Investigation of Deep Learning Dose Estimation Robustness for MR Guided Radiotherapy

Master - Thesis

Eingereicht von: Simon Gutwein

Studiengang: Medizinische Strahlenwissenschaften

Matrikelnummer: 5419474

Gutachter 1:   Prof. Dr. Daniela Thorwarth

Gutachter 2:   Dr. Christian Baumgartner

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Dedication

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Declaration

I hereby declare that I have written the submitted Master's thesis without outside help. I have marked as such all statements taken verbatim or in spirit from other works, if applicable also from electronic media. I am aware that my Master's thesis in digital form will be checked for plagiarism if unmarked texts are used in whole or in part. A violation of the basic rules of scientific work is considered as an attempt to deceive or cheat and will result in the corresponding consequences according to the valid examination regulations of the Medical Radiation Sciences degree programme. For the purpose of comparing my Master's thesis with existing sources, I agree that my Master's thesis will be entered into a database where it will remain after my graduation. I do not cede any further rights to reprint and use it. I further declare that the thesis has not been the subject of any other examination procedure, neither in its entirety nor in essential parts, and that it has not been published.

_____   _____
        Signature   Date

Contents

List of Figures

List of Tables

List of Abbreviations

1   Introduction

Cancer is one of the most contributing diseases leading to death under the age of 70 years in most countries [1]. A multitude of modalities exists for cancer treatment. The application of radiotherapy (RT) is widespread and accessible throughout the world [2] and is used for a multitude of tumor entities such as prostate [3–5], breast [6–8], head and neck (HN) [9–12], liver [13–17] and lymphnode (LN) cancer [18–22]. RT is most applied on solid, local and delimited tumors after surgical resection. The target volume, which is defined by an oncologist, is irradiated with ionizing radation during several radio treatment sessions. Radiation is generated using a medical linear accelerator. Most applications use photons as their initial ionizing radiation. Due to the decreased repair capabilities of cancerous cells, compared to healty tissue, when exposed to ionizing radiation the RT is split into fractions, minimizing the bad side effect while destroying cancerous cells in the target volume. In combination with different angles from where the photons are irradiated an optimal radiation dose inside the target volume can be achieved while sparing organs at risk (OAR). There exist multiple application modalities of RT that depend on the tumor site, institutional guidelines and personal preference. With the intensity modulated radio therapy (IMRT) treatment modality, which is the most applied treatment modality, radiation is applied from multiple specific angles around the patient [23]. Prior to a patient's radiation treatment, a treatment plan is created by an oncologist in conjunction with a treatment planner or medical physicist. Due to the varying anatomy of patients, individualization of treatment plans is crucial to achieve an optimal therapeutic outcome. The goal of RT planning is to optimally tailor treatment to the patient's individual tumor location, shape and environment. To adapt for different shapes of tumors shape and weight of each beam can be changed using lead aperatures inside the accelerator head.

Individualization of treatment plans is not only needed to optimally irradiate the target volume but to spare surrounding OAR. The general workflow for treatment plan creation consists of a computed tomography (CT) aquisition prior to treatment followed by tumor volume and OAR delineation and an iterative process of treatment planning involving the onoclogist and the treatment planner. Iterative planning can take up to several days to create an convenient treatment plan.

The newest development regarding RT is the incorparation of an magnetic resonance imaging (MRI) device into a linear accelerator, called MR-Linac. This enables live imaging before or during the treatment of the patient. Offering a lot of opportunities such as quantitative MRI or treatment supervision and online adaption of treatment plans. However, many challenges still need to be overcome to fully exploit the potential of the

MR-Linac. Current research effort is focusing on the development of the needed software tools to enable an MR-only workflow for RT planning, which makes the CT reduntant in the treatment planning pipeline. Regarding an MR-only treatment plan workflow recent work involves the creation of pseudo CT images [24–26], delineation of OAR [27–29], plan optimization [30, 31] and dose calculation algorithms [32]. Ultimately this should lead to a supervised automated real-time treatment plan adaption, which enables the treatment plan to be automatically adapted, when the patients anatomy changes between or during fractions. As previously mentioned the entire process of treatment planning is a time consuming and also tedious task to do. Therefore, current research is aiming towards the improvement of time consumption and ultimately automating most tasks to be applicable in real-time. An important part in the pipeline consists of the dose deposition (DD) calculation, which is needed for the treatment planning software, as well as dose verification processes. Currently DD calculation algorithms are based on Monte-Carlo (MC) simulations [33]. To achieve accurate results, a multitude of particle histories, in the order of $10^7 - 10^{11}$ particles, are simulated, taking into account the patients anatomy and accelerator settings. In order to obtain accurate results, prior knowledge of the interaction processes is required, which are very well described in the literature. DD calculation algorithms with MC yield very accurate results. Disadvantage of these algorithms is that due to the large number of particles required for accurate results, simulation times can become increasingly large with the required accuracy requirements. Even in times of parallelization and GPU acceleration techniques time consumption still poses a problem. To enable real time MR-only treatment planning, the need for a fast and accurate dose estimation tool arises.

The use of deep learning (DL) has shown to be promising in a multitude of applications in computer vision (CV) tasks in medicine [34–38]. Since DD is based on CT images aquired from the patients anatomy and the accelerator settings, it seems promising to apply DL to overcome the challenges of time consuming simulations with MC.

The aim of this project was to improve the state of the art dose estimation tools involving DL and evaluate robustness aswell as the generalization capabilities. To investigate the applicability of DL for DD calculation we trained a 3D-UNet with training data combining linear accelerator settings as well as patient anatomies to yield 3D dose distributions. We then evalutated the accuracy of said predictions by applying the network on a multitude of tumor sites.

2 Medical Knowledge

2.1 Tumor Radiation

As mentioned in the previous section external beam RT is a widespread non-invasive modality for cancer treatment with or without surgical tumor resection. Studies indicate that it should be used in over 50% of all cancer cases once in the course of the illness [39]. Ionizing radiation of RT interacts with the patient's tissue in a way that causes direct or indirect biological damage to cells. Interaction processes release free radicals inside the tissue, which cause deoxyribonucleic acid (DNA) damage in the form of single- or double strand breaks, leading to the induction of apoptosis [40] and therefore the destruction of the radiated cell. Healthy tissue is able to repair such DNA damage under normal conditions. The repair capabilities of cancerous cells are downregulated due to the microenviromental conditions, such as hypoxia, low pH and nutrient deficiency, in the primary cancer region [41]. Radiation takes advantage of these down-regulated repair mechanisms, resulting in the destruction of cancer cells while limiting radiation to a point where healthy tissue can regenerate.

Photon attenuation is described by an exponential decay, meaning that a photon beam can not be entirely shielded. This is of crucial importance for the optimal planning of entire RT. Since tumors are usually located in deeper regions of the body (see the example of prostate cancer in fig. 1), it is inevitable to irradiate healthy tissue that is located in front of or behind the tumor volume with respect to the radiation direction. Aim of a treatment plan optimization process is to achieve an optimal dose coverage over the entire target volume, while minimizing the radation dose in surrounding OAR. Specific limits for dose coverage of tumor volumes as well as OAR are set by institutional standard operating procedures, which account for recent scientific findings in literature. The first step during treatment planning is the definition of an optimal dose distribution over the entire irradiated volume. Here, the planning software tries to find a dose distribution that satisfies all the given margins and limits, without considering the actual achievable dose distribution with respect to machine parameters and settings. To achieve this first optimal dose distribution the planning software takes the tumor volume and all delineated OAR with their defined dose limits in consideration. In the next step the treatment plan is created in an iterative manner trying to achieve the given optimal dose distribution using actual possible accelerator settings. The result from the planning process is a set of segments defined by multi leaf collimator (MLC) configurations, describing the shape of the beam, accelerator gantry angles and a defined amount of particles irradiated per
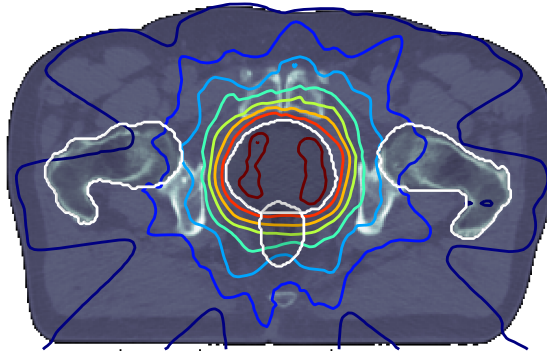
Figure 1: CT image of a prostate cancer patient including the dose distribution of the respective treatment plan shown as iso dose lines. Areas of high dose deposition are displayed in red and orange. Contours of OAR as well as PTV are displayed in white.

segment. This created plan offers an optimal dose coverage of the target volume and sparing of OAR according to the institutional standard operating procedure.

Further decrease of undesired side effects in irradiated healthy tissues is achieved by treatment fractionation. By dividing the treatment into several sessions with lower radiation doses, the increased repair capabilities of healthy tissue are utilized. Time between and dose for fractions is chosen according to findings in current literature yielding information about repair capabilities of cancerous and healthy cells [42, 43]. This leads to a steady decrease in cancer cell population in the irradiated region.

2.2 MR Linear Accelerator

In the current state of the MR-Linac employment the use of a conventional CT as the base for treatment planning is inevitable. On the base of this CT target volumes and OAR are defined by an oncologist. In a conventional setting, this delineation and the following treatment plan on the base of this first CT is used for all treatment fractions. This results in uncertainties during treatment because it is assumed that the position of organs remains constant during the entire course of treatment. Especially for moving OAR such as bladder or sigmoid in the case of a lower abdomen cancer patient this poses a problem because it cannot be assured that they did not move during treatment or between fractions without a control CT. The MR-Linac offers the acquisition of an MR image before treatment. With such position verification using an MR image the attending oncologist can decide wether previous delineations are still valid or not. If the delineations are not valid for the current fraction the MR-Linac offers a solution to adapt the treatment plan to interfraction movement of target volume or OAR. OAR are registered from the

CT onto the MR image and the treatment plan can be adapted to change in position or shape. Ongoing studies investigate if this treatment modality may allow to reduce safety margins accounting for uncertainties regarding patient positioning as well as tissue movement.

Current research is going towards the further exploration of opportunities involving the MR-Linac. The vision behind the MR-Linac would be to adapt the treatment plan based on real-time image acquisition during the treatment. Critical steps to enable this proposed workflow are the optimization of all involved processes of the treatment planning pipeline to be applicable in real-time and to be based on MR images only. An MR-only workflow is promised to ultimately resulting in in-time adaption to intra-fraction movements such as unwanted patient movement as well as breathing motion potentially leading to a significant decrease in safety margins. All necessary steps to be automated or optimized would be the delineation of tumor volumes and OAR on MR images, treatment planning and optimization as well as pseudo-CT creation from MR images to enable DD calculations. Pseudo-CT creation is needed for DD because MRI is not a quantitative imaging modality, meaning that the pixel values of MR images are not correlated to the electron density of the tissue, which are needed for dose calculation algorithms. Therefore a translation from qualitative MRI data to quantitative electron density information is needed to estimate deposited dose inside a given volume or patient's anatomy.

DD is the part of the MR-only treatment pipeline we are focusing on in this study, trying to optimize its computational resources as well as computation time. As previoulsy mentioned MC simulations are used for the simulation of particle histories to give an estimate of deposited dose inside the patient's anatomy. These MC simulations are used in the secondary dose calculation for treatment plan optimization as well as the dose verification.

2.3 Physical Fundamentals

Photons irradiated from the linear accelerator interact with the patient's tissue. All interaction processes are of stochastic nature, meaning it can not be exactly determined when a particle will interact in which manner during is path through a dense volume. Interaction processes of photons as well as electrons are quantitatively well described in literature giving probability distributions for different energies as well as densities of tissues. Main interaction processes of photons in the energy sprectrum of the MR-Linac are elastic and inelastic interactions such as rayleigh scattering, the photo effect or compton scattering, which leads to small changes in trajectory of photons inside a medium. With increased depth inside tissue this change in trajecotry leads to a widening

of the initial beam. During the latter two interaction processes photons transmit all of their or partial energy onto secondary electrons. Shorter mean free path lengths and higher ionization capabilities of electrons lead to dose deposition in the nearby surrounding tissue, which is the main contribution to DD.

MRI requires an strong magnetic field. This magnetic field is present at all times, due to the need of superconduction. Trajectories of moving charged particles such as electrons are affected by the Lorentz-Force first described by Lorentz in 1937 [44]. Changes in trajectory induce a shift in the DD towards the direction of the Lorentz-Force. Electrons have increased mean free path lengths inside low density mediums such as air compared to higher density mediums such as tissue or bones. At the intersection of tissue to air cavities inside the patient's anatomy the Lorentz-Force influences the trajectory of electrons which causes them to return to the surface of the tissue resulting in an increase in deposited dose. This occurence is the so called electron return effect (ERE) [45] and has significant impact on tumor sides where air cavities are present.

