coronal plane for the emulated plan is given as

$$\begin{aligned} \begin{bmatrix} x_{\text{sur}}^{\text{iso}} \\ z_{\text{sur}}^{\text{iso}} \end{bmatrix} &= R_y(\theta_{\text{sur}}) \left(\begin{bmatrix} S_W & 0 \\ 0 & S_L \end{bmatrix} R_y(-\theta_{\text{ref}}) \begin{bmatrix} x_{\text{ref}}^{\text{iso}} - x_{\text{ref}}^{lm} \\ z_{\text{ref}}^{\text{iso}} - z_{\text{ref}}^{lm} \end{bmatrix} \right) + \begin{bmatrix} x_{\text{sur}}^{lm} \\ z_{\text{sur}}^{lm} \end{bmatrix} \\ \text{with } S_W &= \begin{cases} S_{WL}, & \text{if } x_{\text{ref}}^{\text{iso}} - x_{\text{ref}}^{L1} > 0 \\ S_{WR}, & \text{otherwise} \end{cases}, \end{aligned} \quad (4)$$

where $\begin{bmatrix} x_{\text{sur}}^{\text{iso}} \\ z_{\text{sur}}^{\text{iso}} \end{bmatrix}$ and $\begin{bmatrix} x_{\text{sur}}^{lm} \\ z_{\text{sur}}^{lm} \end{bmatrix}$ are the coordinates of the isocenter and the landmark point in the CT coordinate system, respectively. The scaling factor of the right- or left-sided rib width (i.e., S_{WR} or S_{WL}) is applied depending on whether the isocenter is located on the right or left flank of the patient. $R_y(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ is the rotation matrix for rotations around the y-axis, used to correct for the influence of the tilt of the vertebral column on the relative difference in location between the isocenter and landmarks. Therefore, in summary, the difference in location of the isocenter with respect to the landmark is computed, rotated, and scaled, to remove tilting effects [terms in the brackets of Eq. (4)]; the result is then rotated by the angle with which

the surrogate vertebral column is tilted compared to the reference [the left-most term $R_y(\theta_{\text{sur}})$] and added to the coordinates of the landmark point in the surrogate DRR (right-most term).

2.4.6 Isocenter (y)

The y coordinate of the isocenter is set according to the ratio of skin-to-isocenter distance along the AP direction as reported in the RT record. After the (x, z) coordinates of the isocenter are estimated with Eq. (4), we compute the body diameter along the corresponding sagittal plane (on the surrogate CT scan) and apply the aforementioned ratio to estimate the y coordinate.

We repeat the computation of the isocenter for each of the four landmarks and finally take the average as final result, to improve the robustness of the method.

2.4.7 Field borders

The field borders B and MLC leaf positions are defined in the collimator coordinate system $(X, Y, Z)_C$ (Fig. 2).³¹ Note that a vector defined in the patient coordinate system $(x, y, z)_p$ can be transformed to the collimator coordinate system by applying a rotation matrix of the gantry angle and collimator angle as: $(\Delta X, \Delta Y, \Delta Z)_C = R_y(\theta_G)R_z(\theta_C)(\Delta x, \Delta y, \Delta z)_p$, where R_y and R_z are rotation matrices for rotations around the y and the z axes, respectively.

For a typical abdominal AP field ($\theta_G = 0$ deg) where the field orientation follows the vertebral column ($\theta_C \cong 0$), the field borders along Y are denoted by $\{Y^{\text{bottom}}, Y^{\text{top}}\}$, with Y^{bottom} referring to the distance from the isocenter to the field bottom border (patient's feet direction), and Y^{top} represents the distance from the isocenter to the field top border (patient's head direction). Similarly, the field borders along X are denoted by $\{X^{\text{right}}, X^{\text{left}}\}$, with X^{right} referring to the distance from the isocenter to the field right border (patient's right side) and X^{left} referring to the distance from the isocenter to the field left border (patient's left side). Each field border of the emulated plan $B_{\text{sur}} \in \{X_{\text{sur}}^{\text{right}}, X_{\text{sur}}^{\text{left}}, Y_{\text{sur}}^{\text{bottom}}, Y_{\text{sur}}^{\text{top}}\}$ of the emulated plan is calculated by scaling the respective field border of the reference plan B_{ref} by the appropriate scaling factor (S_{WR} or S_{WL} for X , S_L for Y)

$$B_{\text{sur}} = S_B B_{\text{ref}}, \quad B \in \{X^{\text{right}}, X^{\text{left}}, Y^{\text{bottom}}, Y^{\text{top}}\}$$

$$\text{with } S_B = \begin{cases} S_{WR}, & \text{if } B \in \{X^{\text{right}}, X^{\text{left}}\} \text{ and plan is right-sided} \\ S_{WL}, & \text{if } B \in \{X^{\text{right}}, X^{\text{left}}\} \text{ and plan is left-sided} \\ S_L, & \text{if } B \in \{Y^{\text{bottom}}, Y^{\text{top}}\} \end{cases} \quad (5)$$

2.4.8 MLC (shaping of block)

In reference plans, MLCs are used as a substitution of historical blocks, where the beam shaping is achieved by properly placing the leaves of the MLC to resemble the shaping of a block. The width of the leaves we simulate is 1 cm. The leaves are placed such that the block outline touches the leaves' middle point [as show in Fig. 4(R-1)]. In the following, we consider leaves moving in the X direction to explain the block emulation steps.

To capture the shape approximated by the MLC, and later reproduce this in the emulated plan, we fit the X coordinates of the leaves that shape the block in the reference plan with a polynomial function $\hat{X} = \hat{f}(Y)$, where the Y coordinates of these leaves are fixed. Depending on the number n_{ref} of in-field leaves [i.e., the i 'th leaf position $(X_{\text{ref}_i}, Y_{\text{ref}_i})$ satisfies $X_{\text{ref}_i} \in (X_{\text{ref}}^{\text{right}}, X_{\text{ref}}^{\text{left}})$ and $Y_{\text{ref}_i} \in [Y_{\text{ref}}^{\text{bottom}}, Y_{\text{ref}}^{\text{top}}]$] that shape the block [note that leaves can be out of field as in Fig. 2(a)], the polynomial function is fitted (with ordinary least squares) as

$$\hat{X}_{\text{ref}_i} = \hat{f}_{\text{ref}}(Y_{\text{ref}_i}) = \begin{cases} p_1 Y_{\text{ref}_i}^2 + p_2 Y_{\text{ref}_i} + p_3, & \text{if } n_{\text{ref}} > 3 \\ p_1 Y_{\text{ref}_i} + p_2, & \text{otherwise} \end{cases}, \quad (6)$$

where $i \in \{1, 2, \dots, n_{\text{ref}}\}$ is the i 'th in-field leaf, and p_1 , p_2 , and p_3 are the free parameters to fit the function.

