Anatomical Segmentation of CT images for Radiation Therapy planning using Deep Learning

Jan Schreier

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Helsinki 17.08.2018

Thesis supervisor:

Prof. Lauri Parkkonen

Thesis advisor:

Ph.D Hannu Laaksonen

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Jan Schreier

Title: Anatomical Segmentation of CT images for Radiation Therapy planning using Deep Learning

Date: 17.08.2018 Language: English Number of pages: 10+74

Department of Neuroscience and Biomedical Engineering

Professorship: Neuroimaging methods

Supervisor: Prof. Lauri Parkkonen

Advisor: Ph.D Hannu Laaksonen

Radiation therapy is one of the key cancer treatment options. To avoid adverse effect in tissue surrounding the tumor, the treatment plan needs to be based on accurate anatomical models of the patient. In this thesis, an automatic segmentation solution is constructed for the female breast, the female pelvis and the male pelvis using deep learning. The deep neural networks applied performed as well as the current state of the art networks while improving inference speed by a factor of 15 to 45. The speed increase was gained through processing the whole 3D image at once. The segmentations done by clinicians usually take several hours, whereas the automatic segmentation can be done in less than a second. Therefore, the automatic segmentation provides options for adaptive treatment planning.

Keywords: Deep Learning, CT images, Image Segmentation

Preface

Though this work consumed a lot of my effort and strength in the last months, it taught me to keep first things first. Never will I forget the joy of, after working hard, coming back to the ones I love. After years of studying, I am able to reap the fruit of my labor and I can bless those I care about the most.

I am grateful for my son for uplifting me with his joyful personality and for my parents for believing in me.

Helsinki, 17.08.2018

Jan Schreier

Contents

Symbols and abbreviations

Symbols

$\hbar$	reduced Planck constant
σ	cross section
r	radius
Z	number of protons in the atom
ω	angular frequency

Operators

$f(x) \circledast g(x)$	convolution of function f with function g
$\nabla \times \mathbf{A}$	curl of vector $\mathbf{A}$
$\frac{\mathrm{d}}{\mathrm{d}t}$	derivative with respect to variable t
$\frac{\partial}{\partial t}$	partial derivative with respect to variable t
$\sum_i$	sum over index i
$\mathbf{AB}$	dot product of vectors $\mathbf{A}$ and $\mathbf{B}$

Abbreviations

ART	Algebraic Reconstruction Techniques
CBCT	Cone Beam Computed Tomography
CNN	Convolutional Neural Network
CT	Computer Tomography
CTV	Clinical Target Volume
Conv	Convolutional Layer
DAE	Deep Auto-Encoder
DBS	Double-Strand Breaks
FBP	Filtered Back Projection
ELU	Exponential Linear Unit
FC	Fully Connected Layer
GTV	Gross Target Volume
LSTM	Long-Short Term Memory
mRNA	Messenger Ribonucleid Acid
OAR	Organs At Risk
pCBCT	pseudo Cone Bean Computed Tomography
PP	post-processing
PReLU	Paremterized Rectified Linear Unit
PSA	Prostate-Specific Antigen
PTV	Planned Target Volume
ReLU	Rectified Linear Unit
RMS	root-mean-square
RNN	Recurrent Neural Networks
SART	Simultaneous Algebraic Reconstruction Techniques
SGD	Stochastic Gradient Descend
SNN	Sequential Neural Network
SVM	Support Vector Machine
TCGA-UCEC	The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma Collection
VOI	Volume of Interest

1 Introduction

Cancer is the second most common death cause in Europe, causing 262 deaths per 100,000 inhabitants[1]. Therefore, the search for a treatment method has become one of the targets of science in the 21st century. Even though some cancer types can be well controlled when detected in early stage, such as breast and prostate cancer, the treatment of late-stage cancer and difficult cancer types, like pancreatic cancer, remains an issue. For many cancer types, a surgery is a common treatment modality to remove the tumor. However, the risk for recurring tumors is high, when only treated with a surgery. In many cases an adjuvant chemotherapy or radiation therapy is applied in order to ensure the necrosis of remaining tumor cells. Radiation therapy is, therefore, often used as a second treatment modality for reducing the risk of recurring tumors. On the other hand, radiation itself poses a risk for creating tumors, and a high radiation can cause side effects including infertility and incontinence. Therefore, it is important to constrict the radiation as much as possible to the target tissue in order to spare critical structures. For this treatment planning, the anatomical structures need to be segmented from computer tomography scans. The segmentation is currently done by a trained clinical expert, which takes between 10 minutes and several hours depending on the tumor site. Thus, an adaptive planning approach, which uses the anatomy of the day, is unfeasible in many cases.