2.4 Dose Deposition Calculation

There exist multiple algorithms besides MC for DD that are used in reasearch and a clinical setting such as collapsed cone [46] or pencil beam algorithms [47] which is currently used as a primary dose estimation algorithm in the institutional planning software. In this study we focused on the MC simulation algorithm which will build the gold standard of our dose distributions for or training data.

As previously mentioned, IMRT plans are composed of individual segments for certain irradiation angles determined by the planning algorithm. More challenging positions of target volumes and OAR result in higher number of segments, because sparing of organs at risk becomes harder while maintaining dose coverage in the target volum, with less and bigger fields. Each segment is assigned an individual MLC configuration, adapted to the shape of OAR and the target volume and a defined number of irradiated particles. The MLC of a MR-Linac is composed of 80 leaf pairs of width 7.15 mm in the isocenter resulting in a maximum fieldsize of 57.2 $cm \times 22$ cm. The Number of irradaited particles is measured in monitor units (MU). 100 MU are defined as 1 Gy in depth of 10 cm inside a water phantom with a fieldsize of 10×10 cm with a source surface distance of 100 cm. To calculate the DD for our training data, we utilized the open source software MC solution EGSnrc [48] provided under (`https://github.com/nrc-cnrc/EGSnrc`). With the use of EGSnrc we were able to calculate the deposited dose for all segments indivudually. Previous work from our institution by Friedel et al. [49] where the use of EGSnrc for the MR-Linac was investigated enabled us to accurately predict the deposited dose for any

beam configuration. The output yielded by EGSnrc is normalized to 100 MU for the given input parameters such as irradiation volume and accelerator settings.

MC simulations such as EGSnrc are used for dose calculation in both clinical and research applications. Photons transfer their energy in the irradiated volume to electrons or positrons in the energy spectrum of interest for external beam radiotherapy. In the absence of an analytical solution, without significant simplifications and assumptions for the conditions, the MC algorithm is used to simulate a variety of particle trajectories in the desired target volume. Particle trajectories describe the precise path of a photon from the source to the point in the volume where it has lost all of its energy, including energy tranfer to secondary electrons in the course of collisions. The stochastic nature of the interaction processes of photons and electrons requires a large number of simulated particles to obtain an accurate result. Long simulation times are the consequence with a number of histories on the order of 10^7 to 10^{11}.

3 Related Work

MR guided radiotherapy has received increased attention in the last years, with a multitude of studies and reasearch areas involved. The problem of fast DD estimation is becoming increasingly important with the vision of online and even real-time RT plan optimization and adaptation. DD algorithms are involved in primary and secondary dose engines as well as dose verification processes. The aim of this work was to replace MC simulations with DL for secondary dose engine or dose verficiation, because for both the specific accelerator parameters are given.

Different approaches towards the solution of the DD involving deep learning were pursued in previous work from different working groups. Neishabouri et al. [50] applied a long short term memory (LSTM) in the application of DD for proton irradiation. Input to the LSTM network was a sequence of 2D slices from the irradiated volume. Each sequence represented the irradiated volume of the proton beam. Due to the very limited range of protons inside matter and the small beam widening, volumes were small with $15 \times 15 \times 150$ *voxels* resulting in a field of view of $30 \times 30 \times 300$ mm^2 with a isotropic resolution of 2 mm per voxel. Each given slice consisted of size 15×15 *pixels*. Network output was the respective dose distribution for the given input sequence of slices. To deviate from patient anatomies they used the approach of creating areas of increased density inside a $15 \times 15 \times 150$ *voxels* sized volume. By variation in position and size of the area the network was able to map DD processes to different densities translating to good results when applied to patient anatomies. Results showed 98.57% mean gamma passrate (1%/3mm) for artificial cases and an average gama passrate of 97.85% for patient test cases.

A combination of a 3D-UNet and convolutional LSTM networks as skip connections were used by Martinot et al. [51]. By approaching the task as an active denoising problem, they used a set of noisy monte carlo simulations as an input to the 3D-convLSTM-UNet and the network yielded denoised dose distributions for the given volume. Gamma passrates were $94.1\% \pm 1.2\%$ with 3%/3 mm as the gamma criteria for five patient dose distributions.

Approaching the problem as an active denoising problem was also done by Neph et al. [52] that used three input volumes consisting of patient CT, photon fluence map and undersampled dose of a single beamlet. By individual analysis of each volume and later combination they achieved to predict the redisdual dose needed for denoising. Mean absolute error was reduced to 0.106% from initial 25.7% of the undersampled beamlet. A significant decrease in calulation time was also reached with 220 ms compared to the 380 s for the simulation of a fully sampled beamlet.

Inclusion of fluence maps was also pursued by Fan et al. [53]. By projecting a 2D fluence map into 3D space they created 3D fluence volumes from which a 3D UNet like network made predictions of the deposited dose. Dose variation normalized to the prescribed dose was 0.17% ± 2.28%.

Kontaxis et al. [54] used a 3D-UNet with five 3D Inputs to combine information from the accelerator settings as well as the patients anatomy into the input of the network. This enabled them to predict single segments, which can be added up to enitre treatment plans. Application was limited to lower abdomen cancer treatment plans with gamma passrates of 99.9 ± 0.3 for (3%/3mm) for prostate cancer plans. Dose estimation times were short with approximately 3 minutes for a treatment plan with 41 segments.

Work of Kontaxis et al. [54] showed promising results towards a fast DL based DD estimation tool for radio treatment plans. We therefore made adjustments to their proposed network to be able to handle the input data from the MR-Linac and evaluated its performance regarding accuracy, robustness and generalization capabilities.

4 Material and Methods

4.1 Network Architecture

Our prosoed network architecture (fig. 2) consists of a basic 3D-UNet as implemented for the first time by Ronneberger et al. [55]. It is composed of an encoding and a decoding part connected by skip-connections. The encoding part is used to analyse the given input and transform it into a more abstract presentation level. This is achieved by consecutive application of 3D convolutions to the input. Decoding translates this higher level of representation to the interpretable 3D space by applying 3D transposed convolutions. Skip connections are utilized to incorparate a lower level of presentation into decoding part and preserve details from the input masks.

We adapted the downsampling depth of the proposed network due to memory limitations when using 3D data to a factor of 8. Maximum layer depth was 512 in the bottleneck connection. Skip connections were used before each pooling layer. To include a multitude of information from the patient as well as the RT plan, we used five input volumes as proposed by [54] concatenated along the forth dimension, resulting in an input volume of size $5 \times width\ (W) \times height\ (H) \times depth\ (D)$. Resolution for all input volumes were set to the resolution of our CT scans with voxel dimensions of $1.1718 \times 1.1718 \times 3$ mm^3 in the coronal, sagital, and tanversal plane.
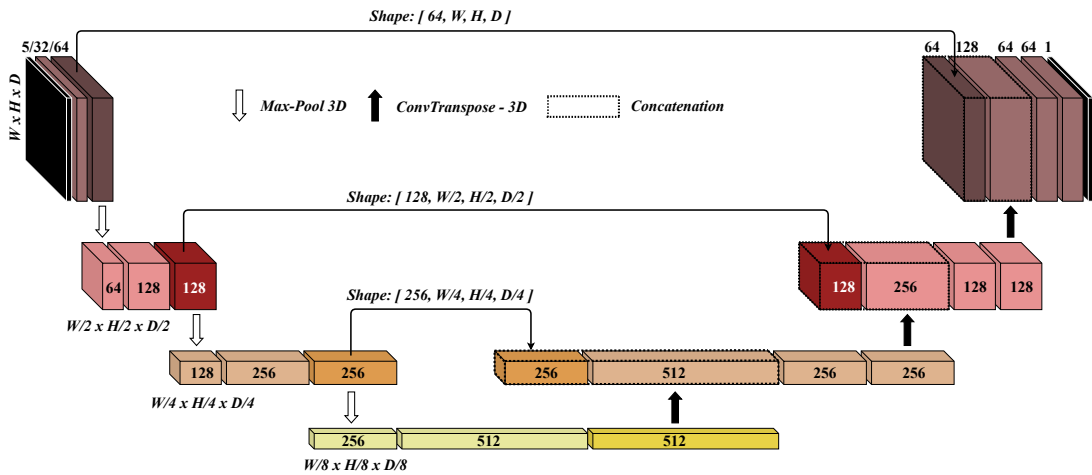
Figure 2: Basic scheme of adapted network architecture with a downsampling ratio of 8 and a maximum depth of 512 in the bottleneck layer. Input dimensions: $5 \times W \times H \times D$; Output dimensions: $1 \times W \times H \times D$

4.2 Network Input & Output

4.2.1 Network Input

The network expects five 3D volumes concatenated along their fourth dimension as resulting in an 4D input. These five 3D volumes (fig. 3) combine different spatial, anatomical as well as accelorator information into the training data. By doing so gantry as well as accelerator head information can be directly translated into 3D space, which makes them interpretable for the network.

(a) **Beam Shape**: The trajectory of the beamshape into 3D space is of crucial importance. Combining information from MLC configurations as well as gantry angle gives the network crucial information about the translation of the beams eye view into 3D space. Information about which voxel is intersected or partially intersected by the beam is stored in this mask. To additionally account for the output factor, the fieldsize of each segment is stored in each voxel intersected with the beam field. Voxel values of partially clipped voxels are scaled down by the percentage of the voxel that lies within the field. The output factor describes the physical process of increasing photon scatter into the central beamline for increasing fieldsizes.

(b) **Center Beam Line Distance**: Flattening filters are used to create uniform field distribution for photon beams from conventional accelerators. Flattening filters are not built into MR-Linac which makes it flattening filter free (FFF). Beamprofiles of the MR-Linac are therefore not uniform. The dose profile peaks in the central beam line from where it flattens out to the edges of the field. Therefore receive fields that are distant from the central beamline slightly lower doses that fields in the direct center. Each voxel in this mask is therefore assigned its minimum distance from the central beamline.

(c) **Source Distance**: The radiation pattern of photons from the accelerator head can be assumed to be the front of a spherical wave. The photon fluence is therefore decreasing with the square of the distance from the source. The source distance mask takes this physical relationship into account. Each voxel within the volume is assigned its distance to the source, taking into account the different voxel dimensions in the different spatial directions.
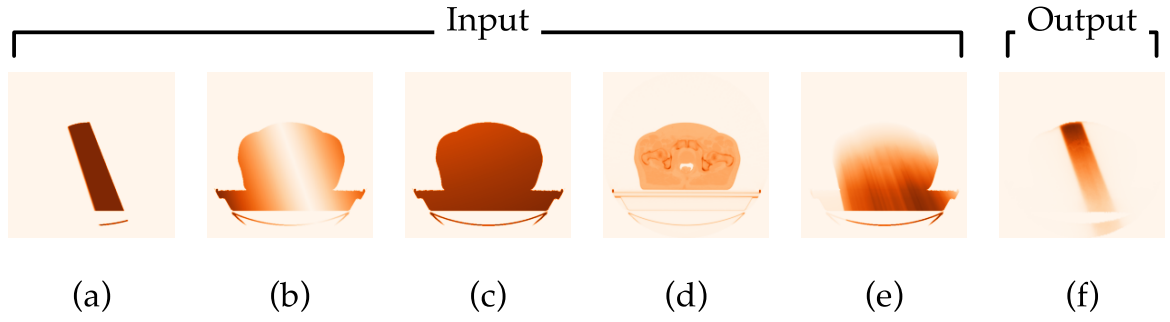
Figure 3: Examplary input and output for a single segment of a prostate plan. (a): beam shape, (b): center beam line distance, (c): source distance, (d): CT, (e): radiological depth, (f): dose distribution,

(d) CT: Dose deposition and interactions of particles depend on their energy as well as the electron density of the affected volume. As the initial energy of the radiated photons is assumed constant, the impact of the electron density of the volume (in this case the patients anatomy) is responsible for the different interaction processes of photons as well as secondary electrons. Hounsfield Units (HU) of the patient's anatomy acquired from a CT scan are stored in each voxel of the volume. Photons radiated from gantry angles below the treatment couch, interact partly inside it, therefore it was included in the CT mask.

(e) Radiological Depth: A photon beam that passes through the interaction medium loses energy on its path. Therefore, the path that a particle follows through a medium is highly relevant for the DD effects that take place. Combining the information of the source distance of each voxel in combination with the density of the patients anatomy results in the radiological depth. It is calculated by the summation of the partial path of a casted ray through each intersected voxel in the medium multiplied by the HU value of each respective voxel. Information about the density of prior tissues is given in each voxel value. Therefore the same spatial depth in a dense material results in a higher radiological depth than the same depth in soft tissue.

Particles interact negligibly in air, therefore all masks were set to zero where the CT masks was 150 HU over the value of air. Air or low density volumes inside the body of the patient were not affected by this step, to account for the ERE on the surface of air cavities inside the patient.