To emulate the shaping of the block by the MLC in the surrogate CT, the parameters of the polynomial function are scaled based on S_{WR} or S_{WL} depending on the plan side, and S_Y . Similar to Eq. (5), we assume

$$Y_{\text{sur}} = S_L Y_{\text{ref}} \text{ and } X_{\text{sur}} = \begin{cases} S_{WR} X_{\text{ref}}, & \text{if plan is right-sided} \\ S_{WL} X_{\text{ref}}, & \text{if plan is left-sided} \end{cases}. \quad (7)$$

So the polynomial function in Y_{ref} coordinates is scaled to the Y_{sur} coordinates as

$$\hat{f}_{\text{sur}}(Y_{\text{sur}}) = \begin{cases} \frac{S_L}{S_W^2} p_1 Y_{\text{sur}}^2 + \frac{S_L}{S_W} p_2 Y_{\text{sur}} + S_Y p_3, & \text{if } n_{\text{sur}} > 3 \\ \frac{S_L}{S_W} p_1 Y_{\text{sur}} + S_Y p_2, & \text{otherwise} \end{cases}, \quad (8)$$

where S_W is S_{WR} or S_{WL} , depending on if it concerns a right-sided or left-sided plan, and n_{sur} is the number of leaves that shape the block on the surrogate CT. After determining the $\{Y_{\text{sur}_j} | j = 1, 2, \dots, n_{\text{sur}}\}$ of the in-field leaves ($Y_{\text{sur}_j} \in [Y_{\text{sur}}^{\text{bottom}}, Y_{\text{sur}}^{\text{top}}]$) that shape the block on the surrogate CT, we can then estimate the corresponding X coordinates of the j 'th leaf as

$$\hat{X}_{\text{sur}_j} = \hat{f}_{\text{sur}}(Y_{\text{sur}_j}), \quad j \in \{1, 2, \dots, n_{\text{sur}}\}. \quad (9)$$

2.5 Dose Calculation and Plan Visualization

The emulated plan parameters are saved in the DICOM RT plan file format. We use a scriptable dose calculation module of Oncentra TPS (version 4.3, Elekta, Stockholm, Sweden) that, given the DICOM data, containing the CT, a plan, and a DICOM RT structure (which stores all the organ delineations) of the surrogate patient, computes the 3-D dose distribution. We scale the MU of the two beams to ensure that the dose at isocenter is the same (14.4 Gy) in the surrogate CT and in the reference CT.

We include an automatic plotting module in our pipeline, to enable quick visualization of the emulated plan on the surrogate DRR. In this module, a DRR in the beam's eye view is generated based on the isocenter of the emulated plan. The effective field shape is plotted on top of the DRR as solid lines. The block made by the use of the MLC is visualized as a solid line that connects the middle location of each leaf's end [e.g., Figs. 4(R-1) and 4(S-5)].

The manipulation of DICOM files described in this section is implemented in python, mostly using the package *pydicom*.³⁷

3 Data and Evaluation

3.1 Data

We considered abdominal RT planning CT scans of patients in supine position between 2 and 6 years old, each including a common abdominal region from T10 to S1. The patients underwent RT between 2002 and 2018 at the Amsterdam UMC, location AMC/Emma Children's Hospital (Amsterdam UMC) ($n = 32$) or at the University Medical Center Utrecht/Princess Máxima Center for Pediatric Oncology (UMC Utrecht) ($n = 39$) for abdominal cancer. Reference data and two sets of validation data were formed as described below.

3.1.1 Reference data

We selected five historical-like Wilms' tumor RT plans associated with five patients' CTs as reference data. The plan set-up includes two beams irradiating the left (two plans) or the right (three plans) flank from AP and PA directions. These plans are clinically approved and were created according to the clinical SIOP WT 2001 protocol,³⁸ in which the treatment field of a typical left-sided Wilms' tumor covers the major part of the left flank including the vertebral column, the spleen, and the left part of the liver, whereas the treatment field of a typical

right-sided Wilms' tumor covers the tumor region including the vertebral column, the iliac crest, and major parts of the right liver. These plans are representative of historical treatment plans, since the RT field treatment protocol described in clinical SIOP WT 2001 did not change significantly from the pre-1990 era.^{33,39}

The plans have different geometries regarding the cranial border location of the RT field (starting from T8, T11, or T12) and the blocked region border (rib 8, rib 10, or rib 11). The isocenter positions with respect to the landmarks also vary according to each patient's tumor bed location.

3.1.2 Validation data

Data set 1. The first data set contains 27 CTs with two subgroups based on whether an intact left or right kidney is present in the CT. Both subgroups consist of 20 CTs each (as 13 CTs are with two intact kidneys). Each reference plan has been manually emulated on the 20 surrogate CTs that have an intact kidney on the same side as the reference CT. These 100 manually emulated plans were made by crafting the plan parameters (Sec. 2.1) in the TPS, by visually comparing the beam's eye view DRRs of the reference and surrogate CTs, under the supervision of and approval by a radiation oncologist (BVB), as described in our previous work.^{17,26}

For data set 1, we performed a direct comparison between the automatically emulated plans and the manually emulated plans. We remark that data set 1 was used to test and refine our landmark detection and plan emulation approach (Sec. 2.3.). Therefore, the results of our automatic plan emulation pipeline on this data set could be positively biased.

Data set 2. The second surrogate data set contains 37 CTs for which manual plan emulations are not available. Therefore, no comparison between manual and automatic plan emulation can be done for this data. Nonetheless, these data have not been used in the process of developing the pipeline and can therefore be used to assess whether our approach generalizes well to new data (no positive bias). We applied the automatic pipeline to this data set and perform qualitative evaluations of the resulting plans.

A summary of the reference data and two validation (i.e., surrogate) data sets can be found in Table 1.

3.2 Evaluation of Automatic Plan Emulations

We validated our approach by performing two qualitative validations that comprised asking two radiation dosimetrists to do two assessments, one of the quality of automatic and manual emulations in a blinded comparison and one of the quality of automatic emulations on a different data set; and two quantitative assessments, one on geometry discrepancy between automatic and manual emulations and one on dose reconstruction differences between automatic and manual emulations.

