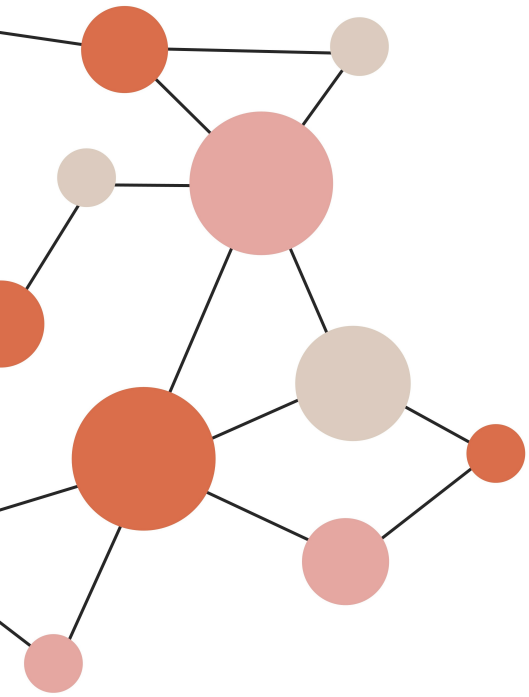
Multi-Task Learning for Optimal Dose and Contour Prediction in Adaptive Proton Therapy

Thomas Landman



Master Thesis
to obtain the degree of Master of Science

Master Applied Physics
Delft University of Technology

June 2022

This report concludes a master's thesis project performed as part of the MSc Applied Physics at TU Delft. The project took place from September 2021 to May 2022 and was a collaboration between the following research groups:

Medical Physics & Technology section, Radiation Science & Technology department, TU Delft
Division of Image Processing, Department of Radiology, LUMC

AUTHOR:

Thomas Landman

Student Number:	4587111
University:	TU Delft
Study Programme:	MSc Applied Physics
Track:	Physics for Health and Life

SUPERVISORS:

Dr. Zoltán Perkó

Institution:	TU Delft
Faculty:	Applied Sciences
Department:	Radiation Science & Technology

Dr. ir. Marius Staring

Institution:	LUMC
Department:	Research Radiology

Drs. Mohamed Elmahdy

Institution:	LUMC
Department:	Research Radiology

Abstract

Adaptive proton therapy (APT) removes one of the most significant sources of inaccuracy in treatment delivery, which is using a treatment plan based on an outdated patient anatomy. Adapting the plan throughout the treatment is crucial for delivering an optimal dose to the patient, whose anatomy is constantly changing. This is especially true for proton therapy, where the delivered dose is highly dependent on the range accuracy. Imaging and plan adaptation must be performed online, immediately before the dose delivery, to take maximum advantage of the benefits of APT. The main problem with online APT is that adaptation of the treatment plan takes too long. Therefore, automation of the processes is required to ensure they can be executed adequately in a short time frame.

Deep learning methods have been successfully applied in two processes required for adaptation, namely the definition of structure contours on a CT scan and determining an optimal dose distribution for a given anatomy. Since a treatment plan is dependent on the locations of the different structures, dose prediction methods rely on manually defined contours, which are not available for daily CT scans in APT due to time limitations. This research aims to develop an approach that determines an optimal dose distribution for prostate cancer patients without using manual structure contours.

We use 3D U-Nets for image segmentation and registration as methods for defining the contours on an image. We use another 3D U-Net to predict an optimal dose distribution, which can use predicted or manually defined contours as input. In addition to this, we use two multi-task learning approaches that allow one network to perform both contour definition and dose prediction, which makes it possible to share information between the tasks. The first approach is a cross-stitch network that allows two networks to share feature maps if this is beneficial and the second approach is a w-net that consecutively performs contour definition and dose prediction, using the predicted contours for the dose prediction.

The manual contour based dose prediction performed well in the area around the structures, resulting in a test set average 2%/2mm gamma pass rate of $93.4\% \pm 3.2\%$ and a D_{mean} prediction error of $0.45\% \pm 0.36\%$ in the prostate. The average errors for predicting measures such as D_{95} and $V_{95\%}$ in the targets range from 1% to 3%.

The best method for predicting optimal dose distributions without manual contours is to first predict the contours on the CT scan and use those contours for the dose prediction. However, dose predictions based on predicted contours are significantly worse than those based on manual contours, having a 2%/2mm gamma pass rate of $83.8\% \pm 6.9\%$ and a D_{mean} prediction error of $0.92\% \pm 0.7\%$ in the prostate. Their average errors for predicting measures such as D_{95} and $V_{95\%}$ range from 7% to 20%, which makes these predicted dose distributions too inaccurate to be helpful for treatment planning. This shows that dose prediction relies heavily on accurate knowledge of the structure locations, considering the predicted contours have similar quality as those from state-of-the-art methods.

Dose predictions have not improved by additionally learning a network the contour definition task. Using feature maps from other networks via cross-stitch units had no advantageous effect on the predicted dose distributions, mainly because dose predictions not based on structure masks were too bad for it to have any effect. The dose predictions from the w-net did not improve after the segmentation and dose prediction networks were trained together, which could be because the dose prediction loss could not improve the segmentation sufficiently. The main conclusion is that multi-task learning can only benefit related tasks if they can already be performed independently to a certain extent. It is not a substitute for missing information required to perform the task.

Contents

List of Figures

List of Tables

List of Acronyms

APT	Adaptive Proton Therapy
CNN	Convolutional Neural Network
CT	Computed Tomography
DSC	Dice Similarity Coefficient
DVF	Deformation Vector Field
DVH	Dose-Volume Histogram
GPR	Gamma Pass Rate
HD	Hausdorff Distance
IMPT	Intensity Modulated Proton Therapy
MSD	Mean Surface Distance
NCC	Normalized Cross-Correlation
OAR	Organ At Risk
PTV	Planning Target Volume
RBE	Relative Biological Effectiveness
ReLU	Rectified Linear Unit
WMSE	Weighted Mean Squared Error

1 Introduction

Radiotherapy is one of the main procedures used to treat cancer patients. Its primary objective is to deliver a sufficient dose to the tumor while minimizing the dose absorbed by important organs. Different radiotherapy methods, such as external radiotherapy, brachytherapy or radionuclide therapy, can be used to deliver this dose to the patient. Proton therapy, a relatively new type of external radiotherapy, is considered because it has a distinct advantage over other types of external radiotherapy. This advantage is that most of the dose of proton beams is absorbed at a certain depth called the Bragg peak, while the dose of photon beams is absorbed more uniformly throughout the body. This property of proton therapy allows it to deposit the dose locally at the target tissue while minimizing the dose absorbed by surrounding healthy tissue.

A drawback to this precision is that it is highly dependent on the precision of the locations of the Bragg peaks. A change in density somewhere along a proton beam path has a drastic effect on the dose distribution of that beam. This means that any anatomical difference between the treated patient and the image used for treatment planning can have a devastating effect on the resulting dose distribution. Since treatment delivery happens in multiple fractions that take place on different days, inter-fractional variations, such as tumor shrinkage or bladder filling changes, reduce the effectiveness of the treatment plan. Some measures, such as adding target margins or incorporating robust optimization, try to ensure the treatment plan delivers a sufficient dose to the target despite these uncertainties. However, since these measures also increase the dose to surrounding organs [1], a better method for dealing with these inter-fractional differences is necessary.

Adaptive Proton Therapy (APT) tries to improve treatment planning by imaging the patient multiple times throughout the treatment and adapting the plan accordingly. Adaptation requires imaging, contour definition, plan assessment, re-planning and plan verification. Because this process can take hours to days, adaptation usually happens between treatment fractions (offline) [2]. Although this can account for slower inter-fractional changes such as tumor shrinkage and weight loss, it fails to account for faster, day-to-day changes like organ fillings. Therefore, to take maximum advantage of the benefits of APT, imaging and adaptation must be performed immediately before the fraction (online). Some recent studies have shown that an online APT workflow is possible [3], but automation of the processes is still required to ensure they can be executed adequately in a short time frame.

In recent years, deep learning methods have been successfully applied to many processes in the radiotherapy workflow. Advancements in computational power allow the application of these methods to large 3D images, something previously not possible. Their ability to identify features from large sets of data and the speed with which they can apply this knowledge to new images makes deep learning methods very useful. Convolution neural networks (CNNs), for example, are an accurate and fast method for defining the contours of important structures in a computed tomography (CT) scan [4]. A CNN can define the contours required for prostate cancer treatment planning in a few seconds, while it takes a skilled oncologist more than an hour to do this manually [5]. In the APT workflow, this method could be used to quickly determine the contours of the structures on the daily CT scan that is made just before the treatment delivery.

Predicting optimal dose distributions is another area where CNNs are successfully applied. Given the patient's anatomy, the network can determine the dose distribution of an optimal treatment plan. Knowing this dose distribution can guide a dosimetrist in optimizing the treatment plan, saving time and potentially ensuring higher-quality plans. Knowing what an optimal dose distribution looks like for a daily anatomy can have multiple uses in APT treatment planning. For example, it could help determine if re-planning is necessary, it could speed up the process of re-planning and it could serve as quality assurance for a newly created treatment plan. The problem with determining the optimal dose distribution for a daily CT scan in APT is that most dose prediction networks described in literature use the contours of the structures in the CT scan for the prediction. Since it takes too much time to determine the contours manually for a daily CT scan, these contours are usually not available.

1.1 Goals of the Research

This research aims to determine an approach that can predict the dose distribution of an optimal treatment plan for a daily CT scan in APT. The challenging thing is that this must be done without using manual contours. Therefore, incorporating the knowledge of the structure locations into the dose prediction network without knowing the exact contours will be vital to answering the first research question of this study.

Research Question 1

In adaptive proton therapy, what is the best approach for predicting an optimal dose distribution for a daily CT scan?

A potentially interesting way of incorporating the structure information into the network is by letting one network perform both the contour prediction task and the dose prediction task. This learning of multiple tasks is called multi-task learning and it could lead to improved competence in both tasks. The idea is that the network can exploit commonalities and differences across tasks, allowing it to extract more information from the available data. Recently it has been shown that joining a segmentation and registration network in a single joint architecture via so-called cross-stitch units results in superior contour prediction compared to the single-task networks [6]. Similarly, the dose prediction task could be learnt together with the contour definition task. The second goal of this research is to investigate the use of multi-task learning for combined dose prediction and contour prediction.

Research Question 2

Is it beneficial to perform the contour prediction and dose prediction task together in a multi-task network?

1.2 Structure of the Report

In Chapter 2, the theory behind the concepts and methods described in this introduction are discussed in more detail. First, a detailed explanation is given about the APT treatment planning process, and different methods for contour definition and optimal dose prediction are discussed. After that, a detailed explanation of deep learning is given, covering the fundamentals of deep learning, convolutional neural networks and multi-task learning. Chapter 3 contains details about the CT scans and treatment plans used in this research. It also gives specific information about the networks developed for contour definition and the different approaches used for optimal dose prediction. Chapter 4 discusses the quality of the created treatment plans used for training and testing the dose prediction network. It also shows results for the predicted contours and discusses their quality. Finally, the predicted dose distributions of the different approaches are compared, followed by a discussion of these results. To conclude the report, our findings are summarized in Chapter 5.

2 Theory

This chapter will describe the concepts underlying this research and discuss the literature. Section 2.1 will explain everything from the basics of radiotherapy up to adaptive proton therapy. Section 2.2 and Section 2.3 give details about two things that can speed up the treatment planning process, namely automatic contour definition and optimal dose prediction. Finally, Section 2.4 and Section 2.5 explain the deep learning methods and concepts required to understand the methods we propose in this report.

2.1 Adaptive Proton Therapy

2.1.1 Fundamentals of Radiotherapy

Radiotherapy is a cancer treatment that fights a tumor by using radiation to damage cancer cells. It is one of the main treatment methods for cancer, used in over 50% of all patients, either on itself or in combination with chemotherapy or surgery [7]. Radiation ionizes the DNA molecules inside the cancer cells, which means the cells can not divide and the tissue can not grow. Since cancer cells tend to divide quicker than normal cells, they are usually more sensitive to radiation and die off quicker [8]. This difference in radiosensitivity is amplified by the fractionation of the dose delivery, which is optimized for damaging cancer cells and sparing normal cells [9]. Dose, a definition for the amount of absorbed radiation, is delivered over 35 fractions of 2 Gy, for example, instead of delivering 70 Gy to the patient in one treatment session. Although this helps to spare normal cells to a certain extent, damaging them and thereby risking treatment-related complications is unavoidable.

Because of this risk of complications, the main goal of radiotherapy is to control the tumor while sparing healthy tissue as much as possible. The entire radiotherapy process (Figure 2.1) is designed to achieve this goal [10]. The radiotherapy process starts with making a 3D CT scan of the patient to characterize their anatomy. Then in contour definition, several structures, such as the tumor and important organs, are defined on the CT scan (Section 2.2). Next, a clinician sets the treatment plan objectives, such as the desired dose in the tumor and the maximum dose allowed in important organs, based on the CT scan and the locations of the different structures. Given these objectives, a treatment planning system calculates the optimal treatment plan (Section 2.1.3). Finally, before the treatment is delivered, the quality of the created plan must be assured. This assurance includes a thorough plan review and a measurement-based test of the delivered dose.

Figure 2.1: *An overview of the different steps of the radiotherapy process.*

Delivering the treatment can be done in a variety of ways. The radiation can be delivered from outside the body (external radiotherapy) or from inside the body (internal radiotherapy). Some examples of internal radiotherapy methods are brachytherapy, which implants radiation sources near the target area, and radionuclide therapy, which injects radioactive molecules into the patient's bloodstream designed to target the cancer cells in the tumor. While these methods do little damage to surrounding tissue, they are only applicable in certain cases [11]. External radiotherapy, applied more often, uses a source of radiation outside the body to irradiate the patient. This radiation can be beams of high-energy photons or beams of accelerated ionizing particles such as protons or carbon ions. High energy photon beams are created by a linear accelerator, which is present in many medical centers. Charged particles, however, require a large cyclotron or synchrotron to be accelerated to the required velocity, which is available in far fewer places.

2.1.2 Advantages and Disadvantages of Proton Therapy

Proton therapy can potentially be a superior method to conventional radiotherapy because of the unique depth-dose characteristics of protons. A proton can interact in the body by Coulomb scattering with the electrons or the nuclei of atoms. The energy loss of protons is inversely proportional to the square of their velocity [12]. This leads to a local peak of absorbed radiation in the body, because protons deposit most of their energy right before they come to rest. This local peak is called the Bragg peak (Figure 2.2) and its location depends on the proton energy. This allows proton therapy to concentrate the dose on the tumor while minimizing the effect on surrounding tissue. Proton therapy is especially useful for treating tumors that have not spread and are close to important organs, such as head and neck cancers. Also it is often applied in the treatment of children, since they are particularly susceptible to late adverse effects of radiation [13].

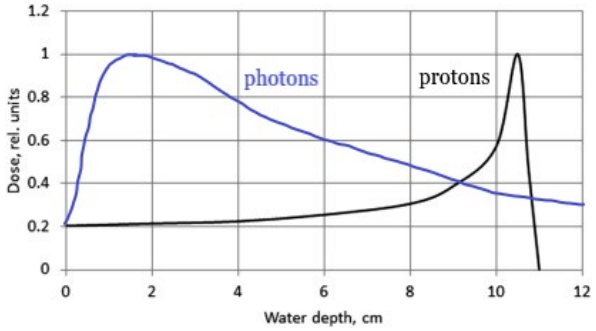


Figure 2.2: *Depth dose distributions for protons and photons.*

Besides these advantages, proton therapy also has some disadvantages. The main problem with proton therapy is that the position of the Bragg peak is strongly dependent on the density of the material the beam has traveled through. This dependency means that any change in the anatomy of the patient can have a devastating effect on the resulting dose distribution, making proton therapy less resilient than photon therapy. Also, there is still much discussion about the relative biological effectiveness (RBE) of protons. This is usually assumed to be 1.1 for all tissue, but multiple studies have reported lower or higher values for different tissues [14]. Besides these physical disadvantages, there are also some economic problems. Since proton therapy facilities are large and expensive, the cost of treatment is considerably higher than for photon therapy. This while, for most cases, there is no clear evidence that proton therapy is a significant improvement over photon therapy. Because of the lack of clinical evidence indicating its superiority over photon therapy and its vulnerability to uncertainties, more research still needs to be done on proton therapy [14].

2.1.3 How an Optimal Treatment Plan is Created

Proton therapy treatment planning aims to find the machine parameters, such as beam position and beam intensity, that result in a treatment that gives the patient the highest chance of recovery without any complications. To find this optimal treatment plan, we can determine multiple criteria that such a plan should meet. The planning target volume (PTV), which is the volume we want to irradiate containing the tumor, should obtain enough dose to ensure the tumor is controlled. Organs at risk (OARs), which are the organs causing most complications when overdosed, should receive as low a dose as possible. Some of these criteria are fixed in guidelines, but often a trade-off between opposing criteria has to be made. Which criteria are given priority is patient-specific, but can also differ between clinicians.

Once the objectives for the treatment plan are decided, they are converted to a numerical optimization problem as follows [15]. The patient is discretized in voxels, so the dose in the voxels d can be calculated for a certain beam set-up x. The relation between the beamlets and the voxels is a linear relation: $d(x) = Ax$, where A is the dose influence matrix. The vector of voxel doses that belong to a region i is denoted by d_i. The objective is represented by a cost function $f(d_0)$ and the constraints are represented by the cost functions $[g_1(d_1), \ldots, g_n(d_n)]$. An objective could be that the mean dose of all voxels belonging to the rectum must be as low as possible. A constraint could be that the dose in all voxels belonging to the PTV must be higher than 70 Gy. Equation 2.1 gives an example of an optimization problem.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(d_0) \\ \text{subject to} \quad & g_1(d_1) \geq 70 \text{ Gy}, \\ & g_2(d_2) \leq b_2, \\ & \vdots \\ & g_n(d_n) \leq b_n, \\ & x \geq 0. \end{aligned} \tag{2.1}$$

Computational methods are used to find the beam set-up x that minimizes $f(d_0)$ given the constraints $[g_1(d_1), \ldots, g_n(d_n)]$. The result is a Pareto optimal solution where the objective criterion can not be improved given the constraints on other criteria.

One of the problems in treatment planning is how uncertainties should be taken into account. Especially proton therapy treatment plans are sensitive to range uncertainties, as discussed in Section 2.1.2. Many uncertainties must be taken into account, such as set-up uncertainties, changes in patient anatomy, imaging or delivery uncertainties, and variations in the defined contours. One way to deal with these uncertainties is to increase the PTV by adding larger safety margins around the tumor. Another way is incorporating robust treatment planning. By creating a plan that performs well for a range of possible anatomies similar to the anatomy on which the planning is based, the plan will be less sensitive to changes in the anatomy [14]. Although both these methods reduce the chance of under-dosing the tumor, they do increase the dose to surrounding organs [1]. A method that reduces uncertainties without increasing the dose to surrounding organs that has shown much potential over the past few years is APT. In APT, the patient is imaged throughout the treatment and the treatment plan is adapted to the patient's anatomy.