4.2.2 Network Output

The network combines the information from the five 3D volumes and yields a single volume with the respective dose distribution. Spatial dimensions for width height and depth as well as voxel dimensions are preserved during inference. A respective dose distribution (f) for the input ((a) - (e)) from a segment taken from a prostate cancer treatment plan is depicted in fig. 3.

4.3 Training Data Generation

All training data was generated from patient data taken from the institutional database of radiotherapy treatment plans. Information for all input masks are given in the CT scan, the dose file as well as the plan file. The patients anatomy for the CT masks is taken from the CT files and adjusted to the right slice thickness, due to different aquisition protocols using 2 mm or 3 mm slice spacing. Dose distribution used as training target for each segment were calculated using the EGSnrc open source software package provided under [48]. The work of Friedel et al. [49] enabled us to accurately simulate single segment dose distributions for the MR-Linac. Simulation of all segments was done using a remote high-performance computing (HPC) solution provided by the state of Baden-Württemberg. Coordinate system orientation of the patient's anatomy as well as MLC and gantry positions are given in the dose and plan file. Simulation of 10^7 particles took around 4 hours on average. The 3D input volumes to the network were calculated using in-house developed python scripts with the provided information from CT, dose and plan files. The central beam line distance and source distance were calculated using the isocenter information from the DICOM-plan file and the respecitve coordinate system provided from the CT-DICOM files. The CT mask was directly taken from the CT images and resized to fit the spatial resolution. Ray tracing was used to calculate the radiological depth mask with information from the CT and the gantry position which was taken from the DICOM-plan file. Information about each segments MLC configuration and gantry angle were used to project the beams eye view into 3D space for the beam shape mask.

4.4 Network Training

Four Nvidia GTX 2080 Ti, provided by the cloud of the Tübingen Machine Learning Cluster of Excellence, were used to increase the GPU memory to 44 GB. This enabled a batchsize of 128 with a patch size of $32 \times 32 \times 32$ voxels resulting in a spatial field of view of $37.5 \times 37.5 \times 96$ mm^3. After every epoch, the model performance was assessed on

the validation set and the top 5 models were saved. The ADAM optimizer was used with 0.9, 0.99, 10^{-8} for β_1, β_1 and ϵ respectively. An initial learning rate of 10^{-4} was used. If the validation loss did not decrease or the gamma passrate on the valdiation set did not increase over 50 epochs the learning rate was decreased by a factor of 10. When a learning rate of 10^{-6} was reached and no performance increase was noted over 50 epochs the training was stopped to prevent overfitting and to save computational recources.

4.4.1 Dataloading

Memory usage is a serious concern in the application of deep learning and espeacially when dealing with 3D data sets. In our case not only the memory usage in the graphics processing unit (GPU) but in the random-access memory (RAM) is of importance. In conventional applications the network training is divided into epochs, where the network sees every single instance before validation. The present training data is therefore loaded into the RAM and can be quickly accessed from there, to pass it onto the GPU memory. Our here posed problem of dose deposition prediction is a 3D problem with input volumes of multiple hundreds of megabytes per input segment. Loading all data into the RAM was therefore not applicable. A solution to this problem is to load individual segments on the fly into the RAM and then process it. This solution was not applicable for us because loading times were long compared to processing the segment during training, which would significantly slow down training. We therefore developed a partial on the fly dataloading inspired by the *Queue* class from Pérez-García et al. [56] open source python libary *TorchIO*. Classical augmentations in the form of shearing, warping or zooming were not applicable in our case because of the trivial relations between the augmentation and the DD. To still achieve additional variation in training data we employed patch based training. With this approach a multitude of patches can therefore be extracted from the same segment volume, increasing the variety in training data. To combine the semi-on-the-fly dataloading and the patch based training, multiple volumes are loadded into the RAM and then processed together as depicted in fig. 4. A fixed amount of training data is loaded from the randomized training data pool into the RAM. From this set of training volumes a subset of patches from each volume is extracted and stored in a queue of patches, which is then shuffled and batch wise passed onto the GPU and the network.

4.5 Evaluation Metrics

Conformality of dose distributions are clinically assessed using the gamma-index metric first introduced by Low et al. [57] in 1998. The evaluation metric is composed of two parametric values that set the criteria for which dose distributions are analysed. Spatial
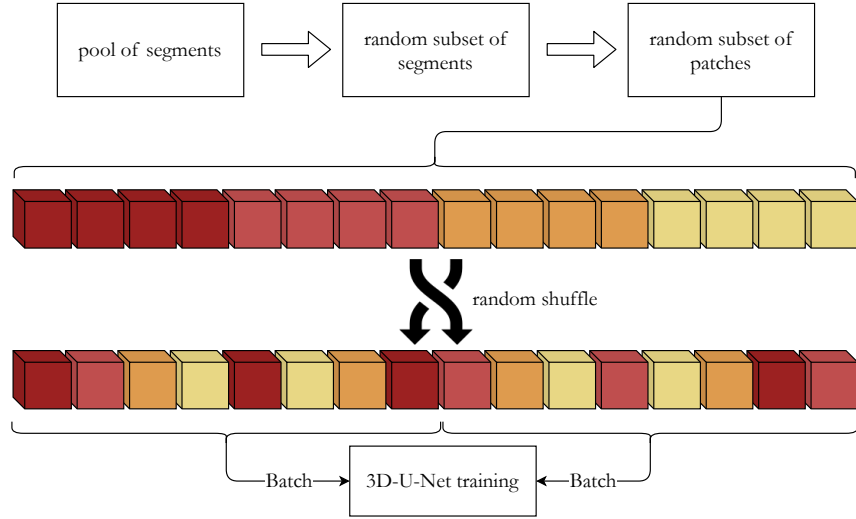
Figure 4: Dataloading scheme for memory efficient patch based dataloading for 3D training volumes. A subset of patches from one training volume is depicted in one color shade.

deviations as well as deviations of dose are respected when analysing. Dose conformality is assessed by analyzing each individual voxel of a given dose distribution using the following equation:

$$\Gamma(r_m, r_c) = \sqrt{\frac{r^2(r_m, r_c)}{\Delta d_M^2} + \frac{\delta^2(r_m, r_c)}{\Delta D_M^2}} \tag{1}$$

where ΔD_M and Δd_M are the dose difference and a spatial criterion respectively, in our case 3% and 3 mm. $r^2(r_m, r_c)$ and $\delta^2(r_m, r_c)$ are the squared spatial distance and dose difference from the reference point r_m to evaluation point r_c respectively. If $\Gamma \leq 1$ the criterion has passed. By evaluating the entire volume in that manner an overall gamma passrate can be calulated with the following formula.

$$\gamma = \frac{NoT(\Gamma \leq 1)}{NoT} \tag{2}$$

where NoT is the **N**umber **O**f **T**ests, in this case the number of voxels in the entire volume and $NoT(\Gamma \leq 1)$ is the number of tests that passed the gamma criterion following eq. 1. The equations eq. 1 and eq. 2 hold true for single as well as multiple dimensions. A 2D example for a passed as well as a failed gamma test is given in fig. 5.
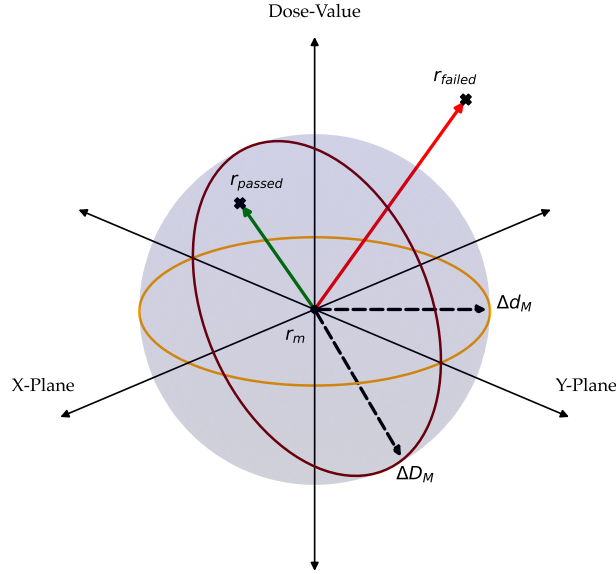
Figure 5: Gamma test for two points r_{passed} and r_{failed} and the reference point r_m for a 2D (X,Y) space. Orange and brown rings indicate distance and dose parameter acceptance margins respectively. The grey sphere represents the set of all points that pass the gamma test. Green and Red arrow indicate a passed and failed gamma test for the given reference point r_m, dose and spatial parameter.

4.6 Hypotheses and Experiments

In the following section we desribe several experiments which aim to investiage a certain hypothesis. All subsequent hypotheses are based on the previous listed hypothesis respectively and thus describe the developement process of the here proposed network.

4.6.1 General Applicability

Hypothesis 1

The network is able to learn the general dose deposition process from the given input taken from a multitude of RT segments of one tumor site. Dose predictions for segments of the same entity are accurate and robust.

Experiment 1

The proposed 3D-UNet was trained on a mutlitude of prostate cancer patient radio treatment segments. Dose conformality was assessed with the gamma passrate for each individual test segment as well as entire test plans.

4.6.2 Poor Initial Translatability

Hypothesis 2

Different tumor sites vary drastically in segment shape, orientation and source to surface distance (SSD). The SSD is, compared to other entities such as HN, approximately constant for prostate cancer patients. The same holds true for tissue homogenity aspects of the respective body regions, when comparing lower abdomen to e.g. HN. Therefore, we hypothesize that the network will not achieve the gamma pass rates achieved so far nor sufficient ones when tested at different tumor sites and patient anatomies.

Experiment 2

The from *4.6.1 General Applicability* trained network was then used for inference on additional testing data from a multitude of tumor entities as well as patient anatomies to test its translational capabilities to a varying set of input characteristics.

4.6.3 Increased Robustness

Hypothesis 3

The inclusion of a broader variety of segment shapes, sizes and positions inside the patients anatomy, as well as differing body regions and therefore varying SSD values and tissue densities results in a better robustness of the networks prediction accuracy. We hypothesize that significant increases in accuracy regarding new included entities can be achieved while the accuracy for the previously present remains stable.

Experiment 3

Network architecture as well as training procedure of the in *4.6.1 General Applicability* trained network were kept, while adding data from liver, breast and HN segments into the pool of training data. The total amount of segments remained approximately constant and the ratio of treatment plans per entity was equal. The translational capabilities were then assessed by testing the newly trained network on the entities that

were included in the training data as well as additional lymph node test cases. Significane of results was assessed using the Wilcoxon singed-rank test. With this test design we investigated if the network performance decreased on prostate data by including new tumor sites into the training data and we assessed its performance on seen as well as unseen treatment plan data. We used lymph node plan data as the unseen test data because lymph nodes are present all over the human body, which makes this anatomies aswell as RT plans of this tumor site especially heterogeneous. Due to the small size of lymph nodes, segment fieldsize was therefore especially small, which we assumed to be the biggest challenge to the network.

4.6.4 Underlying Physics

Hypothesis 4

We hypothesized that basic physical dependencies, such as accurate dose depth profiles, correct mapping of penumbra and impact of the distance squared law will be learned by the network from the includation of a multitude of SSD, patient anatomies, segment shapes, angles and sizes into the training data. Therefore we assumed that the predictions of dose deposition in basic homegenous phantoms will be accurately predicted with the training on the given heterogeneous dataset.

Experiment 4

Artificial phantoms are used to investigate the physical accuracy of the dose predictions of the networks from *4.6.1 General Applicability* and *4.6.3 Increased Robustness* at various positions. Phantoms consist of a water volume placed inside a $600 \times 600 \times 600 \ mm^3$ air volume at differing positions. With the default voxel dimensions of $1.1718 \times 1.1718 \times 3 mm^3$ this results in a volume of shape $512 \times 512 \times 200 \ voxel$ for the phantom. Water phantom dimensions will be $100 \times 512 \times 200 \ voxels$. Comparing dose cruves in the different spatial directions of a $10 \times 10 cm^2$ square field and $0°$ beam angle and the respective dose prediction yields information about the networks capability to learn the underlying physics.

4.7 Patient Data

All data used in the course of this study was taken from patients previously treated at the MR-Linac in our institution. All patients gave their informed written consent

to this study, which was approved by the local ethical committee (ethics approval No. 659/2017BO1). All treatment plans were created by a medical physicist in agreement with an oncologist. The training data consisted of two main datasets. One was used for the training of the in *4.6.1 General Applicability* proposed network, which was only trained on prostate treatment data and consisted of 45 treatment plans with a total number of 2342 segments. The second dataset consisted of 15 treatment plans each for prostate, liver, breast and HN tumor sites with 720, 819, 656 and 929 segments respectively. Additional 15 LN cancer treatment plans with a total of 659 segments were used only for testing. A comprehensive summary of all training and testing data is given in tab. 1 including training, validation, test split and information about the fieldsize distribution of training and test data.