3.2.1 Blinded observer assessment of manually and automatically emulated RT plans

To assess the quality of the emulated plans, two experienced radiation dosimetrists evaluated the emulated plans by visually comparing the effective field shape of the reference plan and the emulated plans plotted on top of the reference and surrogate DRRs, respectively. Data set 1 was used for this purpose, where for each surrogate CT a manually and an automatically emulated plan were available. We prepared the plotted fields with our automatic plan visualization tool and asked the radiation dosimetrists to grade each emulated plan. Crucially, we did not specify which plans were emulated automatically and which manually. Three grades were possible: "1" means the emulated plan is of good quality, "2" means the emulated plan is of sufficient quality, but details can be improved, and "3" means the emulated plan is of insufficient quality. For the latter two cases, the radiation dosimetrists were asked to list the imperfections that justify the decision.

Table 1 Overview of reference data and validation data. For reference data, each patient's age and gender as well as the field features of the reference RT plans are included. F, female; M, male; T1 through T12 represent the 12 thoracic vertebrae; L1 through L5 represent the five lumbar vertebrae; S1 through S5 represent the sacral vertebrae; D1, data set 1; D2, data set 2.

Reference data							
Case	Age	Gender	Tumor site	Field	Field border cranial/caudal/ caudal borders	Blocked region	Source
1	3.8	M	Right kidney	Right	T12/L4	Right rib 11	Amsterdam UMC
2	3.9	F	Right kidney	Right	T8/L4	Right rib 8-T9	
3	4.2	F	Left kidney	Left	T10/L4	Left rib 10	
4	4.7	M	Right kidney	Right	T11/L4	Right rib 10	
5	5.2	M	Left kidney	Left	T11/L5	Left rib 10	
Surrogate data							
	N total	Age (years) mean (range)	Gender (N: F/M)	N intact right kidney	N intact left kidney	Source	
D1	27	3.7 (2.2 to 5.6)	16/11	20	20	Amsterdam UMC (<i>n</i> = 25), UMC Utrecht (<i>n</i> =2)	
D2	37	3.7 (2.0 to 6.0)	18/19	25	29	UMC Utrecht	

3.2.2 Observer assessment of automatically emulated plans

We used the same reference data and applied our plan emulation pipeline to data set 2. Each reference plan was automatically emulated on 10 surrogate CTs resulting in 50 emulated plans. We asked the radiation dosimetrists to grade these automatic emulations similarly to our first experiment (Sec. 3.2.1), the only difference being that manual emulations are not present.

3.2.3 Quantitative plan parameter differences

We investigated differences in plan parameters between the automatically emulated plans and the manually emulated plans quantitatively. To this end, the plan isocenter position, the field sizes, and the collimator angle of the AP field were compared.

3.2.4 Quantitative dose comparison

We compared the reconstructed dose distribution on each surrogate CT based on manually emulated plans and automatically emulated plans. We considered metrics typically employed in epidemiological studies on late adverse effects: organ mean dose (D_{mean}) and the minimum dose received by the most exposed 2 cm³ ($D_{2\text{cc}}$), for liver, spleen, contralateral kidney, and spinal cord. We performed a two-sided Wilcoxon signed-rank test paired by patient to establish whether automatic and manual plan emulations lead to different dose metrics (alternative hypothesis).⁴⁰

To understand the difference in terms of dose reconstruction accuracy (with respect to the reference dose), we calculated dose metric deviations (i.e., the reconstructed dose metrics minus the reference dose metrics) for D_{mean} and $D_{2\text{cc}}$ and will refer to these as DE_{mean} and $\text{DE}_{2\text{cc}}$, respectively. To test if our automatic pipeline leads to a worse dose reconstruction accuracy, we performed a one-sided Wilcoxon signed-rank test on the magnitude of the dose metric deviations (i.e., $|\text{DE}_{\text{mean}}|$ and $|\text{DE}_{2\text{cc}}|$) associated with the automatic plan emulations and manual plan

emulations. The alternative hypothesis for this test is that the automatic plan emulations lead to larger mean absolute dose metric deviations with respect to the reference dose metric. Note that we analyzed the OAR dose metrics for left-sided and right-sided plans separately because dose distributions to different organs considered here very much depend on the side of the plan. The spinal cord is an exception to this due to its central location and the fact that it is always an in-field OAR.

4 Results

4.1 Visual Evaluation of Plan Emulations

The plan evaluation results from the two radiation dosimetrists are summarized in Table 2. Out of the 100 cases based on data set 1, radiation dosimetrist A (PG) approved 100% of both automatic and manual emulations. Radiation dosimetrist B (KFC) approved 92% of automatic and 91% of manual emulations. Further, radiation dosimetrist A gave grade 2 (plan approved but with small flaws) to 14% and 16% of the automatic and manual emulations, respectively; while radiation dosimetrist B gave this grade to 37% and 20% of the automatic and manual emulations, respectively. In Fig. 5, RT plan emulations are illustrated for five cases. Interestingly, in some cases automatic plans scored better than manual ones (e.g., Fig. 5, row 5).

Out of the 50 automatically emulated cases based on data set 2, radiation dosimetrist A approved 100% of the emulations, of which 32% of the cases were assessed with a grade 2. For radiation dosimetrist B, the approval rate was 94%, and 36% cases received grade 2.

Radiation dosimetrist B found a similar number of plans of sufficient quality for manual and automatic plan emulations, but there are more cases assessed as grade 2 for the automatic plan emulations. Radiation dosimetrist A instead graded them very similarly. Although evaluation of the automatically emulated plans by dosimetrist A resulted in a similar approval rate (100%) for the emulations based on data set 2 and for the emulations based on data set 1, the ratio of assessing plan emulations with grade 2 was found to be twice as large for the emulations based on data set 2 compared to the emulations based on data set 1. The ratios of the three different grades given by dosimetrist B for the emulations based on data set 2 were found to be very similar to the ratios of the grades given for the emulations based on data set 1.

4.2 Quantitative Comparison of Geometrical Plan Parameters

The average difference and standard deviation of multiple plan parameters between manually and automatically emulated plans are summarized in Table 3. On average, the difference in iso-center location between the automatically emulated plans and the manually emulated plans is 3.1 mm (with standard deviation 1.9 mm). The field size along RL and FH directions differs by ~4 mm in both directions, and the collimator angle differs on average by 1.4 deg of the automatically emulated plans compared with the manually emulated plans.

Table 2 Summary of plan evaluation results for data set 1 and data set 2.