2.1.4 The Adaptive Proton Therapy Workflow

APT has shown promising results [16, 17, 18] compared to other methods, and an adaptive workflow has already been implemented successfully [3]. Basing a treatment plan on an outdated patient anatomy is one of the most significant sources of range inaccuracies, so updating the treatment plan to the current anatomy should ideally happen as often as possible. This means the patient must be imaged at the start of the appointment, ideally already in the treatment position to minimize set-up inaccuracies. When plan adaptation is required, this must happen on the spot. An example of this adaptive workflow is given in Figure 2.3. The process pre-treatment is the same as in non-adaptive radiotherapy, but the daily treatment delivery process now also includes adapting the treatment plan.

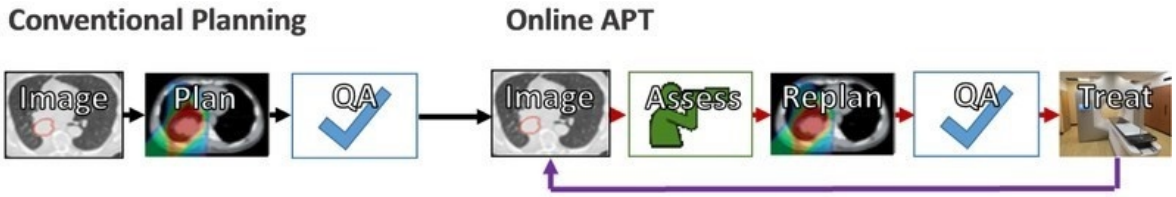


Figure 2.3: *An example of a workflow for the online APT processes. The workflow pre-treatment is the same for conventional planning and online APT, but in online APT extra steps are performed before the treatment delivery. These extra steps must happen in a short time frame. Black arrows represent a time scale of days, while red arrows represent a time scale of minutes. Adapted from [19].*

The bottleneck of this daily process is the time it takes to adapt the treatment plan. Adapting the plan manually is not feasible, since clinicians' time is costly, but more importantly, the treatment appointment would take too long. Making a daily image of the patient requires either an in-room CT or can be done using cone-beam CT. Although the image quality of a cone-beam CT is often not acceptable for dose calculation in proton therapy [1], the planning CT could then be registered to the daily CT, with the added advantage that the planning contours can be propagated to the daily scan. An alternative to propagating the planning contours is defining the contours on the daily CT scan directly using automated methods. However, both these methods (Section 2.2) still require correction or verification by a clinician [2].

When the daily anatomy is characterized, it must be determined if and how the treatment plan must be adapted. Some fast dose calculation methods have been developed, which, although being less accurate than Monte Carlo methods, are able to calculate dose distributions of acceptable accuracy within seconds [20]. This helps to quickly evaluate a treatment plan's performance on a given anatomy. When the adaptation of the plan is required, there are multiple ways to do this. One can chose to re-optimize the original treatment plan, to select a plan from a library of predetermined treatment plans [21] or to do a restoration of the spot positions [22]. The quality of the new treatment plan must be guaranteed, but measurement-based quality assurance is not practical when the patient is kept in the treatment position. However, since most errors in the adaptive radiotherapy process are not those caught during measurement-based quality assurance, an independent dose calculation method may be used instead [23].

2.2 Contour Definition

2.2.1 The Importance of Automated Methods

Correctly defining the contours of the different organs and other structures in the image is an important aspect of radiotherapy since the locations of these structures are essential for treatment planning. Traditionally, contours are defined manually by highly trained radiation oncologists, a process that takes a substantial amount of time. Some of the challenges of contouring correctly are image regions with low contrast or structures from which the edges are ambiguous. For these reasons, among others, there is some inter-observer variation in how certain contours are defined [24]. Because of this inter-observer variation and the long duration of a manual approach, an automated contouring method is necessary. Especially time reduction of the procedure is essential for contour definition in an online adaptive workflow. Automated methods have been studied for a long time, but unfortunately, they still require correction by a clinician in most cases [19]. In recent years, deep learning methods (Section 2.4) for contouring medical images are getting more and more accurate. These fast methods use previously contoured images to learn the correlation between CT scans and the different structures. Using this knowledge, they can either define the contours on an image directly or propagate known contours to the image via registration.

2.2.2 Image Segmentation Using Deep Learning

Image segmentation is a method where all the pixels of an image are classified into predetermined classes (Figure 2.4). This is used for many applications, including computer vision, object detection, recognition tasks or medical imaging. In medical image segmentation, often the goal is to identify the set of voxels that belong to organs or other structures of interest. Performing image segmentation using CNNs (Section 2.5.1) has become a widely used method in recent years and it has been shown to outperform conventional methods [4]. In 2015 Ronneberger et al. published their CNN architecture for image segmentation called u-net [25], from which most state-of-the-art segmentation networks used today are derived. It uses downsampling layers as an encoder and upsampling layers as a decoder, with skip connections between the downsampling and upsampling layers (Section 2.5.2). The encoder extracts features from an image, which the decoder uses to link different image regions to different structures. Image segmentation networks could be helpful in APT since they can be used to identify structures on the daily CT scan that are necessary for treatment planning. This works especially well for structures that have a lot of contrast with the background, such as the bladder.

Figure 2.4: *An example of image segmentation used in computer vision for self-driving cars.*

2.2.3 Image Registration Using Deep Learning

Image registration is the determination of a geometrical transformation that aligns one image of an object (moving image) with another image of that object (fixed image) (Figure 2.5). This is useful when comparing or combining information from both images. It has multiple medical applications such as image guidance, image reconstruction or contour definition. In APT, for example, it can be used to align the planning CT scan onto the daily CT scan, thereby allowing information from both images to be used for treatment planning. Traditional registration algorithms try to find a deformation vector field (DVF) that maps the moving image to the fixed image by an iterative approach. This means that an objective function related to the similarity between the moving image and the fixed image is optimized step by step until the optimal DVF is found. Since these iterations have to be performed for every new image, iterative methods generally take a lot of time. Models based on deep learning, however, only perform this iterative optimization in the training phase and afterward can predict a DVF by a single forward computation.

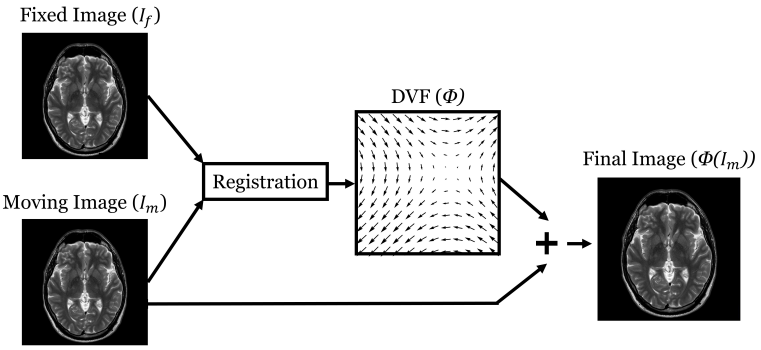


Figure 2.5: *Aligning a moving image with a fixed image by determining a DVF using a registration method.*

Similar to segmentation networks, a common choice for registration networks is a CNN that uses a encoder-decoder architecture (Section 2.5.2). Usually, the moving image and the fixed image are given as input to the network and the transformation parameters are given as an output. The loss of such a network includes a part that penalizes dissimilarity between the transformed moving image and the fixed image and a part that penalizes high gradients in the DVF to ensure a realistic transformation. Although deep learning models for registration have some problems, e.g., lack of training data, difficulty with predicting large deformations and high dimensionality of the output, they show comparable accuracy to iterative registration methods [4].

2.2.4 Contour Propagation Through Image Registration

Many deep learning registration models use supervised training, either by comparing the predicted DVF with a ground truth DVF from a conventional registration method or by comparing reference labels, such as organ contours, on both images. Since mimicking a DVF from another method means the network can not outperform that other method, more commonly, reference labels on both images are used to guide the network. Multiple studies [26, 27, 28] have shown that using anatomical contours to weakly supervise registration networks improves their performance. The contours of the structures on the transformed moving image and the contours of the structures on the fixed image should overlap, so DVFs resulting in a transformation where there is much overlap are favored. Especially when registration is performed to propagate the contours from one image to another, this supervision has been shown to be beneficial [28].

Propagation of contours means transforming the contours of structures on one image, so they overlap with the structures on another image. The transformation parameters obtained by registration of the two images can be used to transform the previous contours so they can be applied to the new image. This is useful in computer vision, for example, when new images are made constantly and information from the previous images could be useful. It could also be applied to adaptive radiotherapy since the contours defined on the planning CT scan can be propagated to the daily CT scan. A recent study which was the inspiration for this research [6], showed that registration of the planning CT scan with the daily CT scan could be used successfully to define contours on the daily CT scan. Instead of only supervising the registration network via the loss, it also showed that combining a segmentation network and a registration network at an architectural level improves the performance of both networks (Section 2.5.3).

2.3 Predicting Optimal Dose Distributions

2.3.1 Knowledge-Based Planning

Optimizing a treatment plan is a time-consuming process that requires manually adjusting objective parameters to find the optimal treatment plan. Even for a skilled and experienced clinician, it takes a lot of time and effort to produce high-quality plans. Another problem, besides that optimization takes a long time, is the variation in treatment plan quality between planners [29]. In an effort to overcome these problems, knowledge-based planning strategies have been proposed [30]. Knowledge-based planning tries to incorporate historic treatment plans and anatomical data to predict what kind of dose-volume histograms (DVHs) and dose distributions are possible. The DVHs and dose distributions can serve as optimization objectives and provide other guidance in the treatment planning process. The aim is to ensure more consistent, high-quality plans and reduce time spent optimizing a plan by trial-and-error. Another potentially practical application is using this knowledge to decide the type of treatment best suited for a particular patient. For example, it could be applied to choose between photon and proton therapy.

Predicting accurate DVHs and dose distributions is useful in APT because the daily optimization of a treatment plan must happen in a short time frame. Knowing the optimal dose distribution for a daily anatomy can help decide if re-planning is necessary. When re-planning is necessary, it can serve as guidance or quality assurance for the new treatment plan. Due to the time limitations in an online APT workflow, the dose distribution must be predicted quickly. Fast deep learning methods that predict the dose distribution based on the patient's anatomy are excellent for this. These deep learning methods can determine geometrical and anatomical features, which they use to predict the resulting dose distribution. Although deep learning methods have only been used in recent years, they have already surpassed traditional methods that use anatomical features handcrafted by researchers and clinicians in terms of accuracy [31].

2.3.2 Deep Learning Methods for Optimal Dose Prediction

The immense progress of deep learning in recent years has substantially impacted dose prediction methods. The development of u-net [25], a deep CNN used for semantic segmentation, was a starting point for much of the research on deep learning methods for dose prediction (see Section 2.5.2). In 2017, two years after the publication of the u-net, Nguyen et al. [32] successfully used this 2D u-net architecture for a contour-based prediction of the optimal dose distribution in prostate cancer patients. It was later found that this approach could be improved by using dosimetric information as input in addition to the structure contours [33]. In 2019, a densely connected 3D u-net was used to predict dose distributions for head and neck cancer patients [34], combining architectures from u-net and DenseNet. This was successful, even though head and neck cancer treatment plans are very complicated due to multiple prescription dose levels and many radiation-sensitive structures. A similar study later confirmed the benefits of using 3D inputs over 2D inputs [35]. Three recent studies on predicting dose distributions for prostate cancer have done this using a 3D u-net. Kontaxis et al. [36] used a 3D u-net with geometrical information about the anatomy as additional input. Another study investigated a general model that could be applied to predict different treatment planning styles [37]. The third study tried to predict the multileaf collimator positions that would lead to an optimal dose distribution in addition to predicting the dose distribution [38]. While most of these studies involved predicting photon dose distributions, a recent study focused on predicting proton and photon dose distributions to assess their clinical usage for the treatment of a particular tumor [39].

2.3.3 Dose-Volume Histograms

A common way of analyzing dose distributions is with a DVH. They are a simple method to visualize a 3D dose distribution in a 2D format. A DVH is generally plotted as a cumulative histogram with dose on the x-axis and relative volume on the y-axis. An example is given in Figure 2.6. The height of a line at a certain dose represents the volume percentage receiving a dose greater than or equal to that dose. Usually, a DVH includes multiple important structures, which are represented by different lines. DVHs can be used to evaluate a plan or compare different plans with each other. In this study, they will be used to compare a predicted dose distribution with the actual dose distribution. Comparisons can be made by looking at specific DVH points. For example, V_x defines the volume percentage that receives x dose or more and can be visualized as the height of a specific line at dose x. D_x is the x^{th} percentile highest dose in a structure and can be visualized as the dose value of a line at volume percentage x. From this definition follows it that D_0 is the same as the maximum dose in a structure.

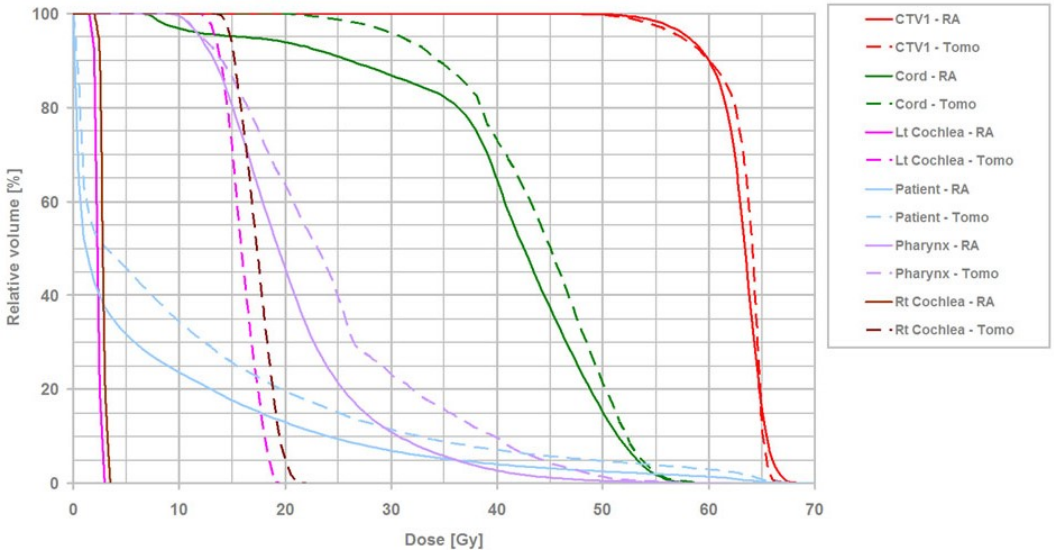


Figure 2.6: *Example of a cumulative DVH comparing Rapid Arc (RA) and Tomotherapy (Tomo) treatment plans [40]. The dose is given on the x-axis and the relative volume on the y-axis. Different lines represent the dose in different structures.*

2.4 Concepts in Deep Learning

2.4.1 An Introduction to Deep Learning

Deep learning is a subset of machine learning where multiple layers are used to learn patterns from input data. The difference between deep learning and other machine learning methods is that deep learning algorithms learn how to represent the data with self-learned features instead of features designed by humans. These features are then used to classify the data or make predictions based on the data. The advantage of this is that deep learning algorithms can process large amounts of unstructured data, such as images or text, without the need for pre-processing. Deep learning concepts have been around since the 1960s, although many applications have only been realized recently due to advancements in computing power. The theory behind deep learning is based on the working of the human brain, which is where the name neural network comes from. The different layers in the network pass information, similar to how layers of neurons are activated in the brain. An example of how different layers in a neural network convert input data to output data using fully connected layers is given in Figure 2.7.

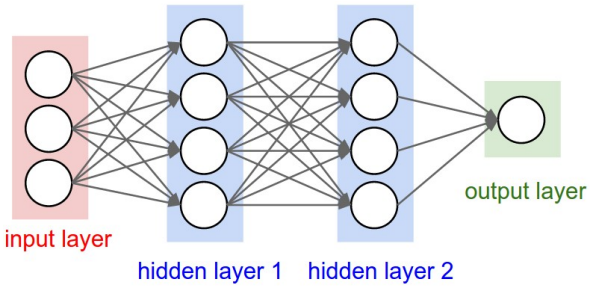


Figure 2.7: *Example of a fully connected neural network.*

Input values, each with a numeric value x_i, are given to the neurons on the first hidden layer. Each neuron has a specific weight w_i for each input x_i and a specific bias h. A weighted sum is calculated using the weights, inputs and bias, which are converted by the activation function f to an output of that neuron y (Equation 2.2). The output of all neurons in the first hidden layer can be calculated in this way. The same process is repeated to calculate the outputs of the second hidden layer. Finally, these outputs are used to calculate the output of the network. This progress of propagating the input values through the network to the output value is called forward propagation.

$$y = f\left(\sum_{i=1}^{n} w_i \cdot x_i + h\right) \tag{2.2}$$

The activation function f must be nonlinear since linear algebra shows that otherwise, any number of layers can be reduced to a two-layer input-output model. Some popular activation functions include the sigmoid function and the Rectified Linear Unit (ReLU) function. The weights in the network convert the different types of input to the desired output. Depending on the task, this input and output can have multiple forms. For example, if the task is to recognize a cat on an image, then the input values are the pixel values of the image and the output is a single bit which is one if it is a cat and zero if it is not a cat. However, suppose the task is to detect animals on an image. In that case, the output can be an array of percentages, giving the probability that the animal is in the image.

2.4.2 Training of Deep Learning Models

The network is trained to give the correct output for different inputs by adjusting the weights of the network. The weights are adjusted based on how well the predicted output corresponds to the correct output. The quality of a prediction is determined by the loss function. The loss function will output a high value if a certain prediction is off. The weights in the network are adapted to optimize the loss function. Some standard loss functions are the mean squared error function for regression problems or the cross-entropy loss for classification problems.

The loss calculated by the loss function is propagated through the networks to determine how the weights should change. This process is called backpropagation. Backpropagation identifies which pathways are more influential in the final answer and allows us to strengthen or weaken connections to arrive at the desired prediction. The influence of the weights on the final prediction is calculated and weights are changed based on this influence. Optimization of the loss function is done by gradient descent. This optimization algorithm minimizes the loss function by iteratively moving in the direction of the steepest descent as defined by the negative of the gradient. How much the weights in the model should be changed depends on the gradient size. Stochastic gradient descent can be used if the number of training samples is large or if each training sample contains much data. This means that the network parameters are updated after a few training examples instead of going through all the training examples before updating the parameters. Often, this can be a more efficient method of calculating the gradients.