		Prostate	Mixed				Testing
			Prostate	*Liver*	*Mamma*	*H&N*	*LN*
Patients	#	45	15	15	15	15	15
Segments	#	2342	720	819	656	929	659
Split	#/#/#	36/4/5	8/2/5	8/2/5	8/2/5	8/2/5	0/0/15
Fieldsize Training	cm^2	36.5 (17.9)	35.5 (18.0)	24.1 (18.4)	40.7 (28.5)	63.0 (50.5)	N/A
Fieldsize Testing	cm^2	34.5 (15.8)	34.6 (15.9)	22.8 (14.4)	40.6 (38.0)	68.9 (53.6)	26.0 (25.6)

Table 1: Patient data information for prostate only trained network and mixed entity trained network and test data set consisting of LN. Split is given for training / validation / testing. Fieldsizes are given as mean (standard deviation).

5 Results

5.1 Experiment 1: General Applicability

The training of the model took approximately 11 days until convergence. Training stopped at $6.7 \cdot 10^5$ due to no improvement in gamma and validation loss. The model reaches mean gamma passrates of $94.35\% \pm 5.99\%$ (min: 58.02%, max: 99.96%) for prostate test segments and a mean gamma passrate of $99.1\% \pm 1.3\%$ (min: 96.8%, max: 100%) for the five plans. Very good dose agreement between target dose, caluclated by MC simulation and the dose prediction can be observed as depicted in fig. 6 showing the isocentric slice of a prostate cancer patient with a gamma passrate of 99.9%. As shown, the model is able to accurately predict the beamshape for single segments as well as the penumbra and beamwidening. The dose volume histogram (DVH) shown in fig. 7 gives crucial information about the applied dose inside target volumes as well as OAR for a test plan of the prostate patient cohort. DVH curves overlap for all structures but urethra, which is caused by the high contribution of single voxels to the entire volume. Very good dose agreement for high as well as low dose regions is reached.

Figure 6: Dose prediction example on isocentric slice on prostate cancer patient from test cohort with prostate data only trained model. From left to right: target dose; dose prediction; gamma map for isocentric slice with a gamma pass rate of 99.9%

5.2 Experiment 2: Poor Initial Translatability

When applying the prostate only trained network on body entities deviating from the lower abdomen its accuracy decreases significantly. A summary over the prediction accuracy for all segments as well as the summed plans for all included test data of different entities is given in tab. 2.

The model shows particular problems with breast cancer and HN segments with a mean gamma passrate of $57.8 \pm 18.0\%$ (min: 1.9% max: 94.7%) and $60.5 \pm 11.8\%$ (min: 31.1%
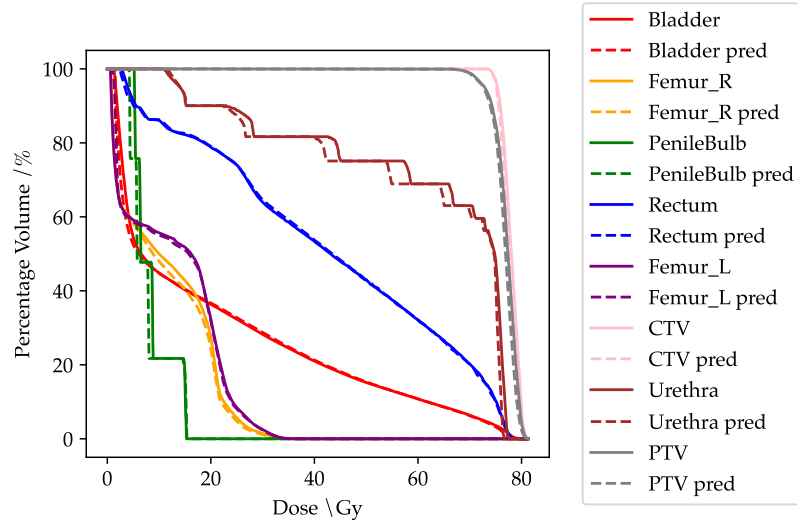
Figure 7: DVH for prostate cancer test plan with gamma passrate of 99.9%.

		Prostate		**Liver**		**Mamma**		**H&N**		**LN**	
		Seg.	*Plans*	*Seg.*	*Plans*	*Seg.*	*Plans*	*Seg.*	*Plans*	*Seg.*	*Plans*
	#	306	5	272	5	419	5	379	5	659	15
Mean		94.3	**99.1**	75.4	89.9	57.8	67.4	60.5	77.4	77.4	93.0
Median		**96.2**	**99.6**	77.4	91.8	56.4	66.2	61.0	78.8	80.2	92.8
STD	/%	6.0	1.3	12.6	4.1	18.0	5.3	11.8	10.2	13.8	6.1
Min		58.0	96.8	38.9	84.0	1.9	60.6	31.1	62.5	36.0	75.3
Max		100	99.9	99.4	94.4	94.7	73.0	91.5	90.2	99.7	99.7

Table 2: Gamma rassrates for segments as well as plans for prostate, liver, mamma, HN and LN predicted with the prostate-only trained model. Gamma test settings: 3 mm/3% lower percentage cutoff 10%

max: 91.5%). An examplary case with a gamma passrate of 60.6% is shown in fig. 8 depicting low accuracy in the high dose as well as lower dose areas of the volume.

5.3 Experiment 3: Increased Robustness

The model training took approximately 9 days until convergence. Training stopped after $5.0 \cdot 10^7$ patches due to no improvement in gamma and validation loss. Testing was done on prostate, liver, mamma and HN plans, which were contained in the training set as well as LN plans which were not included. A summary of the reached gamma passrates is given in tab. 3, showing significantly improved gamma passrates with respect to *Experiment 2*, see fig. 9 *left*. Mean gamma pass rate for single segments changed
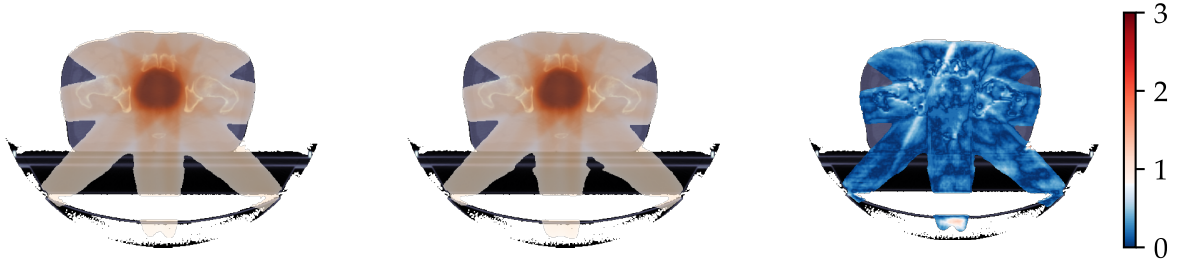
Figure 8: Dose prediction example on isocentric slice on breast cancer patient from test cohort with prostate data only trained model. From left to right: target dose; dose prediction; gamma map for respective slice with a gamma pass rate of 60.6%

by -0.88%, +14.4%, +19.11%, +19.65%, +7.75% for prostate, liver, mamma, HN and LN respectively. The mean gamma passrate shows increased performance for the plan analysis of all tumor entities with an increase of +0.1%, +7.7%,+16.2%, +11.4%, +3% for prostate, liver, mamma, HN and LN plans respectively. A comparison between the results of the individual segments and the overall plans showed that, on average, the segments with the higher contributions to the dose distribution of the entire plan were better predicted than less contributing ones.

		Prostate		**Liver**		**Mamma**		**H&N**		**LN**	
		Seg.	*Plans*	*Seg.*	*Plans*	*Seg.*	*Plans*	*Seg.*	*Plans*	*Seg.*	*Plans*
	#	306	5	272	5	419	5	379	5	659	15
Mean		93.5	**99.2**	89.8	**97.6**	76.9	83.6	80.1	88.8	85.2	**96.0**
Median		**95.0**	99.7	91.6	**97.4**	78.9	83.7	82.8	90.9	87.8	**98.8**
STD	/%	6.3	1.0	7.0	1.0	12.8	4.6	10.9	8.6	10.7	5.8
Min		39.2	97.5	50.8	96.6	18.4	77.0	49.3	76.0	38.7	77.9
Max		99.8	99.9	99.1	99.2	98.5	89.8	97.4	98.5	99.9	99.9

Table 3: Gamma rassrates for segments as well as plans for prostate, liver, mamma, HN and LN predicted with the mixed data trained model. Gamma test settings: 3 mm/3% lower percentage cutoff 10%

To get a deeper insight into the networks method of operation and to enlarge the interpretability of the network, we analysed the the perfomance of the network regarding field size and segment weights as depicted in fig. 10, which gives information about the correlation of field size and prediction accuracy of single segments. Relative occurence of segment sizes as well as weights are displayed as a bar chart in the background of both plots. Around 82% segments have a field size smaller than 60 cm^2 that is also the field size until network accuracy increases with increasing field size for both the prostate-only

and mixed entity trained model. Fieldsizes larger than 60 cm^2 are not predicted as stable with fluctuating gamma passrates for both models. The prostate-only trained model expieriences a larger drop in accuracy for field sizes larger than 60 cm^2 and accuracy is steadily decreasing with increasing field sizes. It was observable that smaller field sizes are less robust and range from 1.9% to 100% and 18.5% to 99.9% for the prostate-only and the mixed entity model respectively on a field size of 0 to 15 cm^2. Gamma passrate analysis with respect to segment weight showed the same trend of increased accuracy and robustness with increasing weight values. Segments with weights under 0.2 made up for almost 80% of all segments. A comprehensive and direct comparison between prostate-only and mixed entity model prediction accuracy for segment and plan predictions is given in tab. 4 and tab. 5 respectively.
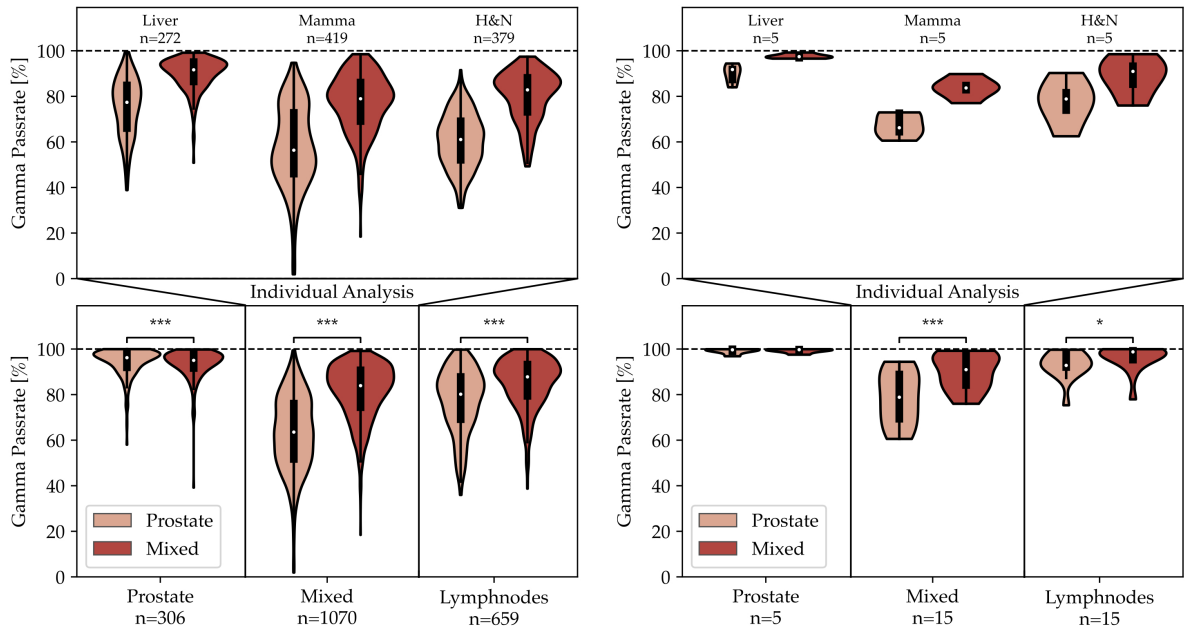
Figure 9: Gamma pass rate comparison for single segments (left) and summed up plans (right) between prostate-only and mixed entity trained models. Boxplot are displayed inside respecitve violin plots. Individual analysis of mixed entity set, composed of liver, mamma and HN displayed on top plot. Significance level was assessed using Wilcoxon singed-rank test with significance level p < 0.05 (*) p < 0.01 (**) and p < 0.001 (***).
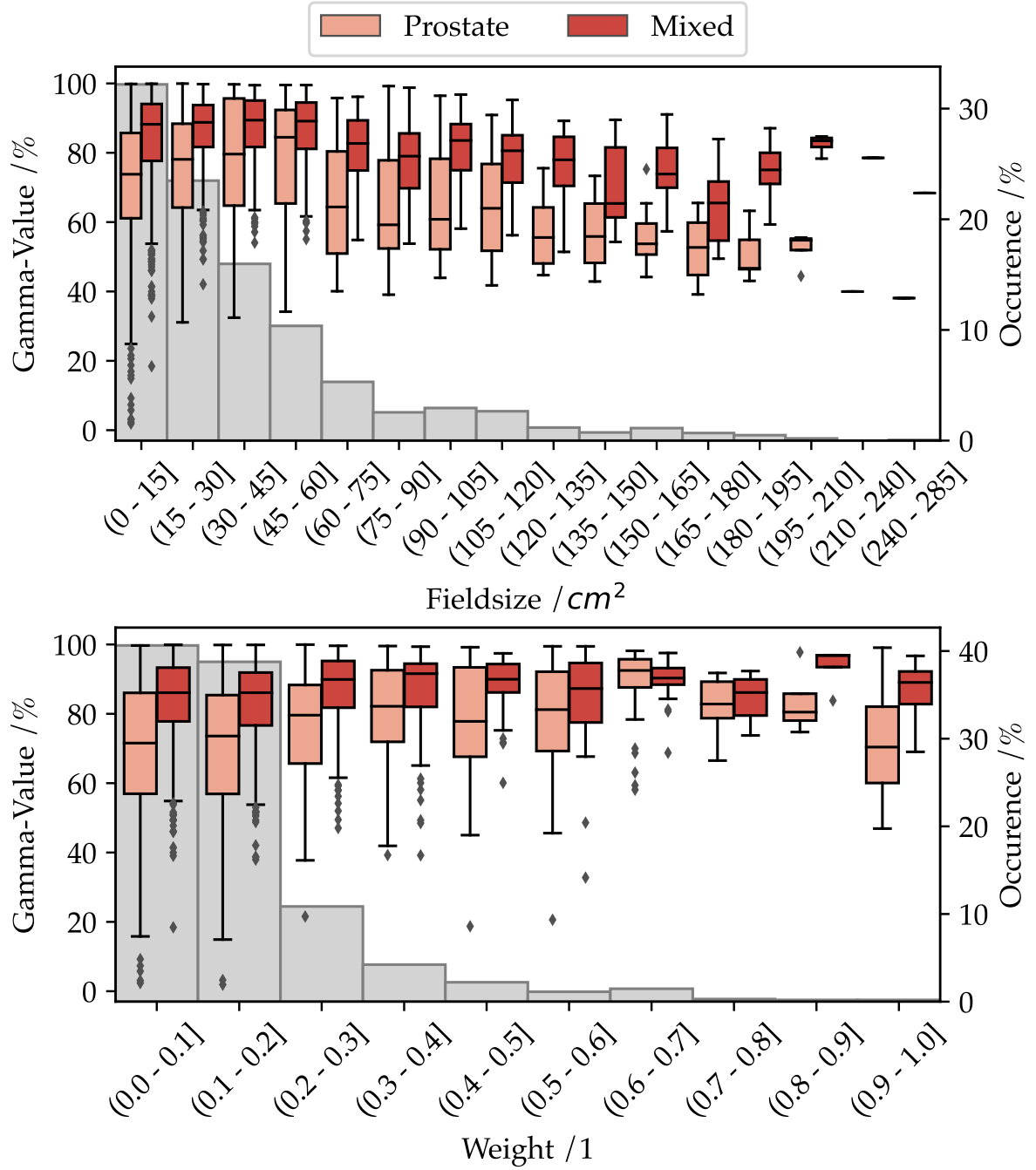
Figure 10: Prediction accuracy of each segment with respect to field size for prostate only and mixed trained model. Relative occurence of discretized field size bin is displayed in the background.