Number of cases	Data set 1						Data set 2		
	Automatically emulated plans			Manually emulated plans			Automatically emulated plans		
Grade	1	2	3	1	2	3	1	2	3
Radiation dosimetrist A	86	14	0	84	16	0	34	16	0
Radiation dosimetrist B	55	37	8	71	20	9	29	18	3

Reference plans on top of reference DRRs	Emulated plans on top of surrogate DRRs with evaluation grades given by two radiation dosimetrists (A/B)	
	automatic	manual
Plan (1) Reference 1	Grade: 2/1	Grade: 1/2
Plan (2) Reference 2	Grade: 1/3	Grade: 1/3
Plan (3) Reference 3	Grade: 1/2	Grade: 2/1
Plan (4) Reference 4	Grade: 1/1	Grade: 1/1
Plan (5) Reference 5	Grade: 2/2	Grade: 2/3

Fig. 5 Examples of reference and emulated plans visualized on DRRs. First column: The reference plans on the reference DRRs. Second column: Automatically emulated plan on one of the surrogate DRRs. Third column: Manually emulated plan on the same surrogate DRR as used for the automatic emulation. The red point indicates the isocenter position. The yellow solid lines (partly overlapped by the blue solid lines) indicate the field borders along the orthogonal coordinate axes of the collimator coordinate system. The blue solid lines indicate the effective shape of the field, including the blocking performed by MLC. Below each emulated plan the grade given by radiation dosimetrist A/radiation dosimetrist B is shown.

4.3 Quantitative Comparison of Dose Reconstruction

In Table 4, the average difference and standard deviation of organ D_{mean} and $D_{2\text{cc}}$ between automatic and manual plan emulations, as well as the p values of the Wilcoxon tests for the alternative hypothesis that automatically and manually emulated plans lead to different dose reconstruction results, are given. The dose at isocenter is scaled to 14.4 Gy by scaling the MU of the beams for all the plans for the sake of comparison. The average difference in D_{mean} between automatically and manually emulated plans for all OARs is <0.9 Gy, i.e., 6% of the prescribed dose of 14.4 Gy. The average difference in $D_{2\text{cc}}$ is <0.8 Gy, except for the left and right kidneys for which a relatively large difference was found (2.66 and 1.80 Gy, respectively). For the left-sided plans, the mean dose for the liver and the right kidney and

Table 3 Absolute differences in plan parameters of the automatically emulated plans compared with the manually emulated plans for the 100 reconstruction cases performed based on data set 1 (average value and standard deviation between brackets).

Plan parameters	Average (standard deviation)			
Collimator angle (deg)	1.4 (1.0)			
Isocenter location (mm)	Euclidean	dx	dy	dz
	3.1 (1.9)	2.3 (1.8)	0.4 (0.3)	1.7 (1.3)
Field size along RL (mm)	$d X^{\text{left}} - X^{\text{right}} $	$d X^{\text{right}} $	$d X^{\text{left}} $	
	3.5 (2.5)	2.9 (2.0)	2.2 (3.0)	
Field size along FH (mm)	$d Y^{\text{top}} - Y^{\text{bottom}} $	$d Y^{\text{bottom}} $	$d Y^{\text{top}} $	
	4.3 (3.3)	3.0 (2.8)	2.2 (1.8)	

the D_{2cc} for the right kidney are found to be significantly different between the automatically reconstructed doses and the manually reconstructed doses. For the dose metrics for the other OARs, the p value of the Wilcoxon test is found to be ≥ 0.08 , which rejects the alternative hypothesis that the two methods lead to different estimations of the dose metrics.

Table 4 also includes the average and standard deviation of the magnitude of the organ DE_{mean} and DE_{2cc} of automatic and manual plan emulations with respect to the reference dose. The p values of the Wilcoxon test for the alternative hypothesis that automatic plan emulation leads to worse dose reconstruction accuracy (with respect to the reference dose) than manual plan emulation are also included. For the left-sided plans, except for $|DE_{\text{mean}}|$ and $|DE_{2cc}|$ for the liver and $|DE_{2cc}|$ for the right kidney, p values ≥ 0.1 were found. For the right-sided plans, p values ≥ 0.07 were found. For cases when p values are relatively large (typically values below 0.05 are considered statistically significant), the hypothesis that the automatic method leads to worse dose metric accuracies can be rejected. By looking at the average $|DE_{\text{mean}}|$ and $|DE_{2cc}|$ in these cases, sometimes the automatic method leads to smaller values than the manual method. Moreover, although for left-sided plans p values smaller than 0.05 were found for $|DE_{\text{mean}}|$ and $|DE_{2cc}|$ for the liver, the actual difference in dose metric deviations is < 0.2 Gy. For D_{2cc} for the right kidney, the difference in $|DE_{2cc}|$ is 1.34 Gy. By observing the boxplots and the distributed points of DE_{mean} in Fig. 6, there is no obvious evidence showing superiority of automatic or manual reconstruction cases in organ DE_{mean} . However, small differences can be observed. For example, the distribution of DE_{mean} values for the liver for left-sided plans of automatic reconstructions covers a range that is slightly lower than that of the manual reconstructions, which indicates that a systematic small underestimation is likely for automatic reconstructions compared to manual reconstructions.

5 Discussion

We are the first to leverage image processing techniques to automate RT plan emulation based on anatomical features available from 2-D radiographs, a key component of RT dose reconstruction for historical plans. Overall, the results of our extensive validation show that our automatic plan emulations can be considered to be of similar quality to the ones made manually. Therefore, our pipeline enables automatic large-scale 3-D radiation dose reconstruction.

In the case of plan emulation on a surrogate anatomy, validation is not straightforward to perform. First, it is hard to define what a “good” plan emulation on a different anatomy is. In this study, bony anatomies are used as landmarks to assess the relative location and shape of the treatment field.^{12,26} However, in some cases, a “good” plan emulation may simply not exist if the surrogate anatomy is too different from the reference anatomy that the plan was originally designed for. For example, in row 2 in Fig. 5, both the manually and automatically emulated plan

Table 4 For the automatically and manually emulated plans based on data set 1, the average (and standard deviation) of absolute differences in the reconstructed organ dose metric for several OARs is provided. Further, for both the automatically and the manually emulated plans, the deviations with respect to the reference dose metrics are provided for several OARs. The results of the reconstructed dose metrics for the OARs in left-sided and right-sided plans are reported separately except for the spinal cord. Also, p values of two statistical tests are provided. (p values less than 0.05 were indicated by bold characters) Test 1 is for the alternative hypothesis that dose metrics are different between automatic and manual emulations. Test 2 is for the alternative hypothesis that the magnitude of dose metric deviations with respect to the reference is larger for automatic emulations than for manual emulations. auto, automatically reconstructed cases; manual, manually reconstructed cases; diff (auto, manu), difference between automatically and manually reconstructed cases.