2.4.3 Multi-Task Learning

Multi-task learning is a deep learning model that performs multiple tasks simultaneously while exploiting commonalities and differences across tasks. This can lead to improved competence in the tasks and higher learning efficiency when compared to training the models separately. The model uses all of the available data across the different tasks to learn generalized representations of the data that are useful in multiple contexts. In general, multi-task learning is useful when there are some commonalities between the tasks. At least there should be some underlying principles or information that can be shared between the tasks. Multi-task learning is commonly used in natural language processing, computer vision, and recommendation systems.

Multi-task learning can be either by hard parameter sharing or soft parameter sharing. Hard parameter sharing means some layers of the network are used by both tasks. Soft parameter sharing means that each model has its own parameters, but they can share information between the models. Hard sharing is more useful when parts of the task are very similar. Soft sharing allows some dissimilarity between the tasks but requires more memory because two networks are trained simultaneously.

2.5 Different Deep Learning Architectures

2.5.1 Convolutional Neural Networks

A CNN is a neural network that uses convolution operations instead of fully connected layers. A visualization of a convolution operation is given in Figure 2.8. They are crucial for many deep learning applications because they are computationally effective. They require far fewer parameters than a fully connected layer with the same number of inputs and outputs. This is because parameters are shared across the convolutional layer and the kernel is much smaller than the input. This enables them to be applied to larger input data, such as images. Also, their ability to only use local information to create feature maps is beneficial for input data with a grid-like topology. Inputs like a time series (1D grid) or an image (2D grid) are usually analyzed using CNNs.

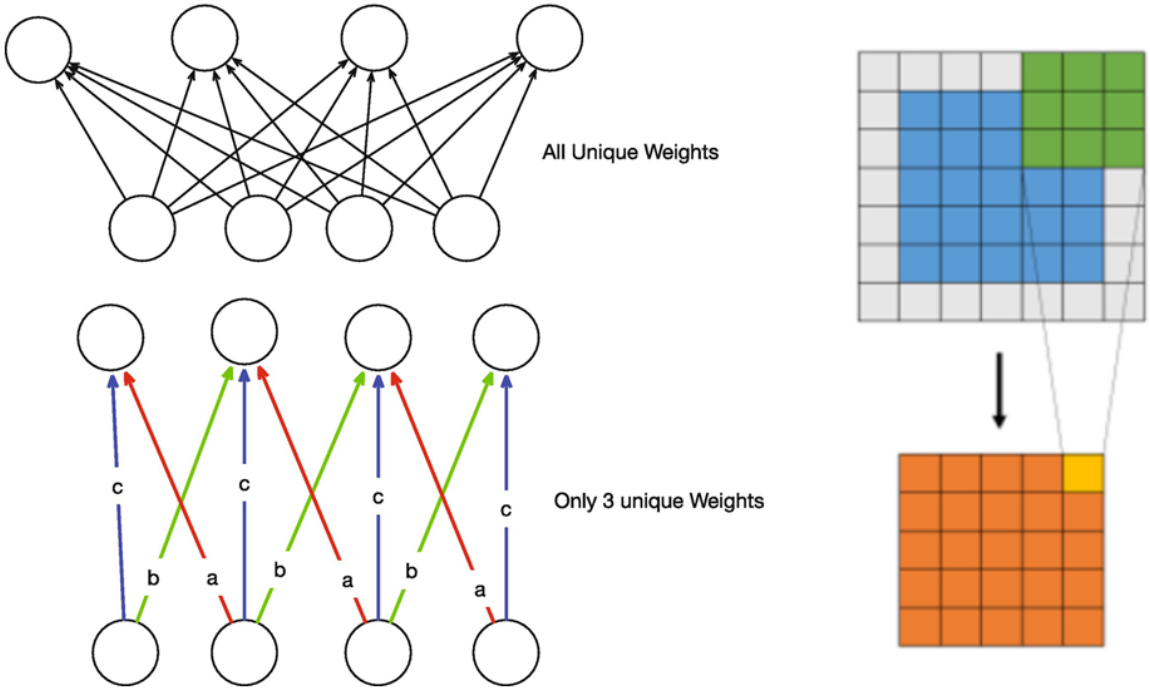


Figure 2.8: *Visualization of the convolution operation. On the left, a 1D convolution that highlights the difference in parameters required for a convolution layer or a fully connected layer. On the right a 2D convolution that shows how only local information is used to create the new feature map.*

An image can be represented in multiple ways, for example, how the color of an image is encoded using three color channels: red, blue and green. A convolution layer creates different image representations, which are called feature maps. This is done by sliding a kernel over the image that uses surrounding pixels to determine the pixel value in the feature map (Figure 2.8). A kernel of size 3 can be thought of as a filter with 3×3 weights. Depending on the weights in the kernel, different features, such as edges or orientations, can be extracted from the image. To ensure the size of the feature map is the same as the size of the image, the number of places the kernel moves between operations (stride) is usually one and an extra layer is added to the boundary of the image (padding). This way, each pixel in the image gets a corresponding pixel in the feature map. Similar, convolutions can be done in 3D, for 3D CT data for example. A convolution layer can have multiple feature maps as input and multiple feature maps as output. If the input is 16 feature maps, the output is 16 feature maps and we use a kernel of 3×3, then the convolution layer has 16×16 kernels with 9 weights each.

An operation that is often used in conjunction with convolution is a pooling operation. An example of a pooling operation is given in Figure 2.9. A pooling layer downsamples the input image. This is done so that different resolution feature maps can be extracted from the image and to decrease the memory required for the network. The pooling operation can be done in multiple ways, for example, with a filter of size two and stride two that takes the maximum or average of each set of pixels as pixel value for the new image.

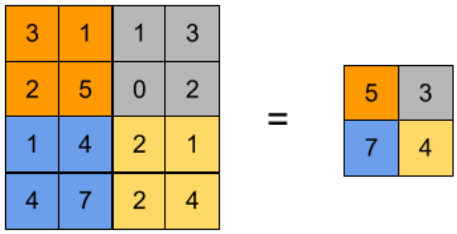


Figure 2.9: *An example of how the resolution of an image is decreased by a max pooling operation.*

2.5.2 U-Net

U-net is one of the most influential CNN architectures with numerous applications. It was first developed for biomedical image segmentation by Ronnenberger et al. [25]. Later, multiple adaptions to U-Net have been proposed to improve it or apply it in a different domain. For example, it was adapted to apply to 3D structures [41] and it was successfully applied for dose prediction [32]. The architecture of the U-Net consists of an encoder to encode the input image into feature maps and a decoder that tries to project the feature maps onto the pixel space of the original image. An architecture diagram of the U-Net is given in Figure 2.10.

In the first half of the network, the image is encoded into different resolution feature maps by convolution blocks that increase the number of channels and downsampling of the image. Downsampling is usually done with a pooling operation. In the second half of the network, the feature maps are decoded and projected onto the original pixel space. This is done by convolution blocks that decrease the number of feature channels and upsampling of the feature maps. Feature channels of the encoding half of the network are concatenated to feature channels of the decoding half that are on the same resolution level to preserve information from higher resolution feature maps. In the final layer, the feature maps that are extracted from the image are converted to a segmentation map by a 1×1 convolution.

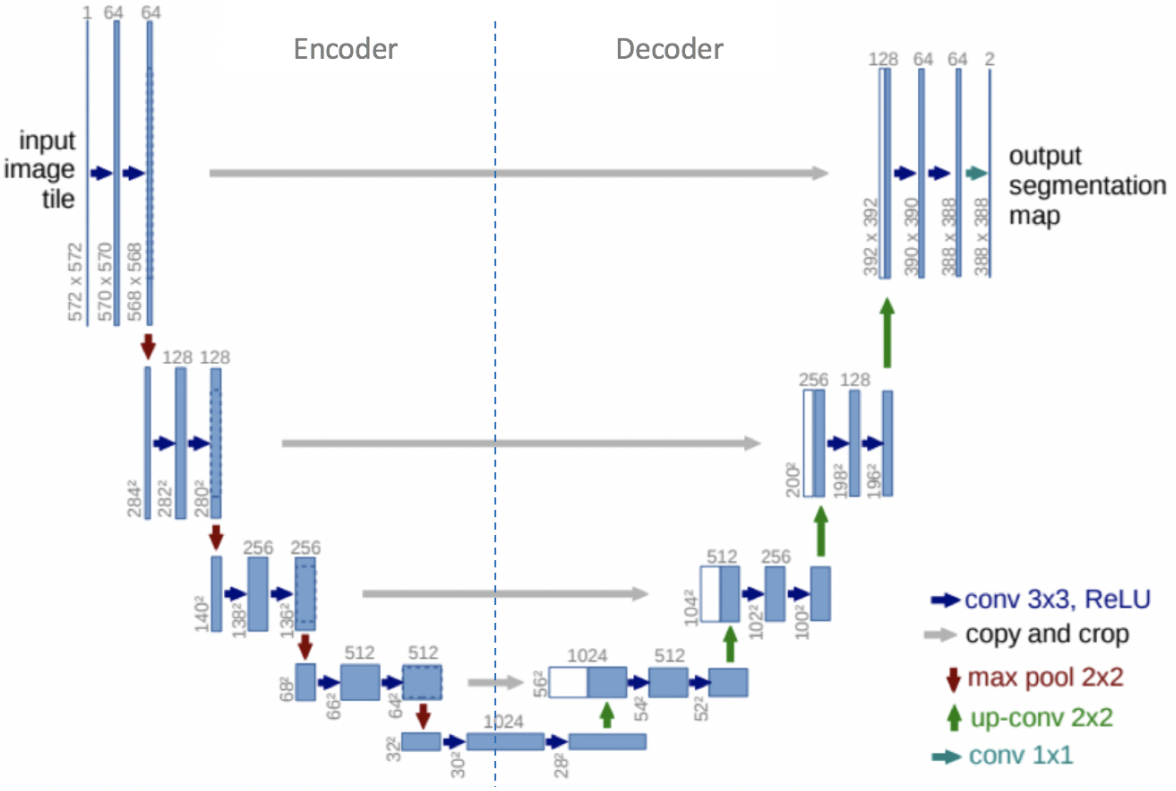


Figure 2.10: *Architecture diagram of the original U-Net. Arrows represent different operations. Rectangles represent the different feature maps created throughout the network. The numbers indicate the image size and the number of channels. Adapted from [25].*

2.5.3 Cross-Stitch Network

A cross-stitch network is a multi-task architecture developed by Misra et al. [42] that allows soft sharing of parameters between different networks. Each task is performed in its own network using its own parameters. Only at some locations, cross-stitch units are inserted that linearly combine feature maps from both networks. Consider that we want to share information between tasks A and B. At some point in the network, we haveu feature maps X_A and X_B of the same size. They are connected to a cross-stitch unit with learnable parameters α_{AA}, α_{BA}, α_{BB} and α_{BA}. The cross-stitch unit linearly combines the feature maps, so that the shared feature maps become $X_A^{'} = \alpha_{AA} \cdot X_A + \alpha_{BA} \cdot X_B$ and $X_B^{'} = \alpha_{BB} \cdot X_B + \alpha_{AB} \cdot X_A$. Since the parameters that determine how much of the feature maps should be shared are learnable, representations can be shared flexibly. If it is beneficial to share the feature maps, the network can do this, but if it is better to use independent feature maps, the network can separate them.

Beljaards et al. used a cross-stitch network to combine a 3D U-Net for image segmentation and registration [6]. Both networks extract feature maps from the input image, which they use to perform their task independently. At four points in the U-Net, after the downsampling and upsampling layers, the feature maps of the two networks are linearly combined by cross-stitch units. Each network can use a combination of feature maps created by the segmentation and registration network. They found that using the cross-stitch units significantly increased the performance of the registration and segmentation network compared to the single-task networks.

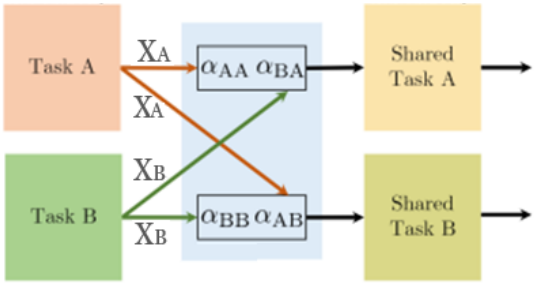


Figure 2.11: *The feature maps X of the two tasks are linearly combined by the learnable weights α, creating shared feature maps.*

3 Methods & Materials

This chapter will describe the data and different methods that were used for contour prediction and optimal dose prediction. Section 3.1 and Section 3.2 give details about the CT scans and the treatment plans used in this research. Section 3.3 gives general information about the used network architecture and the training of the networks. Section 3.4 gives more specific information about the segmentation, registration and cross-stitch networks for contour definition, and Section 3.5 gives specific information regarding the optimal dose prediction networks. Finally, Section 3.6 describes how the contour definition network and the dose prediction network are combined.

3.1 Patient Dataset

A dataset from Haukeland Medical Center is used for this study. It contains a planning CT scan and daily CT scans used for the intensity-modulated radiotherapy treatment of prostate cancer. The dataset includes 18 patients, all having a planning CT scan and 7 to 10 daily CT scans. The CT scans are made with a GE scanner and consist of 90 to 180 slices with a thickness of 2.0 mm. Each slice contains 512×512 voxels having a pixel size of approximately 0.9×0.9 mm. Manually defined contours of the prostate, seminal vesicles, bladder and rectum are available. Since the lymph nodes are not defined for all CT scans in the dataset, they are disregarded in this study. In total, 161 daily CT scans are available for training, validation and evaluation of the networks. The 129 daily CT scans from the first 14 patients are used for training. The 32 daily CT scans from the remaining 4 patients are used for validation and testing.

3.2 Proton Therapy Treatment Plans

3.2.1 Treatment Planning System: MatRad

The open-source treatment planning system matRad [43] was used to generate the proton therapy treatment plans. MatRad is a toolbox built on the MATLAB environment [44] and is meant for educational and research purposes. The created plans are for intensity-modulated proton therapy (IMPT) and consider two gantry angles at 90° and 270°, so the proton beams pass laterally through the body. For the RBE, a constant value of 1.1 is assumed for all tissues, which is considered standard in proton therapy [45]. The dose grid resolution that was used to optimize the treatment plan is 2 mm and the treatment consists of 37 fractions.

In MatRad, multiple constraints and objectives can be set for the dose distribution. All objectives have a weight based on the importance of the objective. The cost function to optimize is a weighted sum of these objective functions. Equation 3.1 is an example of how the list of objectives and constraints in Table 3.1 is converted to an optimization problem using the method described by Equation 2.1. Some of the constraints and objectives available in MatRad are as follows. Constraints can be set on the minimum or maximum dose in a particular structure, which the dose distribution can not exceed. Objectives regarding specific DVH points (Section 2.3.3) can also be set. For example, dose distributions where the 98^{th} percentile highest dose exceeds a certain threshold are penalized. These objectives are beneficial when preventing very high doses from being present in some local regions of a structure. In the same way, the overall dose in certain volumes can be limited by striving for a low mean dose in those structures.

Constraint	Structure	Objective	Weight	Structure
g_1	1	f_1	50	1
g_2	2	f_2	20	3
		f_3	5	3

Table 3.1: *Example of constraints and objectives for treatment planning.*

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & 50f_1(d_1) + 20f_2(d_3) + 5f_3(d_3) \\ \text{subject to} \quad & g_1(d_1) \geq b_1, \\ & g_2(d_2) \geq b_2, \\ & x \geq 0. \end{aligned} \tag{3.1}$$

3.2.2 Objective List Used for Treatment Planning

The objective list used for treatment planning is given in Table 3.2. It is adapted from a wishlist [16] that was used to generate IMPT treatment plans for the same dataset using a treatment planning system called ICycle [46]. Adaptation was required since ICycle has a different method of optimization and different structures have been used for those plans. The target dose for the prostate is 74 Gy and the target dose for the seminal vesicles is 55 Gy. The dose in the rectum and bladder is held as low as possible. The distribution of the weights is ordered similarly to the ranking of the priorities in the original wishlist. By trial and error, the constraints and objectives from Table 3.2 were found to produce the highest quality plans[1], ensuring sufficient dose in the target volumes without overdosing them and as little dose as possible in the organs at risk.

Table 3.2: *The planning constraints and objectives used for the IMPT treatment plans. PTV-high corresponds to the prostate and PTV-low corresponds to the seminal vesicles.*

Constraints

	Structure	Type	Limit
g_1	PTV-high	Minimum	0.97×74 Gy
g_2	PTV-low	Minimum	0.97×55 Gy

Objectives

	Weight	Structure	Type	Limit
f_1	400	PTV-high	D_2 Maximum	1.07×74 Gy
f_2	400	PTV-low	D_2 Maximum	1.07×55 Gy
f_3	50	Rectum	D_2 Maximum	60 Gy
f_4	50	Bladder	D_2 Maximum	60 Gy
f_5	20	Rectum	D_{mean}	0 Gy
f_6	20	Bladder	D_{mean}	0 Gy
f_7	10	Torso	D_{mean}	0 Gy

3.3 Network Architecture and Training Details

3.3.1 Network Architecture

The architecture used for the networks in this research is a 3D deep CNN inspired by the U-Net used by Fan et al. and Beljaards et al. [47, 6]. An architecture diagram of the network used for single-task contour definition is given in Figure 3.1. Architectures used for the other networks are similar to this one and specifics about adjustments can be found in the section regarding that network. The network consists of an encoder to encode the input image into feature maps and a decoder that tries to project the feature maps onto the pixels space of the original image. The first half of the network is the encoding path,

[1] For some underperforming plans, objective functions and weights were slightly adjusted.