Through the rise of artificial intelligence, several approaches have been made to automate the segmentation process using deep neural networks [2]. In this work, different neural architectures are designed and compared to current state-of-the-art neural networks. The comparison is done both in terms of quality as well as inference speed. The anatomical sites have been chosen to be the female breast and pelvis as well as the male pelvis. While in the first site breast cancer is the most prevalent cancer, the second site is affected by cancer of the uterine corpus or colorectal cancer. The segmentation of the male pelvis, on the other side, is used for treating prostate cancer. Thus, this work tries to improve the treatment of three of the most common cancer types in females and the most common cancer type in males.

This work has been performed in cooperation with Varian Medical Systems Oy, which provided part of the framework as well as the data.

2 Background

In the following, a brief introduction into cancer, its different types and treatment options is given. After this, relevant imaging modalities and radiation therapy is explained. This section will be concluded with basics of deep learning, which forms the basis of this work.

2.1 Cancer

Even though many books could be written about cancer in itself, here a very brief introduction into what cancer is and what types of cancer exist will be given. In addition to that, the cancer types, which are most prominent, will be explained and their different treatments discussed.

2.1.1 What is Cancer?

The root cause of cancer is thought to be a genetic mutation, which leads to uncontrolled cell growth. Thus, the diseased tissue starts to suppress the healthy tissue and, therefore, can hinder its function. In some cancer types, the plain accumulation of tissue can lead to pain and other adverse effects. For instance, bone metastases can compress the spinal cord. In general, there are more than 200 different cancer types, such as breast cancer, prostate cancer or leukemia.

2.1.2 Causes of Cancer

The causes for developing cancer can be categorized into micro-causes, such as mutations inside one cell, and macro-causes such as risk factors like obesity or inherited gene defects. In the following, a brief introduction into the genetic causes of cancer will be given.

The development of a cell is controlled by its DNA. This DNA contains sequences, which can control gene expression, activate proteins or be translated to messenger ribonucleid acid (mRNA). The vast functions of the DNA are out of the scope of this short introduction, but an interested reader might be refered to Anthony Griffiths' book 'Introduction to Genetic Analysis' [3]. The DNA comprises two polymer strands of deoxyribonuclid acid. These polymers consist of different nucleotids, which are connected through its phosphate deoxyribonuclid backbone. In the human DNA, four different nucleotids are used: The pyrimidine acids thymine and cytosine, and the purin acids adenine and guanine. Each pyrimidine acid forms a pair with a purin acids based on hydrogen bonds. Thymine forms two hydrogen bonds with adenine, whereas, cytosine forms three hydrogen bonds with guanine. Thus, the two strands of the DNA are tightly connected and form the DNA double-helix. The double-helix is then wound up using histones, which are supporter proteins. They not only reduce the length of the DNA, but also play a role in gene expression [4].

One function of the DNA is to provide the 'blueprints' for proteins. These proteins are used for many different things, such as building parts of the cell but also for controlling its growth. Therefore, if the 'blueprints' are damaged, the growth might

become uncontrolled, leading to cancer. For creating proteins, a part of the DNA is first transcribed to an mRNA polymere using the RNA-polymerase. The mRNA, in contrast to the DNA, is able to leave the nucleus. The mRNA is then translated by the ribosomes to create a protein. In order to keep this chapter concise, mRNA processing such as splicing has been omitted here.

Three consecutive base pairs in the DNA form a codon. Each codon exhibits a certain function in the 'blue print': Start codons communicate to the mRNA-polymerase where to start transcribing and the stop codons cause the mRNA-polymerase to terminate the transcription. The other codons are associated with a protein, which is included into the polypeptid by the ribosomes. However, different codons can be used for the same protein, e.g. guanine-cytosine-guanine and guanine-cytosine-adenine are both translated to alanine.

One type of mutations of the DNA is the point mutation. Here, one base pair of the DNA is changed. Three types of point mutations can be distinguished: silent, mis-sense and non-sense mutations. In a silent mutation, the changed codon is still translated to the same protein. In a mis-sense mutation, the changed codon is translated to a different protein and in a non-sense mutation, the initial codon is changed to a stop codon. The base pair can be changed during replication, due to a misinsertion of the corresponding base. The DNA replication itself is considered to be accurate with an error rate less than $2 \cdot 10^{-10}$ [5]. One hot spot for point mutations are methylated-cytosine-guanine base pairs. 35 % of point mutations were found to be within a cytosine–guanine dinucleotid [6] out of which 90 % were either cytosine→thymine or guanine→adenine mutations. When methylated-cytosine is spontaneously deaminated, it creates thymine and, thus, producing a thymine-guanine mismatch. When repairing, either a guanine-adenine or a cytosine–thymine base pair is created, which leads to the mentioned point mutation. It has been found that 18.2 % of pathologic lesions are due to this point mutation located in a 5'—cytosin—phosphate—guanin—3' dinucleotide [7].