Average absolute difference (standard deviation)											
		Average absolute difference (standard deviation)		p value (test 1)		Average (standard deviation)				p value (test 2)	
Plan side	OAR	D _{mean} (Gy)	D _{20cc} (Gy)	D _{mean}	D _{20cc}	DE _{mean} (Gy)		DE _{20cc} (Gy)		DE _{mean}	DE _{20cc}
						Auto	Manual	Auto	Manual		
Left-sided plans (n = 40)	Liver	0.28 (0.21)	0.08 (0.07)	1e-5	0.61	1.25 (0.76)	1.02 (0.65)	0.27 (0.20)	0.25 (0.17)	3e-5	0.03
	Spleen	0.89 (0.73)	0.27 (0.45)	0.62	0.3	2.62 (1.53)	2.52 (1.66)	1.28 (2.83)	1.16 (2.88)	0.3	0.1
	Right kidney	0.41 (0.29)	1.80 (1.32)	0.01	3e-5	0.52 (0.64)	0.57 (0.60)	3.06 (2.11)	1.72 (1.53)	0.66	2e-5
Right-sided plans (n = 60)	Liver	0.35 (0.27)	0.08 (0.07)	0.58	0.42	2.06 (1.22)	2.02 (1.09)	0.27 (0.22)	0.28 (0.20)	0.37	0.89
	Spleen	0.10 (0.12)	0.72 (1.29)	0.53	0.34	0.18 (0.18)	0.16 (0.16)	2.04 (2.45)	2.03 (2.60)	0.29	0.07
	Left kidney	0.71 (0.66)	2.66 (2.95)	0.61	0.76	0.66 (0.44)	0.67 (0.62)	3.72 (4.27)	3.79 (4.00)	0.44	0.43
Both (n = 100)	Spinal cord	0.22 (0.23)	0.07 (0.06)	0.67	0.08	0.42 (0.32)	0.45 (0.32)	0.23 (0.17)	0.23 (0.17)	0.85	0.31

on the surrogate CT are graded as “not approved” by radiation dosimetrist B, as the caudal field border crosses the iliac bones in the surrogate DRR, which is not the case in the reference DRR. However, it is not possible to avoid crossing the iliac bones while keeping the relative field shape with respect to other anatomical landmarks. Second, observer variation exists in assessing the quality of an emulated plan. Taking the same example case we just discussed, different from radiation dosimetrist B, radiation dosimetrist A positively assessed these two emulations. Also, the example case for reference plans (3), (4), and (5) in Fig. 5 received different grades from the two radiation dosimetrists. Based on the reasons they listed, we observed that radiation dosimetrist A is more strict on the similarity of the field border to spine distance (i.e., field size along RL with respect to bony anatomies), while radiation dosimetrist B is more strict on block placement with respect to the ribs.

Despite these difficulties, we nonetheless attempted to provide a comprehensive validation based on reasonable criteria. The evaluation results in general show that the automatic emulations and manual emulations perform similarly well, both achieving good performance (on average >95% approval rate). More specifically, based on the reasons given by the radiation dosimetrists for cases that were graded as “2” and “3,” we found that manually emulated plans had a better similarity in plan field borders (especially along the lateral direction) with respect to bony anatomy, while automatically emulated plans better captured the shape of the blocking. For quantitative evaluation, we reported statistics of plan parameter differences between automatically and manually emulated plans. The relatively low differences (<5 mm) found in isocenter position and field size do not exceed the inspiration-induced diaphragm motion in pediatric patients which is on average 10.7 mm (range 4.1 to 17.4 mm) during treatment.⁴¹ This is a further reason to support the idea that automatic and manual plan emulations are, in fact, similarly accurate.

As the ultimate goal of our study is to achieve accurate dose reconstruction, we further compared the dose reconstruction accuracy (in terms of deviation of reconstructed dose metrics from the reference ones) between our automatically emulated plans and manually emulated ones. Both the statistical test and the graphical comparison showed similar organ dose reconstruction accuracy. Although we observed a slightly worse performance for automatically emulated plans for the $|DE_{\text{mean}}|$ and $|DE_{2\text{cc}}|$ for the liver, and for the $|DE_{2\text{cc}}|$ for the right kidney for left-sided plans, we comment that an average of 0.2 Gy worse $|DE_{\text{mean}}|$ is not a source of uncertainty to be worried about when put into perspective to other uncertainties in dose reconstruction (such as patient anatomy difference⁷ and organ motion). For near field OARs, $|DE_{2\text{cc}}|$ is sensitive to the organ position with respect to the field border.⁴² Since in this study we did not focus on selecting a good surrogate anatomy, we based the dose reconstructions on a relatively large set of different patient anatomies. The results indicate large variations in $|DE_{2\text{cc}}|$ of the spleen and the kidneys (on average 1.2 to 3.8 Gy with ~ 2 Gy standard deviation, see Table 4). We note that our results indicate that the automatic reconstructions do, however, not lead to larger errors than the errors associated with the manual reconstructions for the right-sided plans. The underlying reasons might come from the assumptions taken in our approach, discussed as follows.

Based on the qualitative and quantitative results, there is still room to improve our automatic pipeline. In the following, we look into potential weak points that lead to unwanted results during the automatic process could be. First, landmark detection can introduce errors because it is based on image contrast between the bone and tissue, and assumed prior knowledge of the constellation of the patient’s bony anatomy. For cases with image acquisition artifacts, as well as abnormalities in patient anatomy (e.g., six lumbar vertebrae instead of five), the landmark detection algorithm cannot provide a reasonable outcome. In practice, such “outlier” cases should be known beforehand and handled manually. When large training data are available, deep-learning techniques in detecting spine structures in coronal radiographs could also be applied to improve the robustness of this step.^{43,44}

A second source of uncertainty potentially comes from the following assumptions regarding a particular RT plan geometry with respect to anatomical landmarks that were made: (1) the field sizes along RL and FH on different anatomies are proportional to the right or left half of the rib cage width (depending on the side of the irradiation field) and to the vertebral column length, respectively, (2) position of the isocenter with respect to the landmarks is consistent among

different patients, and (3) the vertebral column is sufficiently straight (i.e., not curved), such that the collimator angle can be estimated by fitting a straight line along the middle of the vertebrae. Although based on our validations, these assumptions were proven to achieve good quality emulations, we were able to identify some outliers in our own data set. For example, the patient's position could be (slightly) rotated around the FH direction, possibly leading to a discrepancy between the actual dimensions of the patient's anatomy and the dimensions of the visualized anatomy. The assumption (1) uses two scaling factors along each side, which is a simplified, however, valid and pragmatic approach considering that the possible rotation angles are small under the currently used clinical setup protocol. Nevertheless, for future work it might be interesting to investigate a way to estimate the rotation angle and correct the DRR accordingly (e.g., by affine transformation).⁴⁵