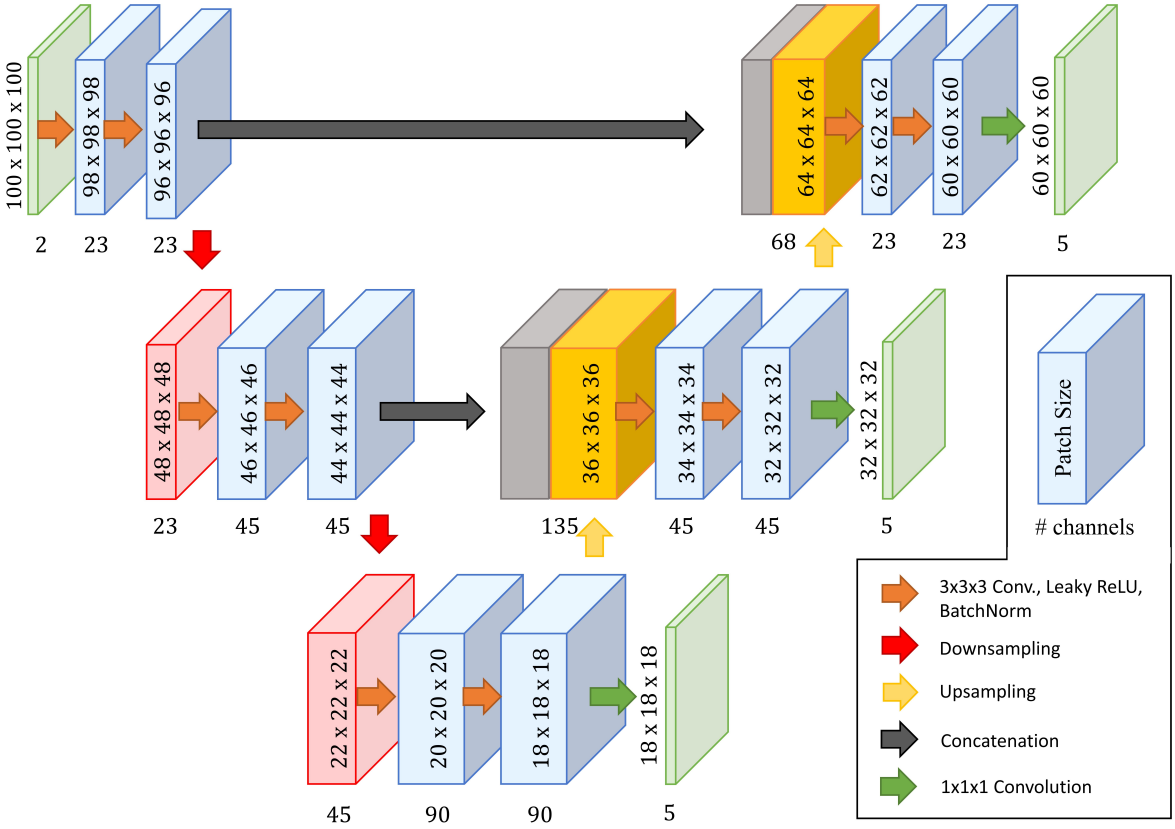


Figure 3.1: *An architecture diagram of the 3D deep CNN used in this study. Blocks are image representations of a given patch size and number of channels. A 3D image of size* $100 \times 100 \times 100$ *is converted to a 3D image of size* $60 \times 60 \times 60$*. Arrows describe the different operations done to the image.*

where convolution blocks that increase the number of feature channels are combined with downsampling of the feature maps to extract feature maps of different resolutions from the image. The second half is the decoding path, where convolution blocks that decrease the number of feature channels are combined with an upsampling of the feature maps to project the feature maps learned by the encoder onto the pixel space of the original image. Feature channels of the encoding half of the network are concatenated to feature channels of the decoding half that are on the same resolution level to preserve information from higher resolution feature maps. For the input a $100 \times 100 \times 100$ patch is taken from the CT scan. Feature maps from different resolution levels are converted to three outputs by $1 \times 1 \times 1$ convolutions. More information on the patches that are used as input and the multi-level outputs can be found in Section 3.3.3.

In total, there are ten convolution blocks present in the network. Six in the encoding path, where they maintain or increase the number of feature channels and four in the decoding path, where they maintain or reduce the number of feature channels. Each convolution block consists of a $3 \times 3 \times 3$ convolutional layer, a Leaky ReLU activation function and batch normalization. The 3D convolutional layer gets a certain amount of feature channels as input and by sliding different $3 \times 3 \times 3$ filters over these feature channels, they produce a certain amount of feature channels as output. Since the input patches are not padded before convolution, the outer pixel at all sides of the patch is lost and the operation reduces the patch size by 2 pixels. A Leaky ReLU with a slope of 0.02 is used as activation function (Equation 3.2). It has the same benefits as the ReLU activation function described in Section 2.4, but it has a small, positive gradient when the unit is not active. This is useful when sparse gradients occur in the network. Batch normalization layers [48] are applied to normalize the inputs of the layers by re-centering and re-scaling them. This ensures the network converges faster and it makes the network more stable.

$$f(x) = \begin{cases} x & x > 0 \\ 0.02\,x & x \leq 0 \end{cases} \tag{3.2}$$

After two convolution blocks in the encoding path, the feature maps are downsampled to half the number

of voxels along each dimension. This downsampling is done by trilinear interpolation using a scale factor of 0.5. To scale the feature maps back up in the decoding path, a trilinear interpolation with a scale factor of 2 is used to double the number of voxels along each dimension. This trilinear interpolation approximates the value of a pixel in a down- or upsampled feature map linearly by looking at using the values of the surrounding pixels. The feature maps from before the downsampling are concatenated to the feature maps after the upsampling to preserve some of the higher resolution features.

3.3.2 Implementation of the Networks

All the networks are implemented using PyTorch [49]. PyTorch is an open-source machine learning and deep learning library developed for Python, which is widely used in the machine learning community. It has easy methods for using GPUs, applying cost or activation functions, performing backpropagation and many more necessities for deep learning. In addition to PyTorch, NiftyNet [50] modules have been used for sampling of the data. NiftyNet is a open-source CNN platform for research in medical image analysis. Data preparation and analysis of the CT scans and contours is done partly using SimpleITK [51]. SimpleITK is an open-source software library for image analysis. All networks are trained, validated and tested on an Nvidia Quadro RTX 6000 GPU with 16 GB of memory.

3.3.3 Training Details

Because of the amount of data in a single 3D CT scan, patch-based networks are used in this research. Patch-based means that a smaller patch from the large 3D image is used as input for the network. The smaller patch size reduces the amount of memory required by the network and the partly random selection of the patch locations serves as data augmentation since many different patches can be taken from one 3D image. For the networks used in this study, windows with a size of $100 \times 100 \times 100$ pixels are taken from the images. Each iteration consists of selecting 4 windows from a single CT scan and use those as input for the network. For contour definition, the centers of these 4 windows are sampled from the 4 structure volumes that must be defined. For dose prediction, 2 windows are sampled from the 2 target volumes and the other 2 windows are sampled from the rest of the dose distribution that must be predicted.

In the conventional U-Net, the loss is calculated only in the final layer. This can result in slow convergence, overfitting and sub-optimal parameters in the first half of the network. To ensure a more direct supervision of the entire network, a loss function is applied on multiple levels [47]. Besides the final layer where a high-resolution prediction is made, also a mid-resolution prediction is made on the 2nd level and a low-resolution prediction on the 3rd level. Because of the downsampling layers and the $3 \times 3 \times 3$ convolution layers without padding, the patch size of these high, mid an low level predictions are $60 \times 60 \times 60$, $32 \times 32 \times 32$ and $18 \times 18 \times 18$, respectively. Since the output image is smaller than the input image, information from voxels surrounding the patch can be used for the prediction. For each level, a combination of isotropic cropping and downsampling is applied to the image which must be predicted. This ensures that the prediction corresponds to the correct representation of the image. The losses of the predictions of all levels are added up evenly to obtain the total loss of the network. Network optimization is done using RAdam [52] with a learning rate of 5×10^{-4}. RAdam uses stochastic gradient descent with an adaptive learning rate, which is an effective and robust method of optimization.

3.3.4 Inference in a Patch-Based Network

Inference is required to make a complete prediction for a previously unseen CT scan using the network. In regular networks, this is simply done by putting a new image into the network, which immediately results in a prediction. However, since this is a patch-based network, only part of the output image can be predicted at a time. The image is divided into blocks of $60 \times 60 \times 60$ and a prediction is made for each block by taking a $100 \times 100 \times 100$ patch around this block. Padding of the image is used for blocks near the boundary. Since the $60 \times 60 \times 60$ blocks have been made to overlap halfway, around 8 predictions are made for each voxel [2]. The average of the different predictions is used. This method of overlapping blocks has been found to give slightly better predictions than placing the blocks next to each other, although the difference was slight.

[2] fewer predictions are made close to the boundary of the image

3.4 Contour Prediction Network

3.4.1 Evaluation Metrics for Predicted Contours

The goal of the contour definition networks is to predict the contours on the daily CT scan. The performance of different networks is evaluated by comparing the predicted daily contours C_d^{pred} with the ground-truth daily contours C_d that are manually delineated by radiation oncologists. Three metrics are used to define the similarity between C_d^{pred} and C_d [53]:

Mean Surface Distance (MSD): The average absolute distance between C_d^{pred} and C_d. For each point on C_d^{pred}, the shortest distance to C_d is calculated (Figure 3.2) and the average is taken of all points. This results in an average distance, where a lower MSD means a better prediction.

Dice Similarity Coefficient (DSC): The overlap ratio between C_d^{pred} and C_d. The overlap between the predicted volume and the actual volume is divided by the sum of their volumes (Equation 3.3). This gives a number between 0 and 1, where 0 means there is no overlap and 1 means the two volumes are the same.

$$\text{DSC} = \frac{2\,|A \cap B\,|}{A + B} \tag{3.3}$$

95% Hausdorff Distance (HD): The 95$^{\text{th}}$ percentile largest distance between C_d^{pred} and C_d. For each point on C_d^{pred}, the shortest distance to C_d is calculated (Figure 3.2), and the 95$^{\text{th}}$ percentile of all points is taken. This is similar to MSD, but more focused on outliers.

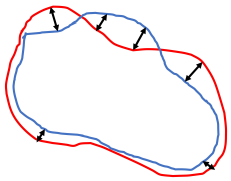


Figure 3.2: *Absolute distance between two surfaces.*

3.4.2 Segmentation Network

The goal of the segmentation network is to define the important structures on the daily CT scan. This is done by classifying all the voxels into categories that represent the different structures. An overview of the segmentation network is given in Figure 3.3. The input to the network consists of two channels, one being

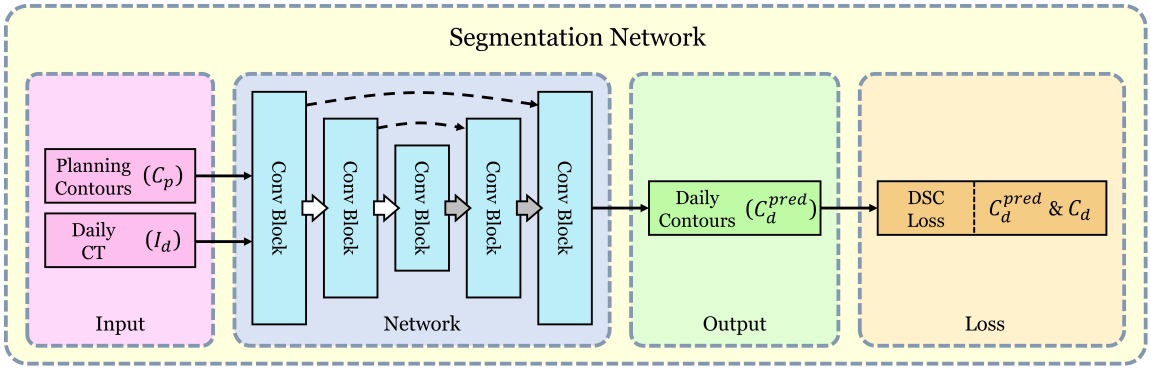


Figure 3.3: *An overview of the segmentation network. C_p and I_d are given as inputs to a U-Net, which then predicts the contours on the daily CT (C_d^{pred}). The loss is based on the difference between the predicted and the ground-truth contours.*

the contour mask of the planning CT (C_p) and the other being the daily CT scan (I_d). C_p has different integer values for the different structures and I_d has voxel intensities scaled from -1 to 1. The network is a U-Net as described in Section 3.3 with three resolution levels, the different levels having 23, 45 and 90 filters, respectively. At the end of the network, a $1 \times 1 \times 1$ convolution layer converts the output channels to 5 channels, each belonging to one of the structures. A softmax function is then applied to determine, for each voxel, the probability that it is part of one of the structures. The DSC between these probabilities and the actual contours is calculated and the loss is given as 1 - DSC.

3.4.3 Registration Network

The goal of the registration network is to define a transformation from the planning CT scan (I_p) to the daily CT scan (I_d). This transformation is then used to propagate C_p to I_d. An overview of the registration network is given in Figure 3.4. The input to the network consists of two channels, one having the voxel intensities from I_p and the other having the voxel intensities from I_d. The network is a U-Net as described in Section 3.3 with three resolution levels, the different levels having 23, 45 and 90 filters, respectively. At the end of the network a $1 \times 1 \times 1$ convolution layer converts the output channels to 3 channels, each belonging to a displacement in one dimension.

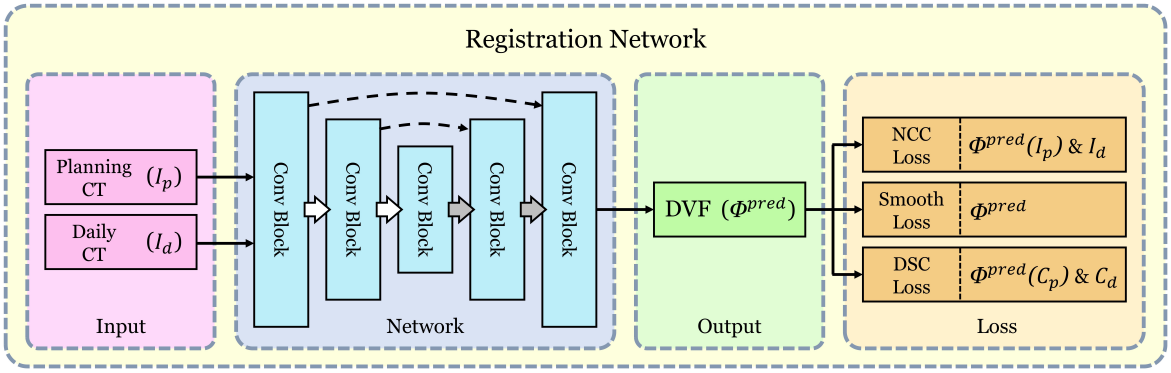


Figure 3.4: *An overview of the registration network. I_p and I_d are given as inputs to a U-Net, which then predicts a DVF that can transform I_p to I_d. The loss consists of the NCC loss between both images, the bending energy of the DVF and the DSC loss between the contours on both images.*

The loss of the registration network is a combination of three parts. The first part penalizes dissimilarity between the transformed planning CT scan and the daily CT scan. The similarity between the two images is calculated using the **normalized cross-correlation (NCC)** given in Equation 3.4:

$$\mathrm{NCC}(f,g) = \frac{1}{\mathrm{n}} \sum_i \frac{(f(k_i)-\mu_f)(g(k_i)-\mu_g)}{\sigma_f \sigma_g} \tag{3.4}$$

where f and g are two images, μ is the average intensity, σ is the standard deviation of the intensity, n is the number of voxels and k_i represents voxel i. The benefits of using the NCC are that there is compensated for brightness differences between images and that broader image similarities are taken into account. The second part of the loss penalizes DVFs with high gradients that do not correspond to physical transformations. The bending energy of the DVF is taken as a loss to guarantee a smooth transformation. This also helps to regularize the network since it is made harder to overfit to the training data. The third part of the loss is the DSC loss between the transformed planning and manual contours. This loss is similar to the loss in the segmentation network, which ensures similar regions in the transformed planning CT and daily CT overlap. The total loss is the sum of these three losses, where the NCC loss and the DSC loss have weight 1 and the bending energy has weight 0.5.

3.4.4 Cross-Stitch Network

The segmentation network described in Section 3.4.2 and the registration network described in Section 3.4.3 are also combined in a cross-stitch network, as previously done by Beljaards et al. [6]. An overview of the cross-stitch network is given in Figure 3.5. Both the segmentation and registration path can be used to define contours on the daily CT scan. The idea is that both networks can share helpful

information, if this improves their performance. Pairs of feature maps from the segmentation and registration path are linearly combined using learnable parameters, as explained in Section 2.5.3. Cross-stitch units are placed after each of the two downsampling layers and each of the two upsampling layers, so at four places in total. The number of filters at the different resolution levels are 16, 32 and 64, respectively, instead of the 23, 45 and 90 filters used in the single-task networks. This is $\sqrt{2}$ times less filters, which ensures that the cross-stitch network has roughly the same number of trainable parameters as the segmentation and registration networks ($7.8 \cdot 10^5$). The loss of the network is calculated by adding the loss from the segmentation network and the registration network.

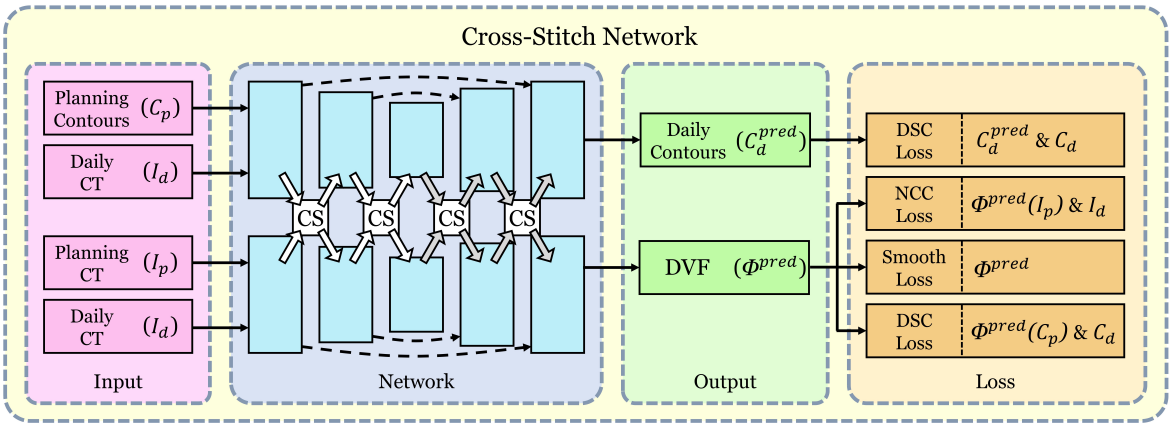


Figure 3.5: *An overview of the cross-stitch network. Both the segmentation path (top) and the registration path (bottom) can define the daily contours. After each down- and upsampling layer, cross-stitch units are placed to linearly combine featuremaps from both networks. The loss is a sum of the segmentation and registration loss.*

3.5 Dose Prediction Network

3.5.1 Evaluation Metrics for Predicted Dose Distributions

The gamma index (γ) is a metric that compares two dose distributions by combining dose difference and distance to agreement. For each point in the reference dose distribution, a nearby point in the evaluated dose distribution is sought with an equal dose. This is an improvement upon metrics only looking at point-by-point dose differences since these can be inaccurate in regions of high dose gradient. Originally γ was used in the quality assurance of radiotherapy delivery, where the dose distribution calculated by a treatment planning system is compared to the measured dose distribution [54]. This study uses it to compare the predicted 3D dose distribution with the ground-truth dose distribution calculated by the treatment planning system.