Hydroxyl radicals can cause an oxidative damage, which leads similary to the deamination of cytosin to mutation. However, several polymers exist to remove oxidative damages, such as 8-oxod-guanin-tri-phosphate-phosphatase, MYH-glycosylase and 8-oxoguanin-DNA-glycosylase.

Other DNA damages are apurinic or apyrimidinic (AP) sites, where the purine or pyrimidine base is removed. Those sites can be repaired during replication through different polymerases. It can also be repaired using base excision repair (BER). Here, the DNA back bone is opened at the AP site and and the matching nucleotide is inserted.

Through ultra violet radiation, bases inside the DNA can be covalently bound, leading to dimers, which distort the double helix. The most prevalent one is the thymin-dimer. In mammals, the neucleotid excision repair (NER) is used to repair the damage. Here, similarly to the BER, the back bone is opened and the nucleotide replaced. However, the BER only removes one nucleotide, whereas the NER removes several nucleotides of the same strand and resynthesizes the matching nucleotide strand.

The two described DNA damages cause the DNA replicase to stop and, thus,

slow down the DNA replication. This is defined as replication stress [8], which can cause a specific cell response through ataxia telangiectasia and Rad3-related protein (ATR).

Ionizing radiation poses a risk for double-strand fracture and inter-strand cross links. Depending on the type of double strand breaks, either the non-homologous DNA end joining (NHEJ) or homology directed repair (HR & SS) can be used for repairing the fracture [9]. One of the proteins related to the homologous DNA repair is encoded by the gene RAD51. A modification inside this gene has shown to be correlated to several different cancer types, such as pancreatic adenocarcinoma [10], head and neck cancer [11], aggressive prostatic cancer [12], non-small-cell lung cancer [13], resectable esophageal squamous cell carcinoma [14], estrogen receptor-positive/progesteron receptor-negative breast cancer [15] and invasive ductal breast cancer [16]. The BRCA2 gene encodes a regulatory protein for the RAD51 and, therefore similarly to the defects in the RAD51 gene, a mutation of the BRCA2 gene can lead to different types of cancers, such as triple-negative breast cancer and ovarian cancer at young age [17].

During each cell cycle, the G1/s checkpoint, the S-phase checkpoint and the G2/M checkpoint exist, which can be activated, when the DNA is damaged. This response of the cell to DNA damage is activated through phosphorylation-dependent events. Several different sensor molecules exists, which signify that the DNA is damaged. Ataxia-telangeictasia-mutated (ATM) is activated at double-strand breaks using different helper proteins, which then itself phosphorylates other substrates, called the transducer proteins, and therefore initializes a signaling cascade. Similarly, the ATR can activate the signal cascade and is itself an essential protein, as its absence is lethal at least for mice. In contrast to ATM, ATR response to replication stress, whereas ATM responses to double-strand breaks [18].

One of the transducer proteins is BRCA1, which is correlated with breast cancer in females. It is phosphorylated by ATM as a response to double-strand breaks. A mis-sense mutation or frameshift inside the BRCA1 gene leads to familial breast and ovarian cancer. ATM itself is a susceptibility factor for breast cancer [19]. Furthermore, a null or hypomorphic ATM gene leads to lymphoid tumors and T-cell leukemia.

Figure 1 shows a coarse sketch of the above mentioned factors increasing the susceptibility for breast cancer and their function in the DNA repair for a double strand breaks. The ATM is used for detecting the DBS and causes the phosphorylation of the transducer protein BRCA1. BRCA1 causes the activation of RAD51, which is one of the factors responsible for the actual repair of the DBS. BRCA2, on the other hand, is one of the regulating factors of RAD51. If one of these proteins is defect, breast cancer can develop. This is by far not exhaustive and only shows the four factors - ATM, RAD51, BRCA1 and BRCA2 -, which were mentioned above to be relevant for breast cancer. The actual repair is far more complex and involves many different proteins. Also, more proteins are known to be related to different cancer types but are out of the scope of this brief introduction.