Like discussed before, apart from a specific but broad age range (2 to 6 years) and grouping based on intactness of the kidney, we did not perform a selection strategy that aims to find a somewhat “representative” surrogate patient. Indeed, how to best select representative surrogate anatomies is still an open problem and active field of research,^{17,24,25} which could greatly benefit from analyzing correlations between patient/plan features and dose outcomes, considering both accurate and inaccurate dose reconstruction cases. In our comparisons, both manual and automatic plan emulations use the same, equally well- or poorly representing surrogate CTs, we could essentially assess dose deviation differences in both fortunate and unfortunate scenarios. Because potentially bad representative surrogates are considered, Fig. 6 shows quite a large range of deviations in organ dose, especially for the spleen in left-sided plans and the liver in right-sided plans. If a good selection strategy is designed within a dose reconstruction approach, then a poor representative CT will not be selected as surrogate, and we expect automatic plan emulation to work well, so that large dose deviation values will not be observed for the final dose reconstruction approach.

Although the automatic pipeline is ultimately designed for dose reconstruction for historical patients using 2-D radiographs, we have considered CT scans and DRRs created from these

Fig. 6 Boxplots with points indicating for each dose reconstruction the deviations in organ mean dose based on the manual and automatic plan emulation approaches as compared to the reference organ mean dose. The lower and upper hinges correspond to the 25th to the 75th percentiles, the upper/lower whiskers extend from the upper/lower hinge to the largest/smallest value no further than 1.5 times the interquartile range from the hinge, the thick horizontal line inside the box indicates the 50th percentile (i.e., the median), and the points indicate the values associated with each case. We present separately the reconstruction cases for the two right-sided plans ($N = 40$) and the three left-sided plans ($N = 60$) in (a) and (b), respectively.

CTs, to be able to validate our results. We adapted the contrast for reference DRRs to simulate the historical radiographs as much as possible based on visual inspection, yet differences could still exist between truly historical reference radiographs and the reference DRRs used in our study.

Another limitation of this study concerning the dose reconstructions is that we simulated a photon beam with energy 6 MV as the radiation source to calculate dose distribution, however, for some historical patients, different radiation sources such as cobalt-60 that has different penetration properties were used in RT.^{7,11} Furthermore, we do not consider the use of wedges in our reference plans, which is quite common in historical plans. To make our pipeline work on real historical radiographs (and plans), more steps will be needed. For example, how to digitize and preprocess physical historical radiographs to remove irrelevant information (e.g., writing and annotations of clinicians) without sacrificing image quality needs to be investigated. Further, the simulation of different historical radiation sources and wedges is possibly needed for some historical cases. We do not consider simulating these historical settings to be an issue since the simulation of different historical radiation sources and the use of wedges have been widely implemented in TPSs.

The patient group we focus on is childhood cancer patients, as they are more prone to develop late adverse effects compared to adults.^{46,47} They have more anatomical variations across their age range,⁴⁸ and dose reconstruction approaches for this specific group are not well-studied.¹² We only validated our approach focusing on Wilms' tumor plans. However, we believe our validation results are valid for a larger class of plans, i.e., for other abdominal RT plans with an AP-PA field set-up, such as, e.g., neuroblastoma as similar anatomical landmarks are used for treatment planning. Although in this study, the development of our landmark detection algorithm was focused on use for coronal DRRs of the pediatric abdominal region, many parts of the pipeline can be used to determine relationships within (bony) anatomy and could therefore be extended to other cohorts and regions of interest when needed. It is possible to design landmark detection algorithms for radiographs resulting from different gantry angle projections on different regions, and for adult patients instead of children. The general idea of transforming the RT plan field based on detected landmarks can also easily be extended.

6 Conclusion

We successfully introduced an automatic pipeline to perform dose reconstruction for historical patients that were treated with AP-PA RT fields planned on 2-D radiographs. Our validation results show that our pipeline achieves performance similar to a laborious, time-consuming manual approach, across both qualitative assessments and quantitative ones. Our pipeline enables, for the first time, performing large-scale dose reconstructions, for abdominal RT plans with AP-PA field set-up. In turn, this enables the use of more powerful data analysis and machine learning techniques, to design better dose reconstruction strategies, to validate a dose reconstruction approach, and to study in more detail how cancer treatment including RT causes late adverse effects, which is a key to improve today's pediatric RT.

7 Appendix

An overview of the parameters used in each step of peak detection as described in Sec. 2.3 can be found in Table 5.

Disclosures

Dr. Alderliesten, Dr. Bel, and Prof. Dr. Bosman are involved in projects supported by Elekta. KiKa and Elekta had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Table 5 Overview of parameters used in each step of peak detection.

Step	Input signal	Minimal peak vertical drop	Minimal distance between two peaks (mm)	Peak number	Peak width (mm)
1. Initial estimation of the vertebral column borders	$\text{smooth}[V'_{\text{middle_top_half}}(x)]$	$0.3 * \max(\text{Input})$	12	2	2
2. ^a Initial estimation of the intervertebral disc locations	$-\text{smooth}[H_{\text{init_vc}}(z)]$	0	z-dimension of image /12	9	0
3. ^b Right/left border estimation for each vertebral bone	$\text{smooth}[V'_{vb,j}(x)]$ with $H_{vb,j}(x) = \begin{cases} \sum_z I(x, z), x \in \{x_{\text{init_vc_R}} - 20, x_{\text{init_vc_L}} + 20\}, \\ z \in \{z_{\text{init_j}}, z_{\text{init_}(j+1)}\}, j \in \{1, \dots, 5\} \\ \sum_z I(x, z), x \in \{x_{\text{init_vc_R}} - 25, x_{\text{init_vc_L}} + 25\}, \\ z \in \{z_{\text{init_j}}, z_{\text{init_}(j+1)}\}, j \in \{6, 7\} \end{cases}$	$0.15 * \max(\text{Input})$	12	3	5
4. ^c Second estimation of intervertebral disc locations	$-\text{smooth}[H_{vc}(z)]$ as: $H_{vx}(z) = \sum_x I(x, z)$, for $z \in \{z_{\text{init_j}}, z_{\text{init_}(j+1)}\}, x \in \{x_{vc_R-j} - 4, x_{vc_L-j} + 4\}, j \in \{1, \dots, 7\}$.	0	z-dimension of image /12	9	0

^aPeak locations within 10 mm from the upper border of the image are removed, to filter out unwanted peak locations above T10.^b(a) $j = 8$ (L5) is not considered because near L5, the presence of iliac bones introduces signals that complicate the detection of the vertebrae. (b) When more than two peaks are detected (because of contrasts or noise caused by other structures near the vertebrae), these peaks are divided to "right-sided" and "left-sided" peaks based on their x-location w.r.t. the middle of the vertebrae. For each side, the peak with the largest magnitude is selected as the bone boundary of that side.^cPeak locations within 10 mm from the upper border of the image are removed, to filter out unwanted peak locations above T10.