For each point in the reference dose distribution, the minimum distance to the evaluated dose distribution in the dose-spatial domain must be calculated (Figure 3.6). When the distance $\Delta r(\mathbf{r}_R, \mathbf{r}_E)$ and the dose difference $\Delta D(\mathbf{r}_R, \mathbf{r}_E) = D_R(\mathbf{r}_R) - D_E(\mathbf{r}_E)$ between a reference point $\mathbf{r}_R$ and an evaluated point $\mathbf{r}_E$ are known, the gamma between these two points can be calculated using Equation 3.5:

$$\Gamma(\mathbf{r}_R, \mathbf{r}_E) = \sqrt{\frac{\Delta r^2(\mathbf{r}_R, \mathbf{r}_E)}{\delta r^2} + \frac{\Delta D^2(\mathbf{r}_R, \mathbf{r}_E)}{\delta D^2}} \tag{3.5}$$

where δr is the distance to agreement and δD is the dose difference criterion. The lowest possible Γ between the reference point and the evaluated dose distribution is taken as γ for the reference point (Equation 3.6).

$$\gamma(\mathbf{r}_R) = \min_{\forall \mathbf{r}_E} \Gamma(\mathbf{r}_R, \mathbf{r}_E) \tag{3.6}$$

When a reference point has a low gamma index, the evaluated and the reference dose distribution are similar at that reference point. In our case, this means the predicted dose is similar to the ground-truth dose.

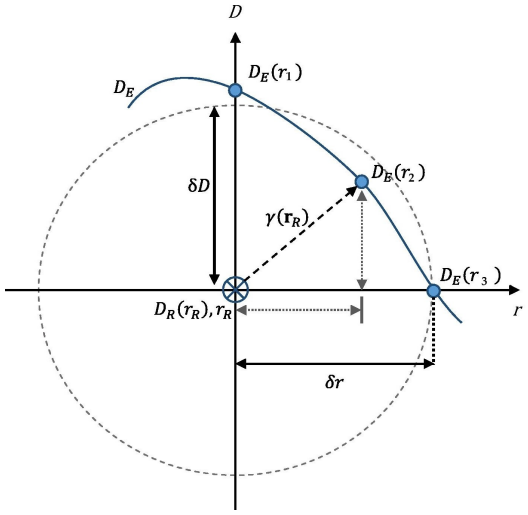


Figure 3.6: *Schematic representation of the gamma index in 1D. The reference point is located in the origin, the dose of a point is given on the y-axis and the distance to the reference point is given on the x-axis. δr and δD create an ellipse around the reference point inside which evaluated points pass the gamma test. In the above example: $\Gamma(\boldsymbol{r}_R, \boldsymbol{r}_1) > 1$, $\Gamma(\boldsymbol{r}_R, \boldsymbol{r}_2) < 1$ and $\Gamma(\boldsymbol{r}_R, \boldsymbol{r}_3) = 1$. $\gamma(\boldsymbol{r}_R)$ is the minimum of all evaluated points and since $\gamma(\boldsymbol{r}_R) < 1$, the reference point $\boldsymbol{r}_R$ passes the gamma test. Adapted from [55].*

.

The two acceptance criteria that determine if the reference and evaluated dose distribution agree are the maximum allowable deviation in dose δD and the maximum allowable distance that such a point can be away δr. If $\gamma(\mathbf{r}_R) \leq 1$, there is an evaluated point that meets these two acceptance criteria and the reference point passes the gamma test. The gamma pass rate (GPR) is the percentage of points from the reference dose distribution that passes the gamma test. In this research, a predicted dose distribution is evaluated by the GPR achieved in the different structures. Usually, δD is given in % and δr is given in mm. The original paper introducing the gamma index [54] recommended 3%/3mm criteria, but recent studies have shown these criteria are outdated for present-day accuracy requirements [56]. Therefore, in this study, δD and δr criteria of 2%/2mm are used. Global gamma analysis is done, which means a nominal value of 2% of the maximum dose is chosen for δD. For each reference point, 65 nearby points from the evaluated dose distribution are examined to find a point that passes the gamma test. To speed up the calculation, searching is stopped when a point has been found that passes.

A second metric used to evaluate a predicted dose distribution is the comparison of DVH points and average doses from both dose distributions. The absolute difference between DVH points of two dose distributions can be calculated in % using Equation 3.7:

$$\Delta\text{DVH} = \left| \frac{D_{true} - D_{pred}}{D_{true}} \right| \cdot 100\,\% \tag{3.7}$$

where D_{true} and D_{pred} are a certain DVH point in both dose distributions. A small ΔDVH means the DVH points of the two dose distributions are similar. This metric is handy when examining if the network predicts a dose distribution of similar quality to the ground-truth dose rather than being a replica of the ground-truth dose. This could indicate what kind of doses are achievable for the different structure

3.5.2 Optimal Dose Prediction Using Daily Contours

The goal of the dose prediction network is to predict what an optimal dose distribution will look like for the daily CT scan (I_d). The network uses the contours of the structures on the daily CT scan to help determine what the dose distribution will look like. In APT manual contours are generally not available for a new daily CT scan, so this method will serve as a benchmark for other approaches that try to predict an optimal dose distribution without knowing the contours. An overview of the dose prediction network is given in Figure 3.7. Four different kinds of input are given to the network. The first input consists of the contours of the structures on the daily CT scan (C_d), which are given to the network in four channels, each containing

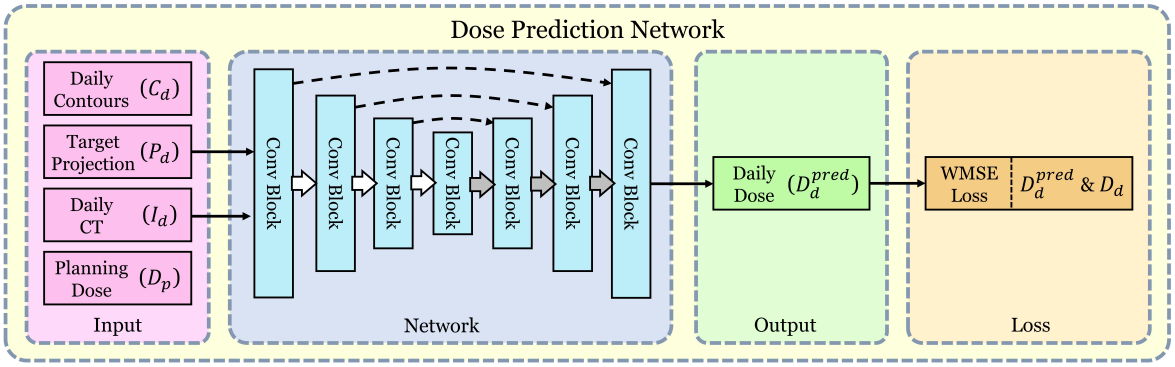


Figure 3.7: *An overview of the dose prediction network. C_d, P_d, I_d and D_p are given as inputs to a U-Net, which then predicts an optimal dose distribution for the daily CT (D_d). The loss is a weighted mean square error function between the predicted- and the ground-truth dose distribution.*

the mask of one structure. The second input consists of the lateral projections of the two target volumes. This is given to the network in two different channels, each containing the mask of one target volume projection. The projection of the target volumes is in the same direction as the proton beams, so this input helps the network determine the location of the targets in places that lie far away from the structures. Without this knowledge, the network has no way of determining the height of the proton beam in patches that do not contain any structure masks. The third input is a channel containing the voxel intensities from the daily CT scan I_d. This information might help the network determine how deep proton beams reach since, for protons, this depends highly on the density of the material it crosses through. The last input is a channel containing the normalized dose values of the planning dose distribution D_p. The shape and intensity of a dose distribution similar to the optimal dose distribution can be helpful for the network.

The network used for dose prediction is similar to the one described in Section 3.3, only it is four levels deep instead of three. The convolutional layers added to create the extra low-resolution level are zero-padded to ensure the convolutions do not decrease the patch size. This ensures the patch size does not get too small and that the patch size in the rest of the network stays the same. The modifications made to the low-resolution level of the network of depth three shown in Figure 3.1 are depicted in Figure 3.8. At the three highest resolution levels, the last feature map is converted to one output channel via a $1 \times 1 \times 1$ convolution layer. These output channels contain the predicted dose for each voxel.

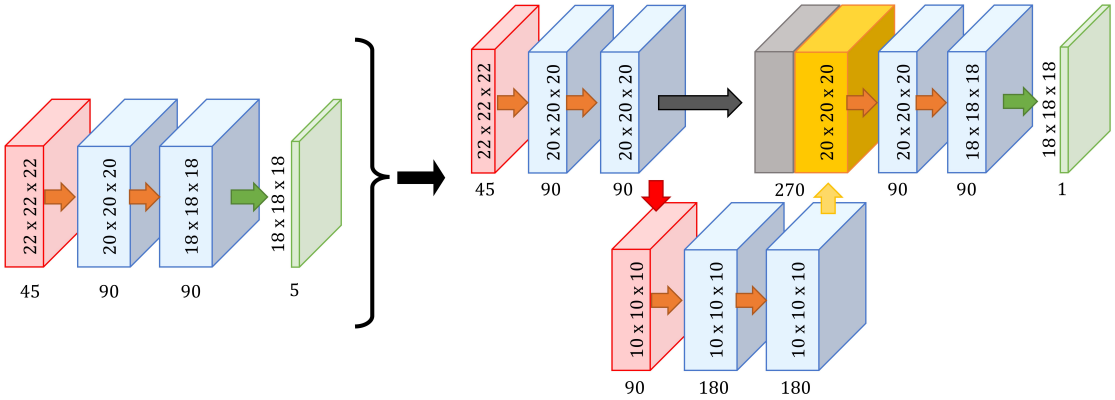


Figure 3.8: *A visualization of how the 3rd level of the network depicted in Figure 3.1 is adjusted to make the network a level deeper. The patch size in the rest of the network stays the same because the convolutions are zero-padded.*

The loss for the dose prediction networks is calculated by a weighted mean squared error (WMSE) function. This loss is based on the difference between the predicted and the ground-truth dose distribution, with larger differences penalized harder. Errors in the structures are weighted higher to encourage the network to make good predictions there, especially when an error comes from under-predicting the dose in the

target volumes. The WMSE loss function is given in Equation 3.8:

$$\text{WMSE}(d_i^{pred}) = \frac{1}{n} \frac{\sum_{i=1}^{n} w_i (d_i^{pred} - d_i^{true})^2}{\sum_{i=1}^{n} w_i} \tag{3.8}$$

where w_i are the weights, d_i^{pred} and d_i^{true} are the predicted and actual dose in voxel i, and n is the number of voxels in the predicted patch. The weights that are used can be found in Table 3.3.

Table 3.3: *Weights for the WMSE loss in different structures. The prostate and seminal vesicle errors from underprediction have a different weight than those from overprediction.*

Prostate		Seminal Vesicles		Rectum	Bladder
Overpred.	Underpred.	Overpred.	Underpred.	Both	Both
40	20	30	10	5	5

3.5.3 Optimal Dose Prediction Without Manual Contours

Since no manual contours are available for dose prediction in APT, this section explains three different methods of predicting an optimal dose distribution for a new daily CT scan without using manual contours.

Predicted Contours: The first method uses the network, as explained in Section 3.5.2, which is trained with manual contours. Only for the inference for CT scans from the test set, we do not use the manual contours, but we use the predicted contours from our contour prediction network. In this study, we used the predicted contours from the segmentation path of the cross-stitch network. One of the disadvantages of training the network on the manual contours that were also used for the treatment planning is that the network might build a dependency on highly accurate structure locations. The network could then under-perform when tested using predicted contours.

No Contours: The second method is training a dose prediction network that does not require structure masks as input. The inputs given to the network are the daily CT scan I_d, the planning CT scan I_p, the planning dose distribution D_p and the contours of the planning CT scan C_p. The same network architecture and WMSE loss function described in Section 3.5.2 are used. The idea is that the network can use what the optimal dose distribution looks like on the planning CT scan and learns to transform that to an optimal dose distribution on the daily CT scan. A potential problem with this method is predicting dose distributions in patches that do not contain any structure masks.

DVF on Planning Dose: The third method uses the deformation vector field determined by the registration network explained in Section 3.4.3 to transform the planning dose onto the daily CT scan. This method does not guarantee that the dose distribution is deliverable to the patient. This method also does not predict an optimal dose distribution since it does not involve a network that is trained on optimal dose distributions. Nevertheless, it could be interesting to see the result obtained by simply transforming the planning dose distribution onto the new image. If other approaches do not outperform this method, it is clear they fail to predict any optimality whatsoever.

3.6 Combining Contour Definition and Optimal Dose Prediction

In Section 3.5, we discussed multiple deep learning methods for predicting an optimal dose distribution. One of the problems with predicting optimal dose distributions for daily CT scans in APT is that the manual contours are not available for the prediction. This means either a prediction must be made without the contours or the contours must be defined first using another deep learning method. In this section we describe two multi-task models that can perform contour prediction and optimal dose prediction simultaneously. The idea is that the dose prediction task could benefit from the information obtained by performing the contour definition task.

3.6.1 Cross-Stitch Network for Contour Prediction and Dose Prediction

The first multi-task model that we use to combine contour definition and optimal dose prediction is the cross-stitch network that was already described for combining segmentation and registration in Sec-

tion 3.4.4. Two different cross-stitch networks are proposed: one combining dose prediction with segmentation and one combining dose prediction with registration. Since the dose prediction network is one level deeper than the contour definition networks, it is not possible to share any feature maps of the lowest level of the dose prediction network. Since the patch sizes and number of channels of the feature maps in the upper part of the networks are the same, they can be linearly combined just as in the cross-stitch network depicted in Figure 3.5. Cross-stitch units are placed after the first two downsampling layers and last the two upsampling layers, so at four places in total.

For the dose prediction network, the same inputs are used as for the dose prediction network that did not use contours as input. So, the daily CT scan I_d, the planning dose distribution D_p and the contours of the planning CT scan C_p. For the segmentation and registration network, the same inputs are used as described in Section 3.4. For the segmentation and registration network, the number of filters at the different resolution levels are 23, 45 and 90, respectively. For the dose prediction network, the number of filters at the different resolution levels are 23, 45, 90 and 180, respectively. The loss from both single-task networks is added to obtain the loss of the cross-stitch network. The WMSE loss is multiplied by 200 to ensure both losses are roughly of the same order. Both networks are trained for 400 epochs after which the validation loss did not decrease anymore.

3.6.2 W-Net: Predicting Contours and Optimal Dose Consecutively

Where the cross-stitch network tries to perform both tasks simultaneously, the w-net tries to perform the two tasks consecutively. An overview of w-net is given in Figure 3.9. In the first half of the network, a segmentation network predicts an image segmentation for the daily CT scan. Then, in the second half of the network, that segmentation is given as input to a dose prediction network. The architecture of the segmentation network is similar to the one described in Section 3.4.2. The architecture used for the dose prediction network is the same as the manual contour-based dose prediction described in Section 3.5.2. The only difference for the dose prediction network is the inputs. The used contours are now predicted instead of defined manually and the lateral projections of the target volumes are not used. The name w-net comes from the shape created when two U-Nets are placed after each other.

The w-net is trained in three stages. In the first stage, only the segmentation part of the network is trained. The daily CT scan I_d and the planning contours C_p are only put through the first network, after which the predicted daily contours are compared against the manual daily contours using the DSC loss. The input patches are sampled from the four different structures that must be predicted, just as in the contour definition networks. The second stage involves freezing the weights of the first half of the network and only training the second half of the network. A segmentation of the daily CT scan C_d is predicted in the first half of the network, which is given as input to the dose prediction network alongside the daily CT scan I_d and the planning dose D_p. The second half of the network is now trained using the WMSE loss. The patches are now sampled from the dose distribution, as is the case for all other dose prediction networks. After the second stage of training, w-net is already able to predict a dose distribution for a daily CT scan. Now in the third stage of training, the first half of the network is unfrozen, which allows the segmentation and dose prediction network to train together. Training the network to perform both tasks simultaneously may result in more dose-conforming contours and thus in better dose predictions. The loss of the network when training together is a sum of the DSC loss and the WMSE loss, where the WMSE loss is multiplied by 600, so the two losses are approximately the same size. This ensures the WMSE loss is roughly four times larger than the DSC loss. Since the goal is an accurate dose prediction, this should be valued higher. Training the segmentation network with both the DSC loss and the WMSE loss encourages it to predict contours that lead to good dose predictions. The networks are first trained separately for 400 epochs and then trained together for 200 epochs, since the validation loss did not improve anymore.

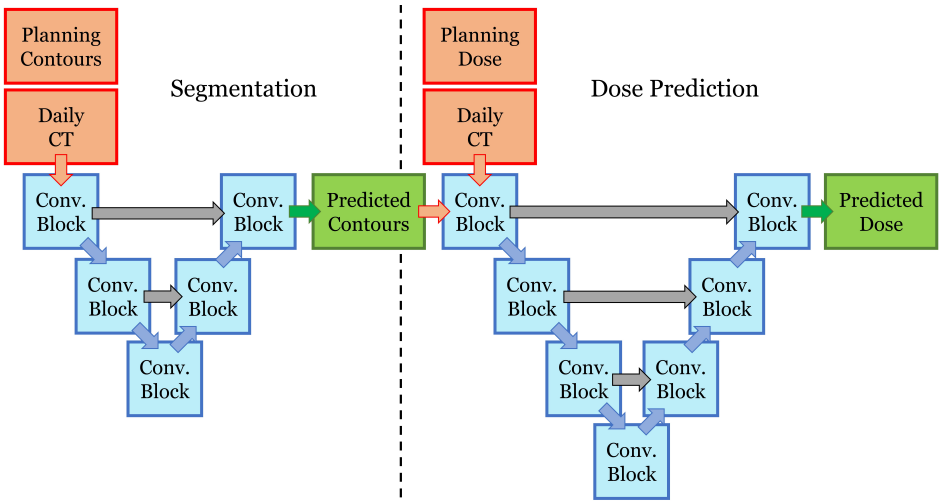


Figure 3.9: *Overview of the w-net. The segmentation network (left) predicts contours for the daily CT scan. These contours are used as input for the dose prediction (right). The networks are trained separately before they are trained together.*

4 Results & Discussion

This chapter will discuss the results obtained for the contour prediction and the dose prediction. First, Section 4.1 will cover the quality of the created IMPT treatment plans used for training and testing of the dose prediction network. Then, Section 4.2 gives results for the accuracy of the predicted contours and discusses these. Finally, in Section 4.3, the predicted dose distributions of the different approaches are compared, and these results are discussed in Section 4.4.

4.1 Quality of the Treatment Plans

The treatment plans aim to deliver 74 Gy to the prostate and 55 Gy to the seminal vesicles, while keeping the dose in the rectum and bladder as low as possible. For clinical plans, the criterion to prevent underdosing is that at least 98% of the target volumes should receive at least 95% of the prescribed dose ($V_{95\%} \geq 98\%$). The criterion to prevent overdosing is that no more than 2% of the target volumes should receive more than 107% of the prescribed dose ($V_{107\%} \leq 2\%$) [57]. In Figure 4.1 (a, b), histograms of the achieved $V_{95\%}$ and $V_{107\%}$ are given for the created treatment plans. For both the prostate and the seminal vesicles, $V_{95\%} < 98\%$ and $V_{107\%} > 2\%$ for many treatment plans. This means that too much of the target volume receives insufficient dose, while also too much of the target volume is overdosed. In Table 4.1, the average target volume percentages that receive at least 95%, 107% and 110% of the prescribed dose are given. This table shows that also, on average, the plans do not meet the criteria described earlier. So, we can conclude from this data that most of these treatment plans are not clinically acceptable, given the criteria described earlier.