		Prostate		Liver		Mamma		H&N		LN	
Model		*Prostate*	*Mixed*	*Prostate*	*Mixed*	*Prostate*	*Mixed*	*Prostate*	*Mixed*	*Prostate*	*Mixed*
Mean		94.3	93.5	75.4	89.8	57.8	76.9	60.5	80.1	77.4	85.2
Median		**96.2**	**95.0**	77.4	91.6	56.4	78.9	61.0	82.8	80.2	87.8
STD	/%	6.0	6.3	12.6	7.0	18.0	12.8	11.8	10.9	13.8	10.7
Min		58.0	39.2	38.9	50.8	1.9	18.4	31.1	49.3	36.0	38.7
Max		100.0	99.8	99.4	99.1	94.7	98.5	91.5	97.4	99.7	99.9

Table 4: Gamma passrate comparison of prostate-only and mixed entity trained models on the test dataset consisting of 306 prostate, 272 liver, 419 mamma, 379 head & neck and 659 lymphnode segments. Gamma test criteria: 3 mm/3% with a lower dose cutoff of 10%.

		Prostate		Liver		Mamma		H&N		LN	
Model		*Prostate*	*Mixed*	*Prostate*	*Mixed*	*Prostate*	*Mixed*	*Prostate*	*Mixed*	*Prostate*	*Mixed*
Mean		**99.1**	**99.2**	89.9	**97.6**	67.4	83.6	77.4	88.8	93.0	**96.0**
Median		**99.6**	**99.7**	91.8	**97.4**	66.2	83.7	78.8	90.9	92.8	**98.8**
STD	/%	1.3	1.0	4.1	1.0	5.3	4.6	10.2	8.6	6.1	5.8
Min		96.8	97.5	84.0	96.6	60.6	77.0	62.5	76.0	75.3	77.9
Max		99.9	99.9	94.4	99.2	73.0	89.8	90.2	98.5	99.7	99.9

Table 5: Gamma passrate comparison of prostate-only and mixed entity trained models on the test dataset consisting of 5 prostate, 5 liver, 5 mamma, 5 head & neck and 15 lymphnode treatment plans. Gamma test criteria: 3 mm/3% with a lower dose cutoff of 10%.

5.4 Experiment 4: Underlying Physics

Both the prostate-only as well as the mixed entity model perform poorly regarding gamma passrates for all water phantom positions tested. The model accuracy is steadily decreasing with increasing SSD with gamma pass rates of 60.1%, 45.5%, 48.4%, and 56.1%, 27.7%, 19.2% for the prostate-only and mixed entity model at depths of 100 pixels, 200 pixels and 300 pixels respectively. A summary of the accuracy is given in tab. 6. The transversal view in fig. 11 shows that the prostate model is better at representing the peak at the beginning of the phantom than the mixed model for a depth of 100 pixels. At higher phantom distances the prostate model resembles the qualitative dose distribution quite well, while failing at achieving a quantitatively correct predictive value. The mixed model cannot reach the correct peak height at the beginning but predicts the later curve very well in terms of height and curvature in the transversal view. Coronal and sagittal view both show that the penumbra is resembled very well by both models. At phantom depths of 200 and 300 pixels the shoulders of the dose profile drop too fast for the mixed trained

Phantom position /pixel	**100**		**200**		**300**	
Model	Mixed	Prostate	Mixed	Prostate	Mixed	Prostate
Gamma Passrate /%	60.1	56.1	45.5	27.7	48.4	19.2

Table 6: Gamma passrates for prostate only and mixed trained models for different water phantom phantom positions inside an air volume. Water phantom thickness was 200 pixels.

model in the coronal view. Slight fluctuations at the phantom's edge can be observed at 100 pixel depth for the mixed model in the sagittal view.
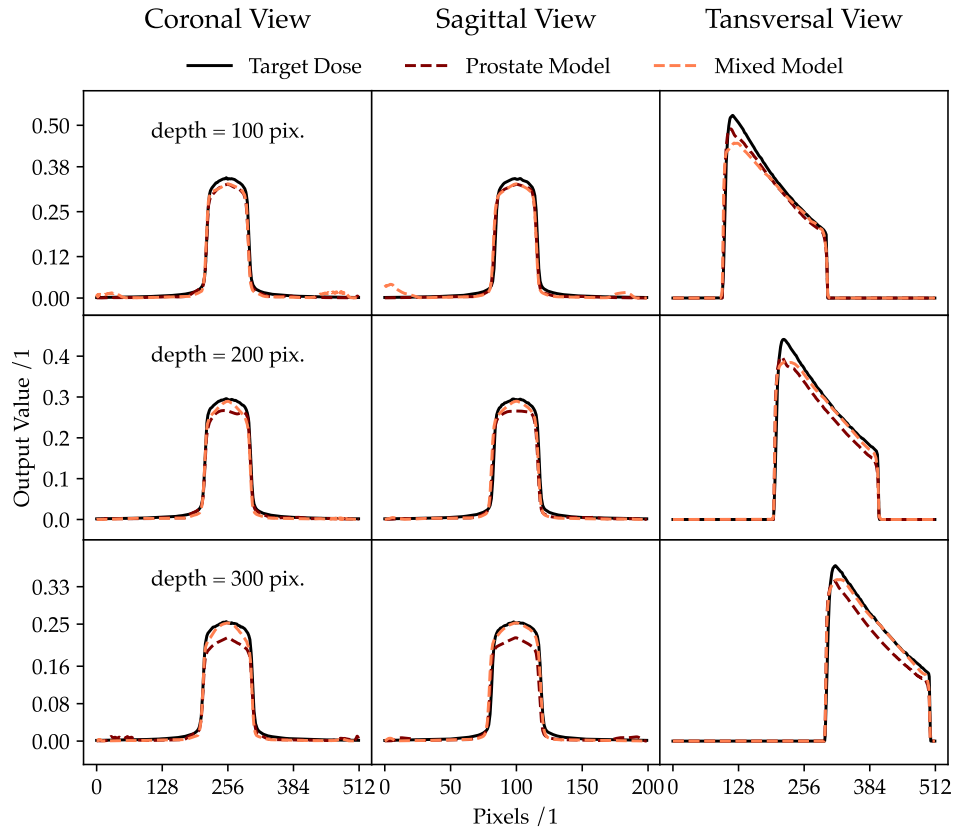

Figure 11: Coronal, sagital and transversal view of the dose distributions for 100, 200 and 300 px depth in the phantom volume

6 Discussion

6.1 Single entity training

We have shown that training the proposed 3D-UNet with the given training algorithm and dataloading scheme can reach very good accuracy concering the provided entity. It reached very high dose conformaty for single tested segments aswell as entire treatment plan dose distributions with mean gamma passrates of 94.4% ± 6.0% and 99.1% ± 1.3% respectively. DVH analysis, depicted in fig. 7, also shows very good agreement in high dose areas, such as CTV and PTV, aswell as OAR such as rectum or femur heads. Accuracy in and DVH conformality of OAR of interest for quality assurance of treatment plans indicate that the proposed network can be used as a secondary dose engine or as a dose verifcation tool. Dose prediction is highly accurate for the entirety of prostate plans and segments with different field sizes, shapes, irradiation angles, and patient anatomies. With prediction times of 3 seconds per segment, the goal of real-time dose prediction is approached. Further developments towards the use of parallelization, the inference time could be further reduced. Nonetheless, 3 seconds is a significant improvement over 4 hours of MC simulation time on an HPC solution that already benefits from paralization using the EGSnrc software tool.

In contrast to other previous works such as Kontaxis et al. [54], that only included segments of five predefined beam angles into the training data, we did not limit us to specific beam angles but rather took all provided angles from the planning system. We further improved on their work by increasing the voxel size in sagital and coronal directions to increase the resolution of the dose distributions and to be able to achieve higher accuracy in high density gradients.

Testing against not included tumor sites have shown the limits of this single entity trained approach. Rapidly decreasing gamma passrates with a lowest passrate of 1.9% for a breast cancer segment were reached. Particularly low gamma-pass rates were obtained at tangential irradiation angles, where the radiation field passed only through part of the patients anatomy, in our case the breast. Beamshape and orientation of the predicted dose are correct, but the network fails at predicting the respective value of each voxel. Predicted values are significantly smaller than the desired target value. Three poorly predicted segments that intersect only parts of the breast are depicted in fig. 12. We assume that due to the offset of the radiation target and the associated larger distance between the target and the central beam line, the effects of the distance mask for the central beam line are overweighted and thus reduce the predicted value.
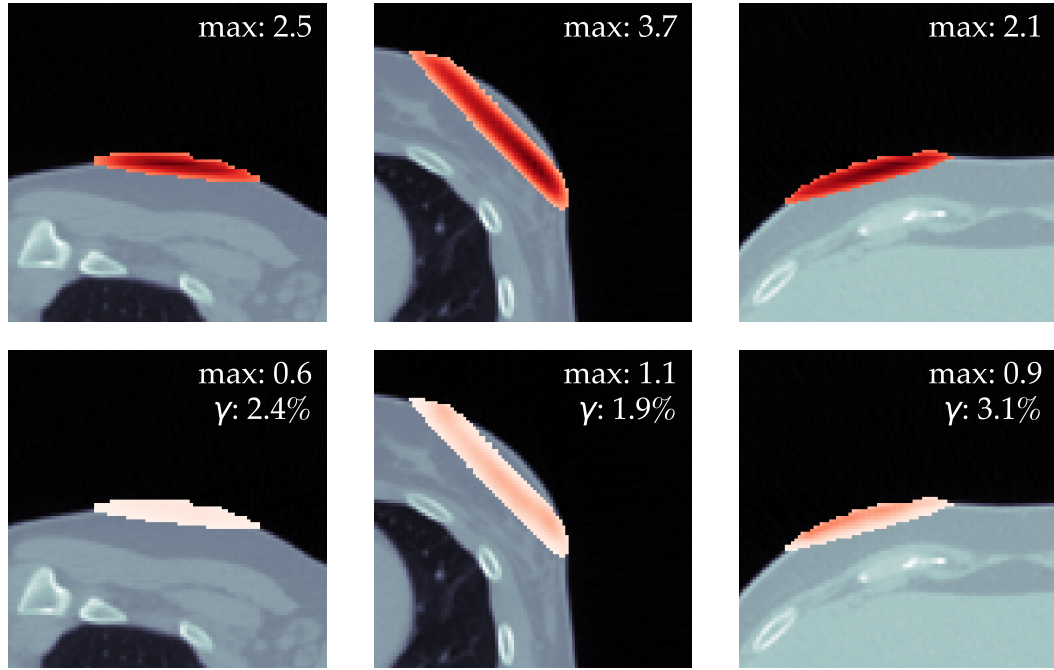
Figure 12: Exemplary images of target dose distribution of tangential segments on breast cancer patients with very small gamma-passrates (top) and the respective dose prediction from the prostate only trained model (bottom).