Acknowledgments

Financial support of this work was provided by Stichting Kinderen Kankervrij (KiKa; project no. 187). The authors further thank Pieter Gangel for his help with plan quality assessment and Petra S. Kroon, PhD and Geert O.R. Janssens, MD, PhD (department of Radiation Oncology, UMC Utrecht Cancer Center, Utrecht, The Netherlands) for sharing the data of 37 patients treated at the UMC Utrecht/Princess Máxima Center for Pediatric Oncology for inclusion in this study. We further thank Dr. Irma W.E.M. van Dijk, PhD for proofreading parts of the manuscript.

References

1. G. T. Armstrong et al., "Aging and risk of severe, disabling, life-threatening, and fatal events in the childhood cancer survivor study," *J. Clin. Oncol.* **32**(12), 1218–1227 (2014).
2. A. B. de Gonzalez et al., "Second solid cancers after radiation therapy: a systematic review of the epidemiologic studies of the radiation dose-response relationship," *Int. J. Radiat. Oncol. Biol. Phys.* **86**(2), 224–233 (2013).
3. K. C. Oeffinger et al., "Chronic health conditions in adult survivors of childhood cancer," *N. Engl. J. Med.* **355**(15), 1572–1582 (2006).
4. W. D. Newhauser et al., "A review of radiotherapy-induced late effects research after advanced technology treatments," *Front. Oncol.* **6**, 13 (2016).
5. C. H. Stokkevang et al., "Estimated risk of radiation-induced cancer following paediatric cranio-spinal irradiation with electron, photon and proton therapy," *Acta Oncol.* **53**(8), 1048–1057 (2014).
6. I. W. van Dijk et al., "Dose-effect relationships for adverse events after cranial radiation therapy in long-term childhood cancer survivors," *Int. J. Radiat. Oncol. Biol. Phys.* **85**(3), 768–775 (2013).
7. J. V. Bezin et al., "A review of uncertainties in radiotherapy dose reconstruction and their impacts on dose-response relationships," *J. Radiol. Prot.* **37**(1), R1 (2017).
8. X. G. Xu, "An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history," *Phys. Med. Biol.* **59**(18), R233 (2014).
9. E. S. Chernak et al., "The use of computed tomography for radiation therapy treatment planning," *Radiology* **117**(3), 613–614 (1975).
10. J. Thariat et al., "Past, present, and future of radiotherapy for the benefit of patients," *Nat. Rev. Clin. Oncol.* **10**(1), 52–60 (2013).
11. D. I. Thwaites and J. B. Tuohy, "Back to the future: the history and development of the clinical linear accelerator," *Phys. Med. Biol.* **51**(13), R343 (2006).
12. M. Stovall et al., "Dose reconstruction for therapeutic and diagnostic radiation exposures: use in epidemiological studies," *Radiat. Res.* **166**(1), 141–157 (2006).
13. A. Ng et al., "Individualized 3D reconstruction of normal tissue dose for patients with long-term follow-up: a step toward understanding dose risk for late toxicity," *Int. J. Radiat. Oncol. Biol. Phys.* **84**(4), e557–e563 (2012).
14. C. Veres et al., "Retrospective reconstructions of active bone marrow dose-volume histograms," *Int. J. Radiat. Oncol. Biol. Phys.* **90**(5), 1216–1224 (2014).
15. C. Lee et al., "Reconstruction of organ dose for external radiotherapy patients in retrospective epidemiologic studies," *Phys. Med. Biol.* **60**(6), 2309–2324 (2015).
16. J. A. Kalapurakal et al., "Feasibility and accuracy of UF/NCI phantoms and Monte Carlo retrospective dosimetry in children treated on National Wilms Tumor Study protocols," *Pediatr. Blood Cancer* **65**(12), e27395 (2018).
17. Z. Wang et al., "How do patient characteristics and anatomical features correlate to accuracy of organ dose reconstruction for Wilms' tumor radiation treatment plans when using a surrogate patient's CT scan?" *J. Radiol. Prot.* **39**(2), 598–619 (2019).
18. C. W. Taylor et al., "Cardiac exposures in breast cancer radiotherapy: 1950s–1990s," *Int. J. Radiat. Oncol. Biol. Phys.* **69**(5), 1484–1495 (2007).
19. M. Stabin et al., "Realistic reference adult and paediatric phantom series for internal and external dosimetry," *Radiat. Prot. Dosim.* **149**(1), 56–59 (2012).