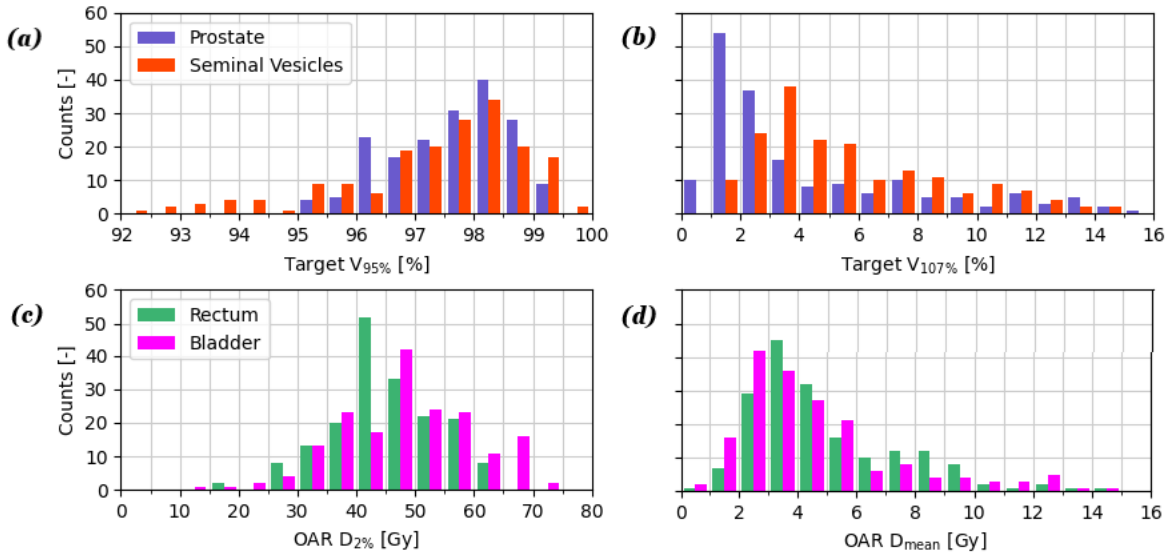


Figure 4.1: *Histograms indicating the performance of each created treatment plan in terms of target coverage and OAR dose. Each plot shows a histogram of a specific DVH point, such as* ***(a)*** $V_{95\%}$ *for the target volumes,* ***(b)*** $V_{107\%}$ *for the target volumes,* ***(c)*** $D_{2\%}$ *for the OARs and* ***(d)*** D_{mean} *for the OARs.*

Since MatRad is not meant for clinical purposes, this result might be expected. Even when the only constraint on the dose distribution was to not underdose the prostate, sometimes $V_{95\%}$ was still below 98%. This could be due to the CT scan having a smaller resolution than the dose grid used for optimization or the optimizer not converging properly[1]. By changing the minimum and maximum doses allowed in the target volumes, the plans would either overdose the targets, underdose the targets or not converge. By tweaking

[1] Admittedly this could also caused due to inexperience with MatRad.

Table 4.1: *Mean (μ) and standard deviation (σ) for volume percentages of the target volumes receiving a certain percent of the prescribed dose and for DVH points indicating the dose in the OARs.*

Target	$V_{95\%}$ (%) $\mu \pm \sigma$	$V_{107\%}$ (%) $\mu \pm \sigma$	$V_{110\%}$ (%) $\mu \pm \sigma$
Prostate	97.6 ± 0.9	4.4 ± 4.1	0.8 ± 0.6
Seminal Vesicles	97.2 ± 1.7	5.8 ± 3.5	3.1 ± 2.5

OAR	$V_{60\,Gy}$ (%) $\mu \pm \sigma$	$V_{45\,Gy}$ (%) $\mu \pm \sigma$	D_{mean} (Gy) $\mu \pm \sigma$	D_2 (Gy) $\mu \pm \sigma$
Rectum	0.6 ± 0.5	2.8 ± 2.3	4.9 ± 2.5	45.2 ± 8.8
Bladder	1.2 ± 1.0	3.1 ± 2.3	4.6 ± 2.8	48.4 ± 11.2

the objectives and constraints for each individual scan, an optimal trade-off between these options could be found, but since there are 179 CT scans in total, it was not feasible to do this for every plan. All the plans for which the prostate $V_{95\%} < 95\%$ were re-planned to guarantee a minimum tumor coverage.

For the rectum and bladder dose, the histograms in Figure 4.1 (c, d) show the $D_{2\%}$ and D_{mean} achieved for the created treatment plans. For the bladder, $D_{2\%} > 60$ Gy in quite a few cases, often when the bladder and prostate were close. This could not be prevented without seriously underdosing the prostate. A few DVH points indicating the dose in the rectum and bladder are given in Table 4.1. Overall the doses in the rectum and bladder are considered sufficiently low.

In Figure 4.2, the DVH for an average performing treatment plan is shown. This plan is chosen since both the prostate's V_{95} and V_{107} are close to the treatment plan average. The target dose, as well as 97% and 107% of the target dose, are indicated for both target structures. For both targets, 95% of the volume receives at least the target dose and only 2.5% of the volume receives less than 97% of the target dose. The volume percentage that receives more than 107% of the target dose is around 5% for both structures. This example shows that although the treatment plans are not of clinical quality, the dose in the target volumes is pretty homogeneous and the dose in the OARs is relatively low.

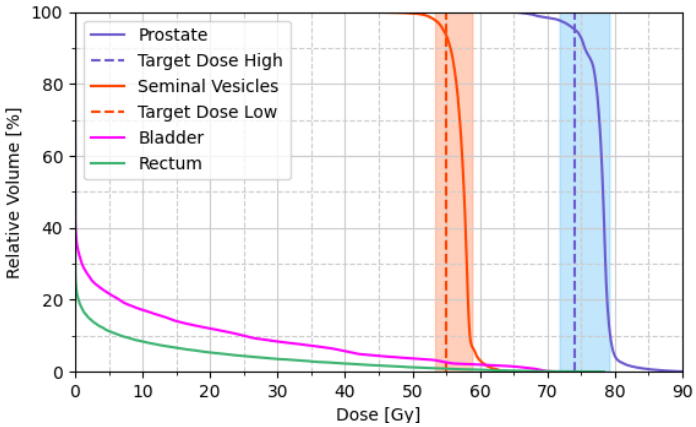


Figure 4.2: *DVH for an average performing treatment plan. The dotted line indicates the target dose for the prostate and the seminal vesicles. The shaded areas indicate the dose between 97% and 107% of the target dose.*

The coverage of the target volumes is not as good as previously achieved with IMPT plans and is certainly not of clinical quality. If this is desired, more effort should be put into optimizing each plan individually, but this was now only done for the worst-performing plans because of the large number of scans. Also, using another treatment planning system is advised if treatment plans of outstanding quality are required since insufficient target coverage also occurred when target coverage was the only objective given to MatRad. Nevertheless, the treatment plans were deemed sufficiently homogeneous to serve as ground truth treatment plans for optimal dose prediction, as this is the actual goal.

4.2 Predicted Contours

The average and median MSD and DSC between the predicted contours and manual contours for the test set are given in Table 4.2 and Table 4.3. Results for the 95% HD are given in the Appendix A. Results are compared against results from literature that use the same method and dataset [6]. The results are also compared against another deep learning method, namely JRS-GAN [58]. This method is similar to the registration network; only it uses a discriminator network for giving feedback on the warped images and contours. If we compare the segmentation network on the 1st row with the registration network on the 3rd row, we see that the segmentation network predicts the contours better for every structure. The segmentation network performs better when there is a lot of contrast between the structure and the background, as is the case with the bladder and the prostate. The registration network struggles with structures that tend to deform between visits, such as the bladder and the rectum. Since the structures differ too much, it fails to transform the contour from the moving image to the correct contour on the fixed image. The cross-stitch network on the 5th row does not seem to improve significantly over the single-task segmentation and registration networks. This is in contrast to the original research, where combining both tasks in a cross-stitch network improved the results over the single-task networks. Contour prediction together with dose prediction is not included in this section as it did not improve upon the single-task networks, and their main goal was to improve the dose prediction.

We now compare our results with the results obtained by Beljaards et al. [6]. Our segmentation network performs significantly better for all the structures. This is probably because the planning contours were given as input to the network, which was not the case in the original research[2]. The segmentation network can probably benefit from prior knowledge about the shape and location of the structures. The registration network performs worse than the previously implemented registration network, both for the single-task network and for the registration path of the cross-stitch network. A convincing reason why this is the case has not been found. The segmentation path of the cross-stitch network performs similarly to the previously achieved results. Although this could be the case because the single-task segmentation network performed better, due to the planning contours that were used as additional input.

Table 4.2: *MSD (mm) values for the different approaches. The best result for each structure is made bold. Results from literature are highlighted in gray.*

		Prostate		Seminal vesicles		Rectum		Bladder	
	Output Path	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median
Segmentation		1.05 ± 0.4	1.03	1.66 ± 1.2	1.27	1.79 ± 0.8	1.57	$\mathbf{0.84 \pm 0.3}$	0.79
Segmentation [6]		1.49 ± 0.3	1.49	2.50 ± 2.6	2.09	3.39 ± 2.2	2.73	1.60 ± 1.1	1.13
Registration		1.40 ± 0.7	1.17	2.22 ± 1.9	1.53	2.66 ± 1.2	2.38	2.55 ± 1.2	2.49
Registration [6]		1.20 ± 0.4	1.13	1.35 ± 0.7	1.16	2.08 ± 1.0	1.82	2.63 ± 2.3	1.90
Cross-Stitch	*Segmentation*	$\mathbf{1.03 \pm 0.4}$	**0.98**	1.52 ± 0.7	1.20	2.05 ± 1.0	1.83	$\mathbf{0.84 \pm 0.3}$	**0.70**
	Registration	1.30 ± 0.6	1.12	2.22 ± 2.0	1.67	2.70 ± 1.3	2.37	2.48 ± 1.2	2.43
Cross-Stitch	*Segmentation*	1.06 ± 0.3	0.99	$\mathbf{1.27 \pm 0.4}$	1.15	$\mathbf{1.76 \pm 0.8}$	**1.47**	0.91 ± 0.4	0.82
[6]	*Registration*	1.10 ± 0.3	1.06	1.30 ± 0.6	**1.13**	2.00 ± 1.0	1.75	2.45 ± 2.1	1.81
JRS-GAN [58]		1.27 ± 0.3	1.25	1.47 ± 0.5	1.32	2.03 ± 0.6	1.85	1.75 ± 1.0	1.26

Table 4.3: *DSC values for the different approaches. The best result for each structure is made bold. Results from literature are highlighted in gray.*

		Prostate		Seminal vesicles		Rectum		Bladder	
	Output Path	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median
Segmentation		0.87 ± 0.04	0.87	0.64 ± 0.15	0.68	$\mathbf{0.86 \pm 0.05}$	0.87	$\mathbf{0.94 \pm 0.02}$	0.94
Segmentation [6]		0.84 ± 0.03	0.84	0.60 ± 0.14	0.62	0.75 ± 0.10	0.77	0.90 ± 0.07	0.93
Registration		0.82 ± 0.07	0.84	0.52 ± 0.23	0.54	0.77 ± 0.07	0.79	0.84 ± 0.08	0.86
Registration [6]		0.87 ± 0.04	0.87	0.67 ± 0.15	0.72	0.83 ± 0.06	0.84	0.87 ± 0.08	0.91
Cross-Stitch	*Segmentation*	0.87 ± 0.04	0.87	0.63 ± 0.15	0.69	0.85 ± 0.05	0.86	$\mathbf{0.94 \pm 0.02}$	**0.95**
	Registration	0.82 ± 0.07	0.84	0.52 ± 0.22	0.53	0.77 ± 0.08	0.78	0.85 ± 0.07	0.86
Cross-Stitch	*Segmentation*	$\mathbf{0.88 \pm 0.04}$	**0.88**	$\mathbf{0.70 \pm 0.11}$	**0.74**	$\mathbf{0.86 \pm 0.05}$	**0.88**	$\mathbf{0.94 \pm 0.02}$	**0.95**
[6]	*Registration*	0.87 ± 0.03	**0.88**	0.68 ± 0.15	0.73	0.84 ± 0.05	0.85	0.88 ± 0.08	0.91
JRS-GAN [58]		0.86 ± 0.04	0.87	0.61 ± 0.20	0.67	0.82 ± 0.06	0.83	0.88 ± 0.08	0.92

[2]The paper did mention that adding planning contours as input improved the segmentation network.

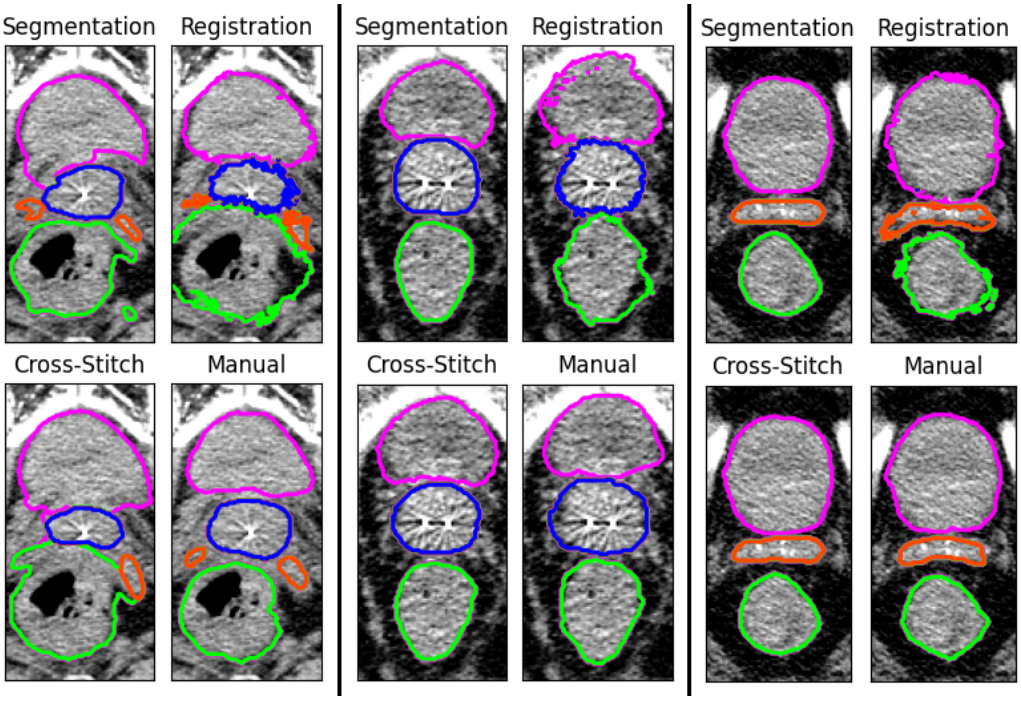


Figure 4.3: *Examples of predicted contours for the segmentation, registration and cross-stitch networks. From left to right, the shown contours are the 1st, 2nd and 3rd quantile lowest prostate MSD of the cross-stitch network.*

Figure 4.3 contains examples of predicted contours for the different networks. The output of the segmentation, registration and cross-stitch network are compared with the manual contours used to train the networks. In the middle and the right image, the segmentation and cross-stitch contours are very similar to the manual contours. On the left image, a slice is chosen that lies on the boundary between different structures, and it is apparent that the networks have more trouble defining the contours there correctly. Although the contours of the registration network are generally located correctly, the boundaries are pretty wobbly, something which is not the case in other studies implementing registration networks. This could indicate that the registration network does not entirely work as it should.

The single-task segmentation network and the segmentation path of the cross-stitch network predict contours of similar quality as state-of-the-art contour definition methods. The registration network predicted worse contours than the segmentation network for every structure and did not achieve results comparable with previous studies. The cross-stitch network did not have the beneficial effect that has been observed in previous research. This can be because the segmentation network performed very well already without the cross-stitch units or because the registration network did not work correctly.

4.3 Predicted Dose Distributions

4.3.1 Gamma Analysis of the Predicted Dose Distribution

For the different dose prediction methods described in Section 3.5 and Section 3.6, the average 2%/2mm GPR of the dose predictions made for the test set is given in Table 4.4. The 3%/3mm GPR of the dose predictions is given in Table A.2. The GPR is highest in the prostate for almost all approaches. This makes sense since the treatment plan objectives create a homogeneous dose distribution in the target volumes and the loss function of the dose prediction networks is weighted higher for the prostate. The average GPR in the entire dose distribution is always lower than the GPR in the different structures. This is because the dose prediction network has a hard time predicting the ray effect of the proton beams that occurs to the left and right of the structures. The dose prediction network is also inclined to predict the dose well inside structures because of the sampling of patches and the weights of the loss function.