Decreasing performance for tumor sites not included into the training data is most probably occuring due to the increased complexity of unseen anatomies. The position of the prostate and the surrounding OAR is quite defined inside the humans anatomy irrespective of the individual patient. Therefore prostate cancer treatment plans have a very similar isocenter, aswell as the majority of highly contributing segments have similiar shape and sizes. In contrast HN, LN or even mamma treatment plans have a wider range of segment shapes and fieldsizes (depicted in tab. 1: row fieldsizes). When looking at the anatomy of a patients lower abdomen the intersubject variation in SSD is small compared to an inter-entity comparison to HN. In fig. 13 three anatomies from prostate, HN and breast cancer are depicted, with the respecitve HU distribution in a slice throught the isocentric slice. The variety in HU in a prostate cancer patient is comparable with the one of a water phantom, it has a "box-like" shape, while crossections from HN and breast cancer show larger variation in width of patient contour as well as HU value variation inside the patients anatomy. Due to increased variation in SSD on HN patient anatomies compared to prostate patients the impact of the distance-squared law is significantly higher. We assume that the network is not able to correctly map the impact of the squared distance-law on the dose distribution due to the small variatzion in SSD when using only prostate
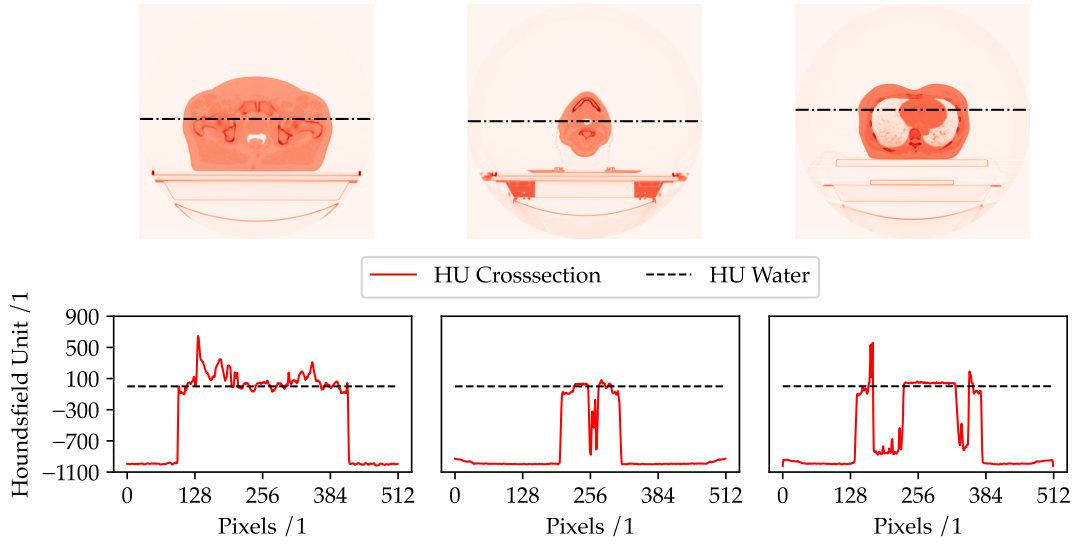
Figure 13: Top: Image of the isocentric slice from a prostate, H&N and breast cancer patient respectively. Bottom: Cross section through the patients anatomy at the marked position. HU value of water indicated with a dotted line.

training data.

There exist a multitude of radiation protocols, which vary depending on the insitutional guidelines aswell as applications or accelerator type. These different treatment modalities have inherent a variety of different parameters, such as radiation angles, field sizes, and beam shapes. Therefore a robust dose prediction inrrespective of the target volume, beam angle and MLC shapes is of crucial importance, and a careful veryfication process is needed.

6.2   Mixed entity training

The number of patches until convergence was lower for the mixed entity model than for the prostate only trained model, at 5.0×10^7 compared to 6.7×10^7 patches. This suggests that the model learns the important aspects of dose deposition more quickly with the greater diversity of training data than with a relatively homogeneously distributed training data set. Model performance and dose prediction accuracy increased for the newly added entities aswell as the LN segments aswell as plans. Slight decreases in model accuracy regarding single segments of prostate treatment plans are noticed. This is most likely due to the increased generalization of the model and the decreased adaption to prostate plans only. Nevertheless the increased generalization leads to a better perfomance in the mean for the entire prostate plans compared to the prostate-only trained model. We have observed that the mean single segment prediction accuracies for all entities are smaller

than the mean plan prediction accuracies. As depicted in th lower plot of fig. 10, the higher weighted segments tend to be predicted better for both models leading to the effect of better predicted plans.

Predictions for breast cancer segments are particularly bad for the prostate-only aswell as the mixed-entity trained model. Beside the challenging anatomy of air cavities as well as tangential fields only partially hitting the patients body, the internal treatment protocol has changed between training and target dataset. Breast cancer treatment plans from the test dataset are significantly higher modulated, with 23.7 ± 5.6 and 83.8 ± 15.3 segments per plan for old and new institutional guidelines respectively. This higher modulation results in more smaller segments and more complex MLC configurations. Nevertheless the newly adapted model showed promising results for all tested entities regarding entire plan dose distributions. As the model needs specific accelerator settings to predict a dose for a specific segment, this model is not usable as a primary dose engine for a radio treatment planning system. However due to its promising accuracy concerning DVH parameters and its short inference times, it may be promising as an online dose verification or a secondary dose engine after further development.

6.3 Dose deposition physics

Water phantoms are used clinically for dose verification and accelerator constancy testing. They are not only the most basic model of a target volume, but dose distributions in them are very well described in literature. Due to the homogenous density distribution inside the phantom physical processes of dose distribution can be analyzed very well with a water phantom. We therefore choose a synthetic water phantom to assess the capability of the model to learn the underlying phyisics of the dose deposition process. Dose curves were resembled by the networks prediction quiet well, dose peaks at the beginning of the phantom were present. The exponential dose decrease aswell as the rapid dose drop to the zero level at the end of the phantom has been mapped well qualitatively. The prostate-only trained model showed particular problems reaching the correct predicted height for entire dose curve. Scaling the curves to the correct maximum, as shown in fig. 14, gives a more comprehensible picture of the curvature for the prostate-only and mixed entity models. As depicted the qualitative curvature is closer resembled by the prostate model. When comapring different entity crosssections, as shown in fig. 13, the prostate density distribution is comparable to the one of a water phantom. Therefore it is probable that the model has not learned the underlying pyhsics but predicts the qualitative dose distribution in a water phantom better due to the similarities between the density distribution of the lower abdomen and the one of a water phantom.
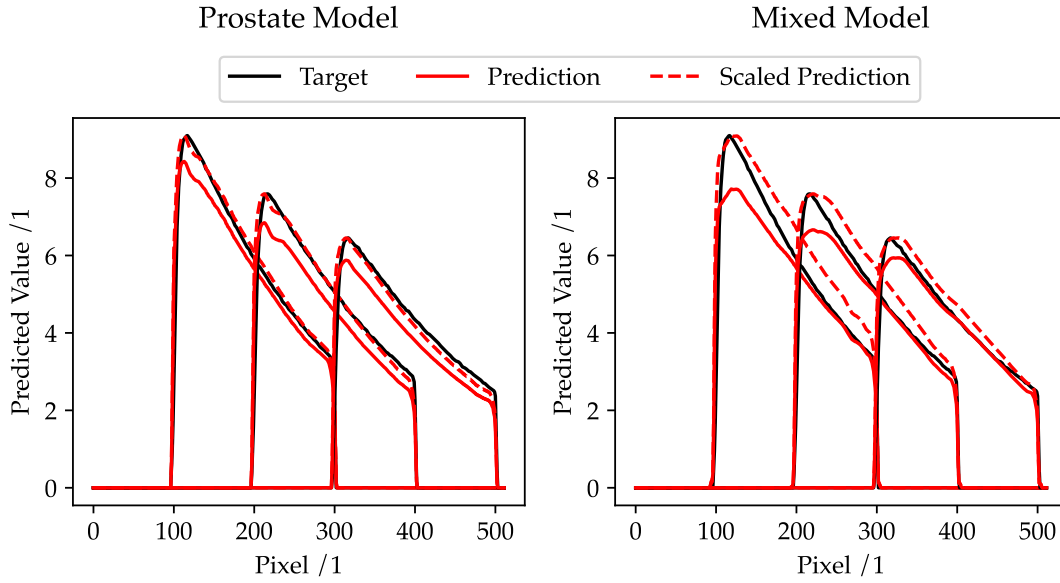
Prostate Model   Mixed Model

Figure 14: Target, predicted and scaled predicted depth dose curve for water phantom positions at 100, 200 and 300 pixels depth, respectively, from left to right. Left: prostate-only trained model. Right: mixed entity trained model.

Both models have problems with the representation of the dose deposition peak at the beginning of the phantom. Taking this into account could explain why the dose from the tangential radiation fields represented in fig. 12 is underestimated in the breast cancer plans. The prediction accuracy of the first 100 pixels of the segment is quite poor, so segments whose radiological path is less than 100 pixels are predicted very poorly. Contrary to the first 100 pixels, the further dose distribution is well predicted by the mixed entity model. The majority of plans have their target volume and therefore the high dose volume located at higher depths. As the models prediction of dose curvature at greater depths become more accurate and the gamma passrates only accounts for high dose areas, the gamma passrates increases for target volumes at higher depths.

6.4   Model improvement

The here proposed model showed promising results as a first step towards a fast, accurate and robust dose enginge for an online treatment workflow using an MR-Linac. We have confirmed that the genereal concept of a spatial information informed 3D-UNet is applicable for the prediction of dose deposition processes in patients anatomy as proposed by Kontaxis et al. Nevertheless the evaluation of dose prediction accuracy has shown outliers, especially on challenging tasks such as dose distributions in highly heterogenous tissues such as air cavities aswell as very small fields with short radiological paths inside the

patients anatomy. The network has shown short estmiation times with approximately 3 seconds per segment. To improve inference time, investigations towards the exploration of other applicable memory efficient network architectures are needed, aswell as the utilization of a multi GPU prediction process at inference. In detail analysis of mask impact for every provided mask, as partially done by Kontaxis et al. [54], might yield benefitial results for the interpretability of the networks performance as well as its limitations towards a robust and reliable dose prediction. As discussed in subsection 6.1, we assume that the center beam line distance, might have a negative effects regarding offset segments.

We have shown that incorporating more heterogenous training data improves the robustness and the prediction accuracy of the model. Our training data consisted of patient specific segments take from entiree radio treatment plans that was genereated by the intitutional treatment planning software. This way we had no control over how segments were shaped or oriented. To achieve desired distributions for fieldsizes and gantry angles and to gain control over the segment shapes and the respective isocenter, we propose to further increase the training data variability by generating artificial segments. By using artificial segments, we can control the exact distributions of gantry angles aswell as fieldsizes over the entire training data pool, from which we promise to achieve an increased representation of otherwise underpresented segment shapes, sizes and areas of uncommon iso centers.

Data from patients that were previously treated at the MR-linac in our institution is very limited. We therefore propose to use, in addition to artificial segments, the anatomy CTs of not only treated patients at the MR-Linac, but any patient CT. In this way, we deviate from the small pool of CTs of cancer patients from our radiotherapy department and expand the pool to all patients who have had a CT of any body region. This decreases the change of the network overfitting to specific anatomic distribrutions of the given patient anatomy pool and contributes to the robustness and generalization process that is desired to be achieved. Implementing this approach would suit as a first proof of concept. Further data individualization can be achieved by deviating from the patient's anatomy and creating artificial target volumes with heterogeneous areas within the irradiated volume. The position, shape and density of these created areas inside the target volume, can be choosen based on the desired distribution as implemented by Neishabouri et al. for their training data generation.

Further development could go towards exploring an LSTM approach that provides 2.5 - D or 3 - D dose distributions in the direction of the irradiation angle, as suggested by Neishabouri et al. in their work on proton dose predictions for heterogeneous tissues [58].

Approching the problem in this manner might lead to the networks incorporating of previously passed tissues. As stated in subsubsection 4.2.1 the radiological depth is of crucial importance for the dose deposition processes on the particle path. Implementing a LSTM would account for information of prior passed tissues, while avoiding the vanishing gradient problem of basic RNNs. Further information incorporated into the model input, such as a vector mask to indicate the Lorentz force acting on the secondary particles, could increase in a more accurate dose deposition prediction.

In contrast to the DeepDose network of Kontaxis et al. [54] we decreased the voxel size of the target volume in the coronal and sagital plane aswell as included the entire possible gantry angle range into the training data. While increasing the complexity of the proposed problem we have reached comparable results for prostate cancer patient plans. Time needed for prediction of single segments is increased compared to DeepDose from approximately 1 second to 3 seconds. his is due to the increased memory usage by a higher resolution grid.

Uncertainty quantification is of crucial importance to applications of machine learning and deep learning in the medical field, especially when dealing with sensitive decisions, such as the creation of a treatment plan for a cancer patient. Medical physicians aswell as the oncologist in charge need to be sure the dose distributions yielded by the software tool need to be accurate and reliable. Therefore should future implementations of the network include the option for uncertainty quantification. As proposed by Kendall et al. aleatoric and epistemic uncertainty could be included when calculating dose distributions. If the network yields a high uncertainty the dose caclulation system would resort to a MC based calculation algorithm instead of the DL solution for single segments or entire plans. Further information about which region the network is uncertain about coulb be used to decide if the uncertainty is of importance for the treatment plan. This could be the case when the network is uncertrain about the deposited dose in the treatment couch or regions where no OAR are present. The planning software could then decide that the uncertaint reported by the network is of minor imprtance and the prediction could still be used.

7 Conclusion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

8 Aknowledgement

Literature

[1] Freddie Bray, Mathieu Laversanne, and Elisabete Weiderpass: *The ever-increasing importance of cancer as a leading cause of premature death worldwide.* In: Cancer 127.16 (Aug. 15, 2021), 3029–3030. ISSN: 0008-543X, 1097-0142. DOI: 10.1002/cncr.33587. URL: https://onlinelibrary.wiley.com/doi/10.1002/cncr.33587 (visited on 11/17/2021).

[2] Zaigham Abbas and Sakina Rehman: *An Overview of Cancer Treatment Modalities.* In: *Neoplasm.* Ed. by Hafiz Naveed Shahzad. InTech, Sept. 19, 2018. ISBN: 978-1-78923-777-1 978-1-78923-778-8. DOI: 10.5772/intechopen.76558. URL: http://www.intechopen.com/books/neoplasm/an-overview-of-cancer-treatment-modalities (visited on 11/17/2021).

[3] Hans Geinitz, Frank B. Zimmermann, Reinhard Thamm, Andreas Schumertl, Raymonde Busch, and Michael Molls: *3D conformal radiation therapy for prostate cancer in elderly patients.* In: Radiotherapy and Oncology 76.1 (July 2005), 27–34. ISSN: 01678140. DOI: 10.1016/j.radonc.2005.06.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167814005002082 (visited on 08/16/2021).

[4] Tan Dat Nguyen et al.: *The curative role of radiotherapy in adenocarcinoma of the prostate in patients under 55 years of age: A rare cancer network retrospective study.* In: Radiotherapy and Oncology 77.3 (Dec. 2005), 286–289. ISSN: 01678140. DOI: 10.1016/j.radonc.2005.10.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167814005004779 (visited on 08/16/2021).

[5] Tom Budiharto, Karin Haustermans, and Gyoergy Kovacs: *External Beam Radiotherapy for Prostate Cancer.* In: Journal of Endourology (), 10.

[6] Joseph Ragaz et al.: *Adjuvant Radiotherapy and Chemotherapy in Node-Positive Premenopausal Women with Breast Cancer.* In: The New England Journal of Medicine (1997), 7.

[7] Mario De Lena, Roberto Zucali, Giovanni Viganotti, Pinuccia Valagussa, and Gianni Bonadonna: *Combined Chemotherapy-Radiotherapy Approach in Locally Advanced (T3b-T4)Breast Cancer.* In: (), 7.

[8] Carolyn Taylor et al.: *Estimating the Risks of Breast Cancer Radiotherapy: Evidence From Modern Radiation Doses to the Lungs and Heart and From Previous Randomized Trials.* In: Journal of Clinical Oncology 35.15 (May 20, 2017), 1641–1649. ISSN: 0732-183X. DOI: 10.1200/JCO.2016.72.0722. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548226/ (visited on 08/16/2021).

[9]   N. R. Datta, A. K. Bose, H. K. Kapoor, and S. Gupta: *Head and neck cancers: Results of thermoradiotherapy versus radiotherapy.* In: International Journal of Hyperthermia 6.3 (Jan. 1990), 479–486. ISSN: 0265-6736, 1464-5157. DOI: 10.3109/02656739009140944. URL: http://www.tandfonline.com/doi/full/10.3109/02656739009140944 (visited on 08/16/2021).

[10]   S.A. Bhide and C.M. Nutting: *Advances in radiotherapy for head and neck cancer.* In: Oral Oncology 46.6 (June 2010), 439–441. ISSN: 13688375. DOI: 10.1016/j.oraloncology.2010.03.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S1368837510000941 (visited on 08/16/2021).

[11]   Pierre Castadot, John A. Lee, Xavier Geets, and Vincent Grégoire: *Adaptive Radiotherapy of Head and Neck Cancer.* In: Seminars in Radiation Oncology 20.2 (Apr. 2010), 84–93. ISSN: 10534296. DOI: 10.1016/j.semradonc.2009.11.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1053429609000769 (visited on 08/16/2021).

[12]   Howard E. Morgan and David J. Sher: *Adaptive radiotherapy for head and neck cancer.* In: Cancers of the Head & Neck 5.1 (Dec. 2020), 1. ISSN: 2059-7347. DOI: 10.1186/s41199-019-0046-z. URL: https://cancersheadneck.biomedcentral.com/articles/10.1186/s41199-019-0046-z (visited on 08/16/2021).

[13]   Morten Høyer, Anand Swaminath, Sean Bydder, Michael Lock, Alejandra Méndez Romero, Brian Kavanagh, Karyn A. Goodman, Paul Okunieff, and Laura A. Dawson: *Radiotherapy for Liver Metastases: A Review of Evidence.* In: International Journal of Radiation Oncology*Biology*Physics 82.3 (Mar. 1, 2012), 1047–1057. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2011.07.020. URL: https://www.sciencedirect.com/science/article/pii/S0360301611030902 (visited on 08/16/2021).

[14]   Jörn Wulf, Ulrich Hädinger, Ulrich Oppitz, Wibke Thiele, Rea Ness-Dourdoumas, and Michael Flentje: *Stereotactic Radiotherapy of Targets in the Lung and Liver:* in: Strahlentherapie und Onkologie 177.12 (Dec. 2001), 645–655. ISSN: 0179-7158. DOI: 10.1007/PL00002379. URL: http://link.springer.com/10.1007/PL00002379 (visited on 08/16/2021).

[15]   Joern Wulf, Matthias Guckenberger, Ulrich Haedinger, Ulrich Oppitz, Gerd Mueller, Kurt Baier, and Michael Flentje: *Stereotactic radiotherapy of primary liver cancer and hepatic metastases.* In: Acta Oncologica 45.7 (Jan. 2006), 838–847. ISSN: 0284-186X, 1651-226X. DOI: 10.1080/02841860600904821. URL: http://www.tandfonline.com/doi/full/10.1080/02841860600904821 (visited on 08/16/2021).

[16] Florian Sterzing, Thomas B. Brunner, Iris Ernst, Wolfgang W. Baus, Burkhard Greve, Klaus Herfarth, and Matthias Guckenberger: *Stereotactic body radiotherapy for liver tumors: Principles and practical guidelines of the DEGRO Working Group on Stereotactic Radiotherapy*. In: Strahlentherapie und Onkologie 190.10 (Oct. 2014), 872–881. ISSN: 0179-7158, 1439-099X. DOI: 10.1007/s00066-014-0714-1. URL: http://link.springer.com/10.1007/s00066-014-0714-1 (visited on 08/16/2021).

[17] Jacob S Witt, Stephen A Rosenberg, and Michael F Bassetti: *MRI-guided adaptive radiotherapy for liver tumours: visualising the future*. In: The Lancet Oncology 21.2 (Feb. 1, 2020), e74–e82. ISSN: 1470-2045. DOI: 10.1016/S1470-2045(20)30034-6. URL: https://www.sciencedirect.com/science/article/pii/S1470204520300346 (visited on 08/16/2021).

[18] Breast Cancer Expert Panel of the German Society of Radiation Oncology (DEGRO) et al.: *DEGRO practical guidelines for radiotherapy of breast cancer IV: Radiotherapy following mastectomy for invasive breast cancer*. In: Strahlentherapie und Onkologie 190.8 (Aug. 2014), 705–714. ISSN: 0179-7158, 1439-099X. DOI: 10.1007/s00066-014-0687-0. URL: http://link.springer.com/10.1007/s00066-014-0687-0 (visited on 08/16/2021).

[19] Haruo Matsushita, Keiichi Jingu, Rei Umezawa, Takaya Yamamoto, Yojiro Ishikawa, Noriyoshi Takahashi, Yu Katagiri, and Noriyuki Kadoya: *Stereotactic Radiotherapy for Oligometastases in Lymph Nodes—A Review*. In: Technology in Cancer Research & Treatment 17 (Jan. 2018), 153303381880359. ISSN: 1533-0346, 1533-0338. DOI: 10.1177/1533033818803597. URL: http://journals.sagepub.com/doi/10.1177/1533033818803597 (visited on 08/16/2021).

[20] John L. Mikell et al.: *Postoperative Radiotherapy is Associated with Better Survival in Non–Small Cell Lung Cancer with Involved N2 Lymph Nodes: Results of an Analysis of the National Cancer Data Base*. In: Journal of Thoracic Oncology 10.3 (Mar. 2015), 462–471. ISSN: 15560864. DOI: 10.1097/JTO.0000000000000411. URL: https://linkinghub.elsevier.com/retrieve/pii/S1556086415316543 (visited on 08/16/2021).

[21] Dan Lundstedt, Magnus Gustafsson, Gunnar Steineck, David Alsadius, Agnetha Sundberg, Ulrica Wilderäng, Erik Holmberg, Karl-Axel Johansson, and Per Karlsson: *Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients*. In: Radiotherapy and Oncology 103.2 (May 2012), 155–160. ISSN: 01678140. DOI: 10.1016/j.radonc.2011.12.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167814011007572 (visited on 08/16/2021).

[22] Barbara Alicja Jereczek-Fossa, Sara Ronchi, and Roberto Orecchia: *Is Stereotactic Body Radiotherapy (SBRT) in lymph node oligometastatic patients feasible and effective?* In: Reports of Practical Oncology and Radiotherapy 20.6 (2015), 472–483. ISSN: 1507-1367. DOI: 10.1016/j.rpor.2014.10.004. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4661354/ (visited on 08/16/2021).

[23] Byungchul Cho: *Intensity-modulated radiation therapy: a review with a physics perspective.* In: Radiation Oncology Journal 36.1 (Mar. 31, 2018), 1–10. ISSN: 2234-1900, 2234-3156. DOI: 10.3857/roj.2018.00122. URL: http://e-roj.org/journal/view.php?doi=10.3857/roj.2018.00122 (visited on 11/17/2021).

[24] Xiao Han: *MR-based synthetic CT generation using a deep convolutional neural network method.* In: Medical Physics 44.4 (Apr. 2017), 1408–1419. ISSN: 00942405. DOI: 10.1002/mp.12155. URL: https://onlinelibrary.wiley.com/doi/10.1002/mp.12155 (visited on 11/26/2021).

[25] Jelmer M. Wolterink, Anna M. Dinkla, Mark H. F. Savenije, Peter R. Seevinck, Cornelis A. T. van den Berg, and Ivana Išgum: *Deep MR to CT Synthesis Using Unpaired Data.* In: *Simulation and Synthesis in Medical Imaging.* Ed. by Sotirios A. Tsaftaris, Ali Gooya, Alejandro F. Frangi, and Jerry L. Prince. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, 14–23. ISBN: 978-3-319-68127-6. DOI: 10.1007/978-3-319-68127-6_2.

[26] Anna M. Dinkla, Jelmer M. Wolterink, Matteo Maspero, Mark H.F. Savenije, Joost J.C. Verhoeff, Enrica Seravalli, Ivana Išgum, Peter R. Seevinck, and Cornelis A.T. van den Berg: *MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network.* In: International Journal of Radiation Oncology*Biology*Physics 102.4 (Nov. 2018), 801–812. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2018.05.058. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360301618309106 (visited on 08/23/2021).

[27] Samaneh Kazemifar, Anjali Balagopal, Dan Nguyen, Sarah McGuire, Raquibul Hannan, Steve Jiang, and Amir Owrangi: *Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning.* In: Biomedical Physics & Engineering Express 4.5 (July 23, 2018), 055003. ISSN: 2057-1976. DOI: 10.1088/2057-1976/aad100. URL: https://iopscience.iop.org/article/10.1088/2057-1976/aad100 (visited on 08/23/2021).

[28] Shujun Liang, Fan Tang, Xia Huang, Kaifan Yang, Tao Zhong, Runyue Hu, Shangqing Liu, Xinrui Yuan, and Yu Zhang: *Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning.* In: European Radiology 29.4 (Apr. 2019), 1961–

1967. ISSN: 0938-7994, 1432-1084. DOI: 10.1007/s00330-018-5748-9. URL: http://link.springer.com/10.1007/s00330-018-5748-9 (visited on 08/23/2021).

[29] Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan, eds.: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II*. Vol. 11765. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-32244-1 978-3-030-32245-8. DOI: 10.1007/978-3-030-32245-8. URL: https://link.springer.com/10.1007/978-3-030-32245-8 (visited on 08/23/2021).

[30] Jiawei Fan, Jiazhou Wang, Zhi Chen, Chaosu Hu, Zhen Zhang, and Weigang Hu: *Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique.* In: Medical Physics 46.1 (Jan. 2019), 370–381. ISSN: 00942405. DOI: 10.1002/mp.13271. URL: http://doi.wiley.com/10.1002/mp.13271 (visited on 03/16/2021).

[31] Zhiqiang Liu, Jiawei Fan, Minghui Li, Hui Yan, Zhihui Hu, Peng Huang, Yuan Tian, Junjie Miao, and Jianrong Dai: *A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy.* In: Medical Physics 46.5 (May 2019), 1972–1983. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13490. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13490 (visited on 03/16/2021).

[32] Umair Javaid, Kevin Souris, Damien Dasnoy, Sheng Huang, and John A. Lee: *Mitigating inherent noise in Monte Carlo dose distributions using dilated U-Net.* In: Medical Physics 46.12 (Dec. 2019), 5790–5798. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13856. URL: https://onlinelibrary.wiley.com/doi/10.1002/mp.13856 (visited on 08/24/2021).

[33] Keyvan Jabbari: *Review of fast Monte Carlo codes for dose calculation in radiation therapy treatment planning.* In: Journal of Medical Signals & Sensors 1.1 (2011), 73. ISSN: 2228-7477. DOI: 10.4103/2228-7477.83522. URL: http://www.jmssjournal.net/text.asp?2011/1/1/73/83522 (visited on 11/18/2021).

[34] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun: *Dermatologist-level classification of skin cancer with deep neural networks.* In: Nature 542.7639 (Feb. 2, 2017), 115–118. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature21056. URL: http://www.nature.com/articles/nature21056 (visited on 11/18/2021).

[35] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen: *Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans*. In: Scientific Reports 6.1 (Apr. 2016), 24454. ISSN: 2045-2322. DOI: 10.1038/srep24454. URL: http://www.nature.com/articles/srep24454 (visited on 11/18/2021).

[36] Mark Cicero, Alexander Bilbily, Errol Colak, Tim Dowdell, Bruce Gray, Kuhan Perampaladas, and Joseph Barfett: *Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs*. In: Investigative Radiology 52.5 (May 2017), 281–287. ISSN: 1536-0210, 0020-9996. DOI: 10.1097/RLI.0000000000000341. URL: https://journals.lww.com/00004424-201705000-00004 (visited on 11/18/2021).

[37] Osamu Iizuka, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and Masayuki Tsuneki: *Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours*. In: Scientific Reports 10.1 (Dec. 2020), 1504. ISSN: 2045-2322. DOI: 10.1038/s41598-020-58467-9. URL: http://www.nature.com/articles/s41598-020-58467-9 (visited on 11/18/2021).

[38] Ryan Poplin, Avinash V. Varadarajan, Katy Blumer, Yun Liu, Michael V. McConnell, Greg S. Corrado, Lily Peng, and Dale R. Webster: *Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning*. In: Nature Biomedical Engineering 2.3 (Mar. 2018), 158–164. ISSN: 2157-846X. DOI: 10.1038/s41551-018-0195-0. URL: http://www.nature.com/articles/s41551-018-0195-0 (visited on 11/18/2021).

[39] Geoff Delaney, Susannah Jacob, Carolyn Featherstone, and Michael Barton: *The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines*. In: Cancer 104.6 (Sept. 15, 2005), 1129–1137. ISSN: 0008-543X, 1097-0142. DOI: 10.1002/cncr.21324. URL: https://onlinelibrary.wiley.com/doi/10.1002/cncr.21324 (visited on 11/18/2021).

[40] Bernd Kaina: *DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling*. In: Biochemical Pharmacology 66.8 (Oct. 15, 2003), 1547–1554. ISSN: 0006-2952. DOI: 10.1016/s0006-2952(03)00510-0.

[41] Lan-ya Li, Yi-di Guan, Xi-sha Chen, Jin-ming Yang, and Yan Cheng: *DNA Repair Pathways in Cancer Therapy and Resistance*. In: Frontiers in Pharmacology 11 (Feb. 8, 2021), 629266. ISSN: 1663-9812. DOI: 10.3389/fphar.2020.629266. URL:

`https://www.frontiersin.org/articles/10.3389/fphar.2020.629266/full` (visited on 11/18/2021).

[42] H. Rodney Withers and Lester J. Peters, eds.: *Innovations in Radiation Oncology*. Medical Radiology. Berlin, Heidelberg: Springer Berlin Heidelberg, 1988. ISBN: 978-3-642-83103-4 978-3-642-83101-0. DOI: `10.1007/978-3-642-83101-0`. URL: `http://link.springer.com/10.1007/978-3-642-83101-0` (visited on 12/03/2021).

[43] G.W Barendsen: *Dose fractionation, dose rate and iso-effect relationships for normal tissue responses*. In: International Journal of Radiation Oncology*Biology*Physics 8.11 (Nov. 1982), 1981–1997. ISSN: 03603016. DOI: `10.1016/0360-3016(82)90459-X`. URL: `https://linkinghub.elsevier.com/retrieve/pii/036030168290459X` (visited on 12/03/2021).

[44] H. A. Lorentz: *Versuch Einer Theorie der Electrischen und Optischen Erscheinungen in Bewegten Körpern*. In: *Collected Papers: Volume V*. Ed. by H. A. Lorentz. Dordrecht: Springer Netherlands, 1937, 1–138. ISBN: 978-94-015-3445-1. DOI: `10.1007/978-94-015-3445-1_1`. URL: `https://doi.org/10.1007/978-94-015-3445-1_1` (visited on 11/19/2021).

[45] Hannah J Lee, Gye Won Choi, Mamdooh Alqathami, Mo Kadbi, and Geoffrey Ibbott: *Using 3D dosimetry to quantify the Electron Return Effect (ERE) for MR-image-guided radiation therapy (MR-IGRT) applications*. In: Journal of Physics: Conference Series 847 (May 2017), 012057. ISSN: 1742-6588, 1742-6596. DOI: `10.1088/1742-6596/847/1/012057`. URL: `https://iopscience.iop.org/article/10.1088/1742-6596/847/1/012057` (visited on 11/22/2021).

[46] Anders Ahnesjö: *Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media: Photon dose calculation*. In: Medical Physics 16.4 (July 1989), 577–592. ISSN: 00942405. DOI: `10.1118/1.596360`. URL: `http://doi.wiley.com/10.1118/1.596360` (visited on 06/22/2021).

[47] R. Mohan, C. Chui, and L. Lidofsky: *Differential pencil beam dose computation model for photons: Differential pencil beam computation for photons*. In: Medical Physics 13.1 (Jan. 1986), 64–73. ISSN: 00942405. DOI: `10.1118/1.595924`. URL: `http://doi.wiley.com/10.1118/1.595924` (visited on 06/22/2021).

[48] *nrc-cnrc/EGSnrc*. original-date: 2012-11-28T19:26:41Z. Nov. 2, 2021. URL: `https://github.com/nrc-cnrc/EGSnrc` (visited on 11/04/2021).

[49] M. Friedel, M. Nachbar, D. Mönnich, O. Dohm, and D. Thorwarth: *Development and validation of a 1.5 T MR-Linac full accelerator head and cryostat model for Monte Carlo dose simulations*. In: Medical Physics 46.11 (Nov. 2019), 5304–5313. ISSN:

0094-2405, 2473-4209. DOI: `10.1002/mp.13829`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13829` (visited on 03/16/2021).

[50] Ahmad Neishabouri, Niklas Wahl, Andrea Mairani, Ullrich Köthe, and Mark Bangert: *Long short-term memory networks for proton dose calculation in highly heterogeneous tissues.* In: Medical Physics (Mar. 11, 2021), mp.14658. ISSN: 0094-2405, 2473-4209. DOI: `10.1002/mp.14658`. URL: `https://onlinelibrary.wiley.com/doi/10.1002/mp.14658` (visited on 03/16/2021).

[51] Sonia Martinot, Norbert Bus, Maria Vakalopoulou, Charlotte Robert, Eric Deutsch, and Nikos Paragios: *High-Particle Simulation of Monte-Carlo Dose Distribution with 3D ConvLSTMs.* In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021.* Ed. by Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert. Vol. 12904. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, 499–508. ISBN: 978-3-030-87201-4 978-3-030-87202-1. DOI: `10.1007/978-3-030-87202-1_48`. URL: `https://link.springer.com/10.1007/978-3-030-87202-1_48` (visited on 11/20/2021).

[52] Ryan Neph, Yangsibo Huang, Youming Yang, and Ke Sheng: *DeepMCDose: A Deep Learning Method for Efficient Monte Carlo Beamlet Dose Calculation by Predictive Denoising in MR-Guided Radiotherapy.* In: arXiv:1908.04437 [physics] (Aug. 12, 2019). arXiv: `1908.04437`. URL: `http://arxiv.org/abs/1908.04437` (visited on 08/02/2021).

[53] Jiawei Fan, Lei Xing, Peng Dong, Jiazhou Wang, Weigang Hu, and Yong Yang: *Data-driven dose calculation algorithm based on deep U-Net.* In: Physics in medicine and biology 65.24 (Dec. 22, 2020), 245035. ISSN: 0031-9155. DOI: `10.1088/1361-6560/abca05`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7870566/` (visited on 11/20/2021).

[54] C. Kontaxis, G. H. Bol, J. J. W. Lagendijk, and B. W. Raaymakers: *DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning.* In: Physics in Medicine & Biology 65.7 (Apr. 2020). Publisher: IOP Publishing, 075013. ISSN: 0031-9155. DOI: `10.1088/1361-6560/ab7630`. URL: `https://doi.org/10.1088/1361-6560/ab7630` (visited on 08/23/2021).

[55] Olaf Ronneberger, Philipp Fischer, and Thomas Brox: *U-Net: Convolutional Networks for Biomedical Image Segmentation.* In: arXiv:1505.04597 [cs] (May 18, 2015). arXiv: `1505.04597`. URL: `http://arxiv.org/abs/1505.04597` (visited on 11/03/2021).

[56]   Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin: *TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning.* In: Computer Methods and Programs in Biomedicine 208 (Sept. 2021), 106236. ISSN: 01692607. DOI: `10.1016/j.cmpb.2021.106236`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0169260721003102` (visited on 11/03/2021).

[57]   Daniel A. Low, William B. Harms, Sasa Mutic, and James A. Purdy: *A technique for the quantitative evaluation of dose distributions.* In: Medical Physics 25.5 (May 1998), 656–661. ISSN: 00942405. DOI: `10.1118/1.598248`. URL: `http://doi.wiley.com/10.1118/1.598248` (visited on 03/16/2021).

[58]   Ahmad Neishabouri, Niklas Wahl, Ulrich Köthe, and Mark Bangert: *Long short-term memory networks for proton dose calculation in highly heterogeneous tissues.* In: arXiv:2006.06085 [physics] (June 10, 2020). arXiv: `2006.06085`. URL: `http://arxiv.org/abs/2006.06085` (visited on 03/16/2021).

[59]   Alex Kendall and Yarin Gal: *What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?* In: arXiv:1703.04977 [cs] (Oct. 5, 2017). arXiv: `1703.04977`. URL: `http://arxiv.org/abs/1703.04977` (visited on 07/23/2021).

Appendix