20. A. M. Geyer et al., “The UF/NCI family of hybrid computational phantoms representing the current US population of male and female children, adolescents, and adults—application to CT dosimetry,” *Phys. Med. Biol.* **59**(18), 5225–5242 (2014).
21. D. Tward et al., “Generating patient-specific dosimetry phantoms with whole-body diffeomorphic image registration,” in *IEEE 37th Annu. Northeast Bioeng. Conf.*, pp. 1–2 (2011).
22. W. Segars et al., “The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization,” *Med. Phys.* **42**(8), 4719–4726 (2015).
23. V. Cassola et al., “Standing adult human phantoms based on 10th, 50th and 90th mass and height percentiles of male and female Caucasian populations,” *Phys. Med. Biol.* **56**(13), 3749–3772 (2011).
24. M. Virgolin et al., “On the feasibility of automatically selecting similar patients in highly individualized radiotherapy dose reconstruction for historic data of pediatric cancer survivors,” *Med. Phys.* **45**(4), 1504–1517 (2018).
25. E. J. Stepusin et al., “Assessment of different patient-to-phantom matching criteria applied in Monte-Carlo based computed tomography dosimetry,” *Med. Phys.* **44**(10), 5498–5508 (2017).
26. Z. Wang et al., “Are age and gender suitable matching criteria in organ dose reconstruction using surrogate childhood cancer patients’ CT scans?” *Med. Phys.* **45**(6), 2628–2638 (2018).
27. P. D. Inskip et al., “Radiation-related new primary solid cancers in the Childhood Cancer Survivor Study: comparative radiation dose response and modification of treatment effects,” *Int. J. Radiat. Oncol. Biol. Phys.* **94**(4), 800–807 (2016).
28. W. Schillemans et al., “SU-E-T-208: automated routine 3D secondary patient dose calculation prior to and during fractionated treatment,” *Med. Phys.* **40**(6 Part 13), 252–252 (2013).
29. Z. Wang et al., “Automatic radiotherapy plan emulation for 3D dose reconstruction to enable big data analysis for historically treated patients,” *Proc. SPIE* **10954**, 109540V (2019).
30. P. Mildenerberger, M. Eichelberg, and E. Martin, “Introduction to the DICOM standard,” *Eur. Radiol.* **12**(4), 920–927 (2002).
31. R. L. Siddon, “Solution to treatment planning problems using coordinate transformations,” *Med. Phys.* **8**(6), 766–774 (1981).
32. F. M. Khan and J. P. Gibbons, *Khan’s the Physics of Radiation Therapy*, Lippincott Williams & Wilkins/Wolters Kluwer, Philadelphia, Pennsylvania (2014).
33. B. Jereb et al., “Radiotherapy in the SIOP (International Society of Pediatric Oncology) nephroblastoma studies: a review,” *Med. Pediatr. Oncol.* **22**(4), 221–227 (1994).
34. G. M. Vujanić et al., “Revised International Society of Paediatric Oncology (SIOP) working classification of renal tumors of childhood,” *Med. Pediatr. Oncol.* **38**(2), 79–82 (2002).
35. R. W. Schafer and A. V. Oppenheim, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey (1989).
36. M. I. Sezan, “A peak detection algorithm and its application to histogram-based image data reduction,” *Comput. Vision Graphics Image Process.* **49**(1), 36–51 (1990).
37. D. Mason, “SU-E-T-33: pydicom: an open source DICOM library,” *Med. Phys.* **38**(6 Part 10), 3493–3493 (2011).
38. R. Furtwängler et al., “Clear cell sarcomas of the kidney registered on International Society of Pediatric Oncology (SIOP) 93-01 and SIOP 2001 protocols: a report of the SIOP Renal Tumour Study Group,” *Eur. J. Cancer* **49**(16), 3497–3506 (2013).
39. G. J. D’angio et al., “Radiation therapy of Wilms’ tumor: results according to dose, field, post-operative timing and histology,” *Int. J. Radiat. Oncol. Biol. Phys.* **4**(9), 769–780 (1978).
40. J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *J. Mach. Learn. Res.* **7**, 1–30 (2006).
41. S. C. Huijskens et al., “Magnitude and variability of respiratory-induced diaphragm motion in children during image-guided radiotherapy,” *Radiother. Oncol.* **123**(2), 263–269 (2017).
42. S. Lamart et al., “Radiation dose to the esophagus from breast cancer radiation therapy, 1943–1996: an international population-based study of 414 patients,” *Int. J. Radiat. Oncol. Biol. Phys.* **86**(4), 694–701 (2013).

43. B. Aubert et al., “Automatic spine and pelvis detection in frontal x-rays using deep neural networks for patch displacement learning,” in *IEEE 13th Int. Symp. Biomed. Imaging*, pp. 1426–1429 (2016).
44. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature* **521**(7553), 436–444 (2015).
45. T. Vrtovec, F. Pernuš, and B. Likar, “A review of methods for quantitative evaluation of axial vertebral rotation,” *Eur. Spine J.* **18**(8), 1079–1090 (2009).
46. Y. T. Cheung et al., “Chronic health conditions and neurocognitive function in aging survivors of childhood cancer: a report from the Childhood Cancer Survivor Study,” *JNCI* **110**(4), 411–419 (2018).
47. M. M. Hudson et al., “Health status of adult long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study,” *JAMA* **290**(12), 1583–1592 (2003).
48. V. Varchena, “Pediatric phantoms,” *Pediatr. Radiol.* **32**(4), 280–284 (2002).

Ziyuan Wang is a PhD student working in the Radiation Oncology Department, Amsterdam UMC, University of Amsterdam. In 2015, she received her MSc degree in applied physics from the Delft University of Technology in The Netherlands. Her research interests include medical imaging analysis and physics of radiation treatment.

Marco Virgolin is a PhD candidate at Centrum Wiskunde and Informatica in Amsterdam, The Netherlands, and is enrolled at the Delft University of Technology, Delft, The Netherlands. He received his MSc degree in computer engineering from the University of Trieste, Italy. He is mostly interested in evolutionary and explainable machine learning, with a special focus on genetic programming and symbolic regression.

Peter A. N. Bosman is a senior researcher in the Life Sciences and Health (LSH) research group at the Dutch National Research Institute for Mathematics and Computer Science (Centrum Wiskunde and Informatica) and professor of Evolutionary Algorithms (EAs) at Delft University of Technology. His research concerns the design of scalable model-based EAs and their application, primarily in the LSH domain. He has (co-)authored over 100 refereed publications, out of which four received best paper awards.

Koen F. Crama is a research radiation therapy technologist in the Radiation Oncology Department, Amsterdam UMC. He received his MSc degree in radiation oncology from Haarlem University in 2015. He is involved in research and clinical implementation of modern treatment planning techniques as image-guided adaptive planning, robust optimization, automatic planning, and also in proton therapy.

Brian V. Balgobind is a radiation oncologist at the Department of Radiation Oncology at Amsterdam UMC-location AMC. He received his MD from the University of Utrecht in 2005 and his PhD from the Erasmus University, Rotterdam, The Netherlands, in 2011. His current focus is on stereotactic ablative therapy, (pediatric) sarcoma, and hematologic malignancies.

Arjan Bel received his MSc degree in computational/experimental physics from Utrecht University in 1990 and his PhD from Netherlands Cancer Institute, The Netherlands, in 1996, on strategies to improve patient setup accuracy during radiation therapy. Currently, he is the head of the clinical physics group at the Department of Radiation Oncology, Amsterdam UMC, The Netherlands. He is involved in research and clinical implementation of image-guided adaptive radiation therapy.

Tanja Alderliesten received her PhD in computer science from the Utrecht University, in The Netherlands, in 2004. Currently, she is an associate professor at the Department of Radiation Oncology of the Leiden University Medical Center, The Netherlands. Her research focus is translational in nature and primarily concerns the development of state-of-the-art methods and techniques from the fields of mathematics and computer science (including image processing, biomechanical modeling, and optimization) for radiation oncology.