Table 4.4: *The average and standard deviation of the 2%/2mm GPR (%) of the test set for the different dose prediction approaches. The GPR is given for the entire dose distribution and for each structure.*

Model		Dose	Prostate	Seminal vesicles	Rectum	Bladder
Manual Contours		79.0 ± 2.5	93.4 ± 3.2	88.6 ± 7.1	95.9 ± 5.0	94.7 ± 5.0
Predicted Contours		65.4 ± 4.9	83.8 ± 6.9	71.1 ± 15.5	82.6 ± 13.6	80.7 ± 12.3
No Contours		46.5 ± 5.7	83.6 ± 6.5	72.7 ± 14.9	65.3 ± 18.2	65.1 ± 18.0
DVF on Planning Dose		55.7 ± 9.4	84.1 ± 6.1	72.1 ± 18.3	59.5 ± 19.6	69.8 ± 18.4
Cross-Stitch	*Segmentation & Dose*	50.4 ± 7.4	82.5 ± 6.9	67.7 ± 15.4	67.6 ± 17.9	65.8 ± 21.3
	Registration & Dose	50.3 ± 7.2	73.3 ± 9.9	67.6 ± 19.4	60.2 ± 19.1	70.0 ± 17.8
W-Net	*Before Training Together*	52.3 ± 6.7	81.4 ± 8.1	75.2 ± 14.0	77.7 ± 21.5	78.4 ± 15.3
	After Training Together	51.5 ± 7.2	82.4 ± 8.9	73.8 ± 14.8	75.7 ± 21.3	73.9 ± 16.3

If we compare the different approaches, it is clear that the manual contour based dose prediction network performs better than all other approaches. This is expected since the ground-truth dose distributions are based on the manual contours. When we look at the results for the model where manual contours are used for training and predicted contours are used for testing, we see that the predicted dose distributions fail the gamma test for up to three times as many points in some structures. It is clear that accurately knowing the locations of the structures used for treatment planning is really important for predicting the resulting dose distribution. W-net consists of a network that predicts the contours for a CT scan and a dose prediction network trained and tested using those predicted contours. This performs worse than the network that used manual contours for training and predicted contours for testing. Especially the predicted dose distribution outside the structures is worse, probably because the lateral projections of the target volumes are not given as input to the dose prediction network of the w-net. Both the network trained without contours as input and the cross-stitch networks fail to predict dose distributions better than a transformed planning dose distribution. These networks that do not use contours as input fail to predict any optimality. Adding the

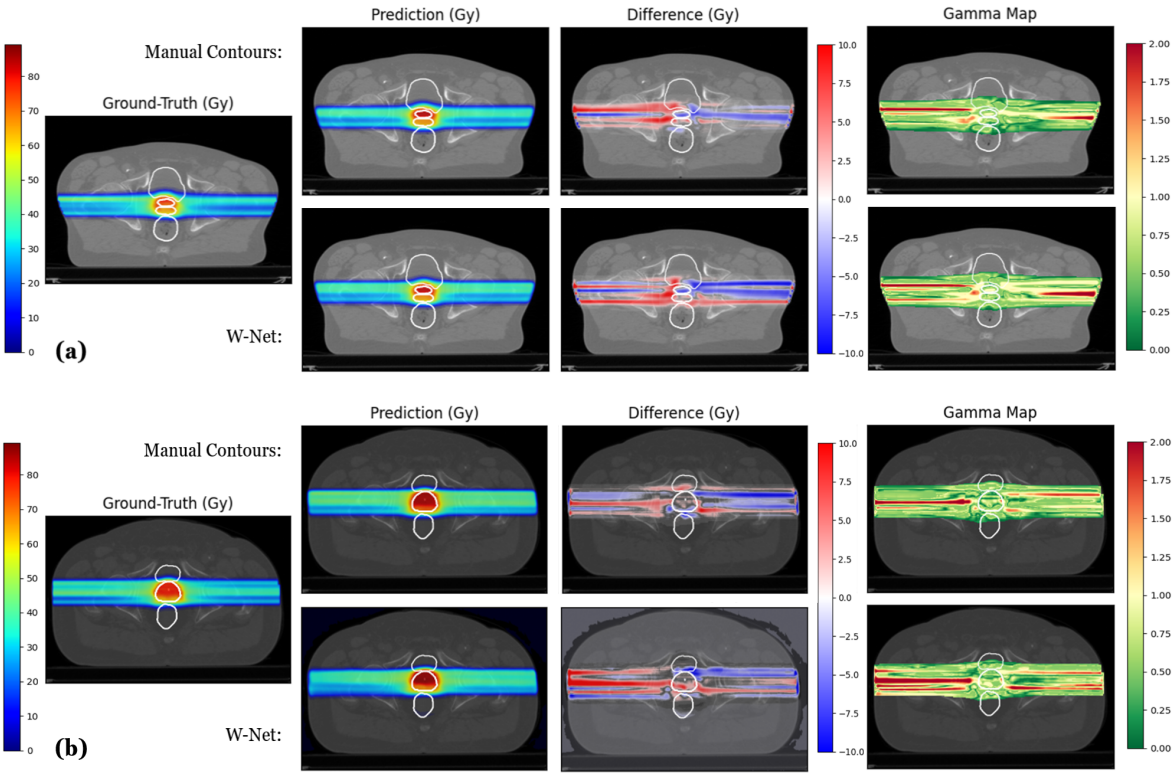


Figure 4.4: *A comparison between predicted dose distributions from the manual contour based network and the w-net. From left to right, the ground-truth dose distribution, the predicted dose distribution, the difference between the two (D_{true}-D_{pred}) and a 2%/2mm gamma map of the prediction are shown.* ***(a):*** *A good prediction (3^{rd} quartile of prostate GPR for the w-net).* ***(b):*** *A bad prediction (1^{st} quartile of prostate GPR for the w-net).*

segmentation network or the registration network to the dose prediction network via cross-stitch units does not seem to benefit the dose prediction network. Also, training the contour prediction network and the dose prediction network of the w-net together does not seem to give any benefit.

Predicted dose distributions from the manual contour based network are compared to those of the w-net in Figure 4.4. Both the difference between the prediction and the ground-truth dose distribution and a 2%/2mm gamma map, which gives the gamma index of each of the predicted points, is given. Gamma values between 0 and 1 indicate that the prediction meets the criteria, and gamma values above 1 indicate that the prediction does not meet the criteria. These images show that the dose prediction networks have difficulty correctly predicting the ray effect of the proton beams, which appears to the left and right of the structures. This is partly because the patch-based network has a hard time predicting the dose distribution in those areas due to the lack of structures. However, there is also something inherently tricky about predicting these rays accurately since their locations in a treatment plan are not self-evident. Around the structures, the predicted dose is close to the ground-truth dose, and as expected, most of the prediction there meets the 2%/2mm criteria. The dose prediction from w-net around the structures sometimes suffers from inaccurately predicted contours, although this is not really visible from these images.

4.3.2 Dose-Volume Histogram Comparisons

For the different dose prediction approaches, the average ΔDVH between the predicted and ground-truth dose distributions of the test set are given in Table 4.5 and Table 4.6. ΔDVH is calculated using Equation 3.7 for D_{mean}, D_{95} and $V_{95\%}$ of the target volumes. Here D_{mean} is the average dose, D_{95} is the 5^{th} percentile lowest dose delivered to a structure and $V_{95\%}$ is the volume percentage receiving at least 95% of the prescribed dose. For the organs at risk, averages of the absolute dose differences are given for ΔDVH instead of percentage differences since the dose in these structures is low. Here differences in D_{mean} and D_2 are given, where D_2 is the 2^{nd} percentile highest dose in a structure.

Just as with the gamma analysis, the manual contour based dose predictions are the most accurate for predicting DVH points. This method outperforms the other approaches, especially for structures where the contours are hard to predict, such as the seminal vesicles. Although the predicted contour based approaches can predict D_{mean} quite well, they fail to predict the D_{95}, $V_{95\%}$ and D_2. Since knowing these DVH points is essential for knowing the quality of a treatment plan, any method besides the manual contour based one does not predict dose distributions accurate enough to be helpful in the treatment planning process.

Table 4.5: *ΔDVH (%) for the different dose prediction approaches. The percentage difference for D_{mean}, D_{95} and $V_{95\%}$ of the target volumes is given.*

Model		ΔDVH Prostate (%)			ΔDVH Seminal Vesicles (%)		
		D_{mean}	D_{95}	$V_{95\%}$	D_{mean}	D_{95}	$V_{95\%}$
Manual Contours		0.5 ± 0.4	1.4 ± 1.0	1.0 ± 0.8	1.1 ± 0.8	2.9 ± 3.0	3.1 ± 3.8
Predicted Contours		1.4 ± 1.1	8.2 ± 5.3	6.7 ± 4.6	4.4 ± 4.3	17.6 ± 15.0	19.6 ± 15.2
No Contours		0.9 ± 0.8	4.4 ± 5.1	2.7 ± 2.9	3.2 ± 4.2	16.5 ± 18.4	16.5 ± 18.4
DVF on Planning Dose		1.8 ± 1.5	9.9 ± 8.6	6.9 ± 4.1	6.0 ± 9.2	21.6 ± 24.3	17.7 ± 19.2
Cross-Stitch	*Segmentation & Dose*	1.2 ± 1.1	6.0 ± 5.9	3.9 ± 3.7	3.2 ± 3.5	13.1 ± 14.8	12.6 ± 14.1
	Registration & Dose	1.2 ± 1.0	8.9 ± 6.5	6.4 ± 4.5	4.0 ± 5.3	15.3 ± 17.1	15.2 ± 18.9
W-Net	*Before Training Together*	1.9 ± 1.2	8.4 ± 5.1	6.7 ± 4.3	4.0 ± 3.7	16.0 ± 13.8	16.5 ± 14.2
	After Training Together	1.5 ± 1.5	9.2 ± 6.3	7.3 ± 5.2	4.3 ± 3.9	17.2 ± 13.4	19.3 ± 14.9

Table 4.6: *ΔDVH (Gy) for the different dose prediction approaches. The dose difference for D_{mean} and D_2 of the organs at risk is given.*

Model		ΔDVH Rectum (Gy)		ΔDVH Bladder (Gy)	
		D_{mean}	D_2	D_{mean}	D_2
Manual Contours		0.49 ± 0.3	2.93 ± 2.6	0.46 ± 0.3	4.22 ± 2.6
Predicted Contours		0.92 ± 0.7	7.34 ± 5.7	0.70 ± 0.6	5.80 ± 4.7
No Contours		2.14 ± 1.8	17.15 ± 8.0	2.06 ± 1.3	12.20 ± 8.9
DVF on Planning Dose		2.74 ± 2.5	16.28 ± 11.2	1.78 ± 1.5	11.84 ± 8.2
Cross-Stitch	*Segmentation & Dose*	1.67 ± 1.1	11.44 ± 8.6	2.09 ± 1.7	14.88 ± 10.5
	Registration & Dose	2.40 ± 2.2	7.72 ± 4.9	1.71 ± 1.4	9.15 ± 6.4
W-Net	*Before Training Together*	1.35 ± 1.4	8.69 ± 6.5	0.98 ± 0.9	7.59 ± 6.2
	After Training Together	1.26 ± 1.2	8.22 ± 6.8	1.12 ± 1.0	9.77 ± 7.0

In Figure 4.5, the DVH of a test patient is plotted for the ground-truth dose distribution and the dose predictions from the manual contour based network and the w-net. The DVH of the manual contour based dose prediction and the w-net dose prediction are both very similar to the DVH of the ground-truth dose distribution. This indicates that although the predicted dose distributions are not very accurate outside the structures used for treatment planning, the dose prediction inside the structures is acceptable. Only the w-net predicts too low a dose for the lowest 10% of the target volumes, which could be because the predicted contours do not align with the manual contours.

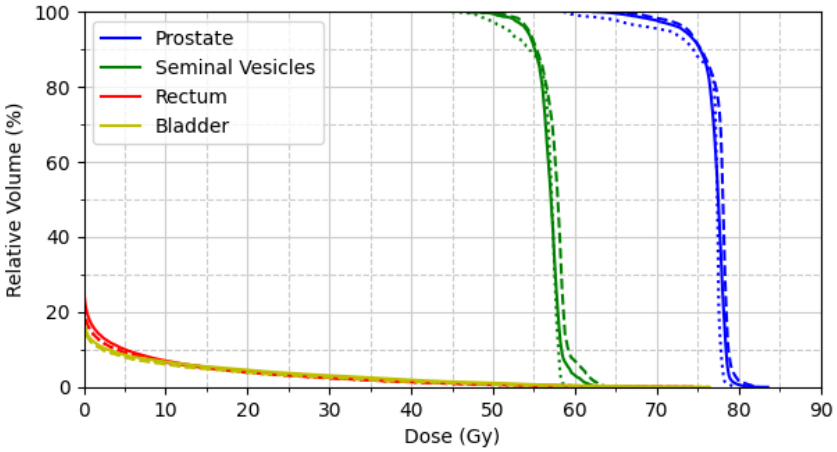


Figure 4.5: *DVH plot comparing predicted dose distributions for a test patient. Dose distributions from the ground-truth (solid line), manual contour based network (dashed line) and w-net (dotted line) are compared.*

4.4 Discussion of the Predicted Dose Distributions

Both the gamma analysis and the DVH comparisons indicate that dose prediction approaches that do not use manual contours do not perform as well as an approach that does use the manual contours. When no contours are available, it is best first to predict the contours and then use those contours to predict the optimal dose distribution. Methods that use no contours for dose prediction or perform dose prediction and contour prediction simultaneously do not result in predicted dose distributions better than the dose distribution of the planning CT transformed to the daily CT scan. This indicates that these approaches can not predict an optimal dose distribution, merely a dose distribution that is somewhat similar to the optimal dose distribution. Table 4.7 compares the average ΔDVH obtained on the test set for D_{mean} and D_{95} of the PTV, and D_{mean} of the rectum with two other dose prediction studies for prostate cancer. The manual contour based network predicts D_{mean} in the PTV better than other results from literature, but D_{95} similar or worse. The predicted contour based network predicts D_{mean} similar to other results, but it predicts D_{95} significantly worse. It has to be noted here that the PTV used in this study is simply the structure volume of the prostate, in contrast to the other studies where a PTV was used that contains target margins. Furthermore, the percentage difference in D_{mean} for the rectum is way worse than the results from literature. This is probably because the average dose in the rectum is very low in these treatment plans (4.9 Gy), which causes the error percentage to be high. The difference between prediction and ground-truth D_{mean} for the rectum is within 0.5 Gy, which is smaller than reported in other studies.

Table 4.7: *Comparison of obtained ΔDVH (%) with results from literature. ΔDVH is given for D_{mean} and D_{95} of the prostate and for D_{mean} of the rectum.*

Model	ΔDVH Prostate (%)		ΔDVH Rectum (%)
	D_{mean}	D_{95}	D_{mean}
Manual Contours	0.45 ± 0.36	1.40 ± 1.02	9.91 ± 6.22
Predicted Contours	1.38 ± 1.07	8.21 ± 5.28	18.59 ± 13.48
Meerbothe, 2020 [38]	0.99 ± 0.67	1.42 ± 1.26	3.39 ± 2.47
Kandalan et al., 2020 [37]	1.60 ± 0.55	0.40 ± 0.60	1.25 ± 0.95

4.4.1 Usefulness for Adaptive Proton Therapy

The difficulty with predicting an optimal dose distribution for a daily CT scan in APT is that the contours of the structures are not known beforehand. It was clear that all attempts to find a substitute for the manual contours did not produce dose predictions nearly as good as the manual contour based approach. Especially errors in DVH statistics like D_{95} and $V_{95\%}$ make that these methods are probably not helpful in the treatment planning process of APT since these statistics are often used to determine the quality of a treatment plan. The best option is to predict the contours based on the CT scan and use those predicted contours for dose prediction. Since the predicted contours in this study are of high quality and still fail to predict dose distribution similar to those from the manual contours, it can be concluded that the quality of a predicted dose distribution depends heavily on accurate knowledge of the structure locations.

Dose distributions from IMPT treatment plans usually have higher dose gradients than those resulting from photon radiotherapy. This is because the depth dose distribution of protons allows for dose distributions that are more conformal to the tumor. Dose distributions with higher gradients are more difficult to predict since slight spatial deviations lead to higher dose differences. Given this fact, it might be more challenging to predict IMPT dose distributions than dose distributions resulting from photons. However, compared to networks predicting VMAT dose distributions, the manual contour based network obtained similar results.

4.4.2 Effect of Multi-Task Learning

Combining the dose prediction network with a segmentation or registration network via cross-stitch units did not benefit the dose prediction network. The main reason for this is the poor performance of a dose prediction network that does not use structure masks as input. Dose prediction is a complicated task that requires many different resolution feature maps of the structures used for the treatment plan. This missing information in the dose prediction network could not be substituted by information from the feature maps used in the segmentation or registration network. Cross-stitch networks should be used to improve competence in tasks that can already be performed correctly to a certain extent.

The w-net did not benefit from training the segmentation and dose prediction network together. A problem is that a small change in a predicted contour only slightly affects the predicted dose distribution. Therefore, any effect the loss of the dose prediction has on the parameters of the segmentation network might be negligible. Probably more could have been done with the way the loss of the network is calculated. Now the losses of both networks are given a weight and added up. Determining which loss should be valued more using dynamic weights could be an interesting way of improving this. A plot of the training and validation loss of the w-net is given in Figure A.1. As can be seen, the validation loss does not improve much after 100 epochs. Even unfreezing the parameters of the segmentation network after 400 epochs and training both networks together does not have much effect on the loss. One of the reasons the validation loss is higher than the training loss is that the network lacks information to generalize well to new examples. A possibly interesting way of increasing the information available for dose prediction in the w-net is adding feature maps from the segmentation network back to the dose prediction, similar to how this is done in the cross-stitch network. The information from the different feature maps in the segmentation network might be helpful in addition to the predicted contours. However, this will not eliminate the error arising from the difference between the predicted and manual contours.

4.4.3 Limitations of the Study

One of the critical problems in most deep learning methods used for medical image applications is the lack of training data. In this study, 129 daily CT scans and treatment plans were available for training. Since these are large 3D images, this should be sufficient for training the network. However, the problem is that this training data is from only 14 different patients. This makes it hard for the network to generalize the model to anatomies of new patients. Augmentation of data in the form of rotations has been tried, but since this changes the angles of the proton beams, this had no positive effect on the dose prediction network. A form of data augmentation is done by randomly selecting patches from the image, which drastically increases the number of different inputs available to the network.

Another advantage of patch-based networks is the memory saved by using smaller inputs. Furthermore, a patch-based network is not dependent on the geometrical location of the structures, which makes it less dependent on the homogeneity of the input. Despite these advantages, patch-based networks also have some disadvantages. Most importantly, the lack of global geometrical information about the patch makes

it difficult for the network to say something useful about what the dose distribution in the patch should be. Especially in this study, where the contours of only four structures are used, it is difficult to predict a dose distribution in patches that contain no structure information. In the manual contour based network, this was solved by adding the lateral projections of the target volumes as input to the network. These projections give some information regarding the position of the proton beams, but it is not sufficient. This is why dose predictions outside the defined structures are generally quite off. Recently, other studies from literature have utilized patch-based dose prediction networks [34, 52], but they did not report the same difficulties. This probably is because those studies involved head and neck cancer patients where there are many more structures present in the image.

The created treatment plans have quite a bit of variation in the resulting dose distributions. Although some variability due to the different anatomies is expected, more homogeneous treatment plans are definitely possible. In addition, only four structures were used to create the treatment plans due to a lack of manual contours. Adding more structures, such as the femoral heads, could lead to better treatment plans. This could lead to better training of the network and better generalization to the test data, which could improve the manual contour based dose prediction.

4.4.4 Suggestions for Future Research

If similar research aims at leveraging multi-task learning for predicting optimal dose distributions and contours together, the dose prediction should probably be based on predicted contours. This is because the dose prediction is highly dependent on the structure information. Some interesting directions to explore are using multiple resolution feature maps from the segmentation network as input for the dose prediction network and using dynamic loss weights to optimally combine the losses of both networks. If further work is carried out with the code used in this research, then the registration network should be checked since the resulting contours had not the same accuracy as in previous work.

This research found that calculating the loss on multiple resolution levels had a positive effect compared to calculating the loss only at the end of the network. This feature is not utilized in most deep CNN. It could be interesting to investigate the effect of multiple losses and how they should be weighted to obtain an optimal result. Another minor finding concerns inference with a patch-based network. Being free to choose the patch for which to predict a dose distribution makes it so multiple predictions can be made for a single voxel. We found that taking the average of 8 predictions improves the dose prediction slightly. Of course, increasing the number of predictions per voxel increases the inference time. However, if there is an optimal way of utilizing multiple predictions, this is something worth exploring.

A promising direction for future research is further investigating the automation of treatment planning. Knowing what dose distribution is achievable based on historic treatment plans gives an idea of what quality should be strived for, but it does not tell what treatment plan results in such a dose distribution. The next step is a network capable of predicting beam angles and intensities of an optimal treatment plan. The treatment planning system can then use these parameters to quickly optimize the treatment plan without much assistance from a clinician. Some exciting research has been done on this problem in the last couple of years [38, 59], but the high number of parameters that must be predicted makes it a complex problem. Nevertheless, with the emergence of new knowledge and the increasing capabilities of neural networks, in the near future, it might be possible to automatically determine an optimal treatment plan based on the patient's anatomy.

5 Conclusion

This research aimed to determine an approach that can predict the dose distribution of an optimal treatment plan without using manual contours of the structures. Multiple methods, such as using no contours, using predicted contours or predicting contours and dose distributions simultaneously, have been tried. We found that basing the dose prediction on predicted contours gives the most accurate results. However, dose predictions based on predicted contours are significantly worse than those based on manual contours, having a test set average 2%/2mm GPR of 83.8%±6.9% compared to 93.4%±3.2%. Furthermore, the errors for predicting DVH statistics, such as D_{95} and $V_{95\%}$, range from 7% to 20%, which makes these predicted dose distributions not accurate enough to be helpful for treatment planning. This shows that dose prediction relies heavily on exact knowledge of the structure locations, considering the predicted contours have similar quality as those from state-of-the-art methods.

The second goal of this research was to investigate the potential benefit of performing dose prediction and contour prediction together in a multi-task network. However, no successful method for combining contour prediction and dose prediction in a multi-task network has been found. Firstly, combining a dose prediction network with a segmentation or registration network via cross-stitch units had no advantageous effects on the predicted dose distributions, mainly because dose predictions not based on structure masks were too bad for it to have any effect. Secondly, a w-net that performs contour definition and dose prediction consecutively did not benefit from training the segmentation and dose prediction networks together. A possible improvement to the w-net could be how the losses of both tasks are combined. The main conclusion is that multi-task learning can only benefit related tasks if they can already be performed on their own to a certain extent. It is not a substitute for missing information required to perform the task.

The developed approaches for predicting optimal dose distributions for daily CT scans do not improve the APT workflow. Further advancements in fast contour definition methods and accurate dose prediction methods are necessary for this to be the case. A possible improvement would be to also determine the machine parameters leading to an optimal dose distribution, in addition to the contour definition and dose prediction. This would allow a network to determine an optimal treatment plan based on a CT scan. Such a predicted plan can drastically speed up the treatment planning process, making online treatment plan adaptation feasible. This adaptation would allow for more optimal dose delivery, which in the end, will result in better patient outcomes.

A Additional Results

Table A.1: *95% HD values for the different approaches. The best result for each structure is made bold. Results from literature are highlighted in gray.*

		Prostate		Seminal vesicles		Rectum		Bladder	
	Output Path	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median
Segmentation		3.4 ± 1.1	3.3	6.8 ± 7.2	4.5	**8.5 ± 4.1**	7.4	3.5 ± 2.0	2.5
Segmentation [6]		4.4 ± 1.0	4.4	8.6 ± 8.6	7.3	16.5 ± 11.0	13.3	6.9 ± 6.6	4.0
Registration		5.2 ± 3.8	4.5	7.0 ± 4.6	5.9	10.0 ± 4.7	8.7	10.9 ± 4.3	11.8
Registration [6]		5.5 ± 4.5	4.0	5.6 ± 4.1	4.3	11.0 ± 6.4	9.4	15.7 ± 9.6	12.1
Cross-Stitch	*Segmentation*	3.4 ± 1.1	3.3	5.4 ± 2.7	4.0	9.6 ± 4.9	8.0	3.4 ± 1.8	2.6
	Registration	4.3 ± 1.8	4.0	6.7 ± 4.0	6.0	10.2 ± 5.0	9.9	10.7 ± 4.5	11.4
Cross-Stitch	*Segmentation*	**3.0 ± 1.0**	**3.0**	**4.3 ± 1.7**	3.9	9.5 ± 6.2	**7.2**	**3.3 ± 2.9**	**2.3**
[6]	*Registration*	3.2 ± 0.9	**3.0**	4.5 ± 3.3	**3.6**	9.8 ± 6.3	8.6	12.2 ± 10.1	9.7
JRS-GAN [58]		3.4 ± 1.2	**3.0**	5.3 ± 3.0	4.6	10.1 ± 6.1	8.4	11.0 ± 9.6	7.6

Table A.2: *The 3%/3mm GPR (%) for the different dose prediction approaches. The GPR and its uncertainty is given for the entire dose distribution and also for each individual structure.*

Model		Dose	Prostate	Seminal vesicles	Rectum	Bladder
Manual Contours		87.0 ± 2.2	98.2 ± 0.8	96.5 ± 3.7	98.4 ± 3.0	98.7 ± 2.1
Predicted Contours		78.3 ± 4.5	93.0 ± 4.8	83.6 ± 12.8	91.4 ± 9.5	91.0 ± 8.8
No Contours		61.5 ± 6.9	93.4 ± 3.3	85.8 ± 11.7	78.5 ± 17.7	81.4 ± 16.1
DVF on Planning Dose		69.1 ± 9.6	92.2 ± 4.8	82.5 ± 16.4	71.8 ± 17.2	81.6 ± 17.1
Cross-Stitch	*Segmentation & Dose*	62.7 ± 7.7	92.5 ± 4.8	83.1 ± 13.3	79.3 ± 15.5	77.6 ± 19.6
	Registration & Dose	64.1 ± 7.8	89.4 ± 6.2	80.8 ± 18.3	77.8 ± 16.2	83.8 ± 15.8
W-Net	*Before Training Together*	64.6 ± 6.9	92.1 ± 5.0	85.9 ± 11.8	86.6 ± 17.2	88.3 ± 11.9
	After Training Together	63.7 ± 7.3	92.3 ± 5.1	84.8 ± 12.4	86.2 ± 17.7	84.8 ± 12.4

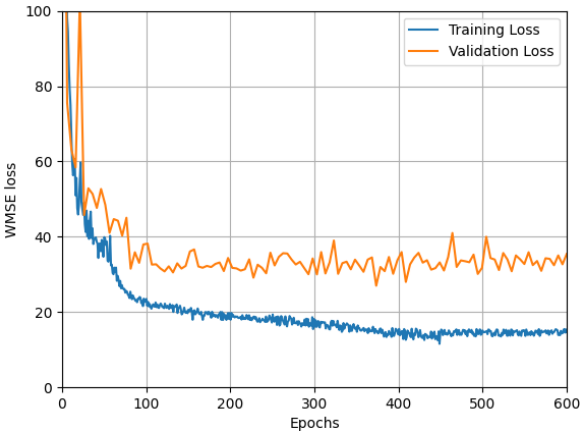


Figure A.1: *Training and validation WMSE loss of the w-net. At 400 epochs the parameters of the segmentation network are unfrozen and image segmentation and dose prediction are trained simultaneously.*

References

[1] H. Paganetti, P. Botas, G. C. Sharp, and B. Winey, "Adaptive proton therapy," *Physics in Medicine & Biology*, vol. 66, 11 2021.

[2] F. Albertini, M. Matter, L. Nenoff, Y. Zhang, and A. Lomax, "Online daily adaptive proton therapy," *Br J Radiol*, vol. 93, 2020.

[3] L. Nenoff, M. Matter, M. Charmillot, S. Krier, K. Uher, D. C. Weber, A. J. Lomax, and F. Albertini, "Adaptronics: Selected papers from the 14th international adaptronic congress," *Phys. Med. Biol*, vol. 66, 2021.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 12 2017.

[5] M. La Macchia, F. Fellin, M. Amichetti, M. Cianchetti, S. Gianolini, V. Paola, A. J. Lomax, and L. Widesott, "Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer," *Radiation Oncology*, vol. 7, 09 2012.

[6] L. Beljaards, M. S. Elmahdy, F. Verbeek, and M. Staring, "A cross-stitch architecture for joint registration and segmentation in adaptive radiotherapy," 2020.

[7] G. Delaney, S. Jacob, C. Featherstone, and M. Barton, "The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines," *Cancer*, vol. 104, pp. 1129–1137, 9 2005.

[8] J. B. Little, *Principal cellular and tissue effects of radiation*. Shelton, CT: B.C. Decker, 2003.

[9] P. P. Connell, S. J. Kron, and R. R. Weichselbaum, "Relevance and irrelevance of dna damage response to radiotherapy," *DNA Repair*, vol. 3, pp. 1245–1251, 8 2004.

[10] S. J. Gardner, J. Kim, and I. J. Chetty, "Modern radiation therapy planning and delivery," *Hematology/Oncology Clinics of North America*, vol. 33, pp. 947–962, 12 2019.

[11] D. Ersahin, I. Doddamane, and D. Cheng, "Targeted radionuclide therapy," *Cancers*, vol. 3, no. 4, pp. 3838–3855, 2011.

[12] W. D. Newhauser and R. Zhang, "The physics of proton therapy," *Physics in Medicine and Biology*, vol. 60, 4 2015.

[13] R. Mohan and D. Grosshans, "Proton therapy – present and future," *Advanced Drug Delivery Reviews*, vol. 109, pp. 26–44, 1 2017.

[14] H. Paganetti, C. Beltran, S. Both, *et al.*, "Roadmap: Proton therapy physics and biology," *Physics in Medicine and Biology*, vol. 66, 3 2021.

[15] S. Breedveld, D. Craft, R. van Haveren, and B. Heijmen, "Multi-criteria optimization and decision-making in radiotherapy," *European Journal of Operational Research*, vol. 277, pp. 1–19, 8 2019.

[16] T. Z. Jagt, S. Breedveld, R. van Haveren, B. J. Heijmen, and M. S. Hoogeman, "Online-adaptive versus robust impt for prostate cancer: How much can we gain?," *Radiotherapy and Oncology*, vol. 151, pp. 228–233, 10 2020.

[17] M. Bobic, A. Lalonde, G. C. Sharp, C. Grassberger, J. M. Verburg, B. A. Winey, A. J. Lomax, and H. Paganetti, "Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy," *Physics in Medicine and Biology*, vol. 66, 3 2021.

[18] H. Tamura, K. Kobashi, K. Nishioka, *et al.*, "Dosimetric advantages of daily adaptive strategy in impt for high-risk prostate cancer," *Journal of applied clinical medical physics*, 2022.

[19] O. L. Green, L. E. Henke, and G. D. Hugo, "Practical clinical workflows for online and offline adaptive radiation therapy," *Seminars in Radiation Oncology*, vol. 29, pp. 219–227, 7 2019.

[20] L. Nenoff, M. Matter, A. G. Jarhall, *et al.*, "Daily adaptive proton therapy: Is it appropriate to use analytical dose calculations for plan adaption?," *International Journal of Radiation Oncology Biology Physics*, vol. 107, pp. 747–755, 7 2020.

[21] T. Z. Jagt, S. Breedveld, R. van Haveren, R. A. Nout, E. Astreinidou, B. J. Heijmen, and M. S. Hoogeman, "Plan-library supported automated replanning for online-adaptive intensity-modulated proton therapy of cervical cancer," *Acta Oncologica*, vol. 58, pp. 1440–1445, 10 2019.

[22] T. Jagt, S. Breedveld, R. van Haveren, B. Heijmen, and M. Hoogeman, "An automated planning strategy for near real-time adaptive proton therapy in prostate cancer," *Phys. Med. Biol.*, vol. 63, July 2018.

[23] B. Cai, O. L. Green, R. Kashani, V. L. Rodriguez, S. Mutic, and D. Yang, "A practical implementation of physics quality assurance for photon adaptive radiotherapy," *Zeitschrift für Medizinische Physik*, vol. 28, no. 3, pp. 211–223, 2018.

[24] A. C. Riegel, J. G. Antone, H. Zhang, P. Jain, J. Raince, A. Rea, A. M. Bergamo, A. Kapur, and L. Potters, "Deformable image registration and interobserver variation in contour propagation for radiation therapy planning," *Journal of Applied Clinical Medical Physics*, vol. 17, 2016.

[25] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351 of *LNCS*, pp. 234–241, Springer, 2015.

[26] S. Hu, L. Zhang, G. Li, M. Liu, D. Fu, and W. Zhang, "Brain deformable registration using global and local label-driven deep regression learning in the first year of life," *IEEE Access*, vol. 8, pp. 25691–25705, 2020.

[27] B. Li, W. Niessen, S. Klein, M. de Groot, A. Ikram, M. Vernooij, and E. Bron, "A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes," 2019.

[28] Z. Xu and M. Niethammer, "Deepatlas: Joint semi-supervised learning of image registration and segmentation," 2019.

[29] O. Nwankwo, H. Mekdash, D. S. K. Sihono, F. Wenz, and G. Glatting, "Knowledge-based radiation therapy (kbrt) treatment planning versus planning by experts: Validation of a kbrt algorithm for prostate cancer treatment planning," *Radiation Oncology*, vol. 10, 5 2015.

[30] K. Nawa, A. Haga, A. Nomoto, R. A. Sarmiento, K. Shiraishi, H. Yamashita, and K. Nakagawa, "Evaluation of a commercial automatic treatment planning system for prostate cancers," *Medical Dosimetry*, vol. 42, pp. 203–209, 2 2017.

[31] S. Momin, Y. Fu, Y. Lei, J. Roper, J. D. Bradley, W. J. Curran, T. Liu, X. Yang, and C. X. Yang, "Knowledge-based radiation treatment planning: A data-driven method survey," *J Appl Clin Med Phys*, vol. 22, 2021.

[32] D. Nguyen, T. Long, X. Jia, W. Lu, X. Gu, Z. Iqbal, and S. Jiang, "Dose prediction with u-net: A feasibility study for predicting dose distributions from contours using deep learning on prostate imrt patients," *ArXiv*, 09 2017.

[33] M. Ma, N. Kovalchuk, M. Buyyounouski, L. Xing, and Y. Yang, "Incorporating dosimetric features into the prediction of 3d vmat dose distributions using deep convolutional neural network," *Phys. Med. Biol*, vol. 64, 2019.

[34] D. Nguyen, X. Jia, D. Sher, M.-H. Lin, Z. Iqbal, H. Liu, and S. Jiang, "3d radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture," *Physics in Medicine & Biology*, vol. 64, mar 2019.

[35] Y. Liu, Z. Chen, J. Wang, X. Wang, B. Qu, L. Ma, W. Zhao, G. Zhang, and S. Xu, "Dose prediction using a three-dimensional convolutional neural network for nasopharyngeal carcinoma with tomotherapy," *Frontiers in Oncology*, vol. 11, 11 2021.

[36] C. Kontaxis, G. H. Bol, J. J. Lagendijk, and B. W. Raaymakers, "Deepdose: Towards a fast dose calculation engine for radiation therapy using deep learning," *Physics in Medicine and Biology*, vol. 65, 4 2020.

[37] R. N. Kandalan, D. Nguyen, N. H. Rezaeian, *et al.*, "Dose prediction with deep learning for prostate cancer radiation therapy: Model adaptation to different treatment planning practices," *Radiotherapy and Oncology*, vol. 153, pp. 228–235, 12 2020.

[38] T. Meerbothe, "A physics guided neural network approach for dose prediction in automated radiation therapy treatment planning," Master's thesis, TU Delft, 2020.

[39] F. Guerreiro, E. Seravalli, G. O. Janssens, J. H. Maduro, A. C. Knopf, J. A. Langendijk, B. W. Raaymakers, and C. Kontaxis, "Deep learning prediction of proton and photon dose distributions for paediatric abdominal tumours," *Radiotherapy and Oncology*, vol. 156, pp. 36–42, 3 2021.

[40] Wikipedia, the free encyclopedia, "Cumulative dose-volume histogram," 2011. [Online; accessed June 6, 2022].

[41] O. Cicek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9901 of *LNCS*, pp. 424–432, Springer, Oct 2016.

[42] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," *CoRR*, 04 2016.

[43] H.-P. Wieser, E. Cisternas, N. Wahl, *et al.*, "Development of the open-source dose calculation and optimization toolkit matrad," *Medical Physics*, vol. 44, no. 6, pp. 2556–2568, 2017.

[44] MATLAB, *version 9.5 (R2018b)*. Natick, Massachusetts: The MathWorks Inc., 2018.

[45] H. Paganetti, A. Niemierko, M. Ancukiewicz, *et al.*, "Relative biological effectiveness (rbe) values for proton beam therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 53, no. 2, pp. 407–421, 2002.

[46] S. Breedveld, P. Storchi, P. Voet, and B. Heijmen, "icycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar imrt plans," *Medical physics*, vol. 39, pp. 951–63, 02 2012.

[47] J. Fan, X. Cao, P. T. Yap, and D. Shen, "Birnet: Brain image registration using dual-supervised fully convolutional networks," *Medical Image Analysis*, vol. 54, pp. 193–206, 5 2019.

[48] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.

[49] N. Ketkar and J. Moolayil, *Introduction to PyTorch*, pp. 27–91. Berkeley, CA: Apress, 2021.

[50] E. Gibson, W. Li, C. Sudre, *et al.*, "Niftynet: a deep-learning platform for medical imaging," *Computer Methods and Programs in Biomedicine*, vol. 158, 09 2017.

[51] B. Lowekamp, D. Chen, L. Ibanez, and D. Blezek, "The design of simpleitk," *Frontiers in Neuroinformatics*, vol. 7, 2013.

[52] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," 2019.

[53] V. Yeghiazaryan and I. Voiculescu, "Family of boundary overlap metrics for the evaluation of medical image segmentation," *J. Med. Imag*, vol. 5, p. 15006, 2018.

[54] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical Physics*, vol. 25, no. 5, pp. 656–661, 1998.

[55] M. Hussein, C. H. Clark, and A. Nisbet, "Challenges in calculation of the gamma index in radiotherapy – towards good practice," *Physica Medica*, vol. 36, pp. 1–11, 4 2017.

[56] L. Yu, T. L. S. Tang, N. Cassim, A. Livingstone, D. Cassidy, T. Kairn, and S. B. Crowe, "Analysis of dose comparison techniques for patient-specific quality assurance in radiation therapy," *Journal of Applied Clinical Medical Physics*, vol. 20, no. 11, pp. 189–198, 2019.

[57] M. S. Elmahdy, T. Jagt, R. T. Zinkstok, Y. Qiao, R. Shahzad, H. Sokooti, S. Yousefi, L. Incrocci, C. A. Marijnen, M. Hoogeman, and M. Staring, "Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer," *Medical Physics*, vol. 46, pp. 3329–3343, 8 2019.

[58] M. S. Elmahdy, J. M. Wolterink, H. Sokooti, I. Išgum, and M. Staring, "Adversarial optimization for joint registration and segmentation in prostate ct radiotherapy," 2019.

[59] T. van der Hoeven, "Physics-guided deep learning for volumetric-modulated arc therapy treatment planning quality assurance," Master's thesis, TU Delft, 2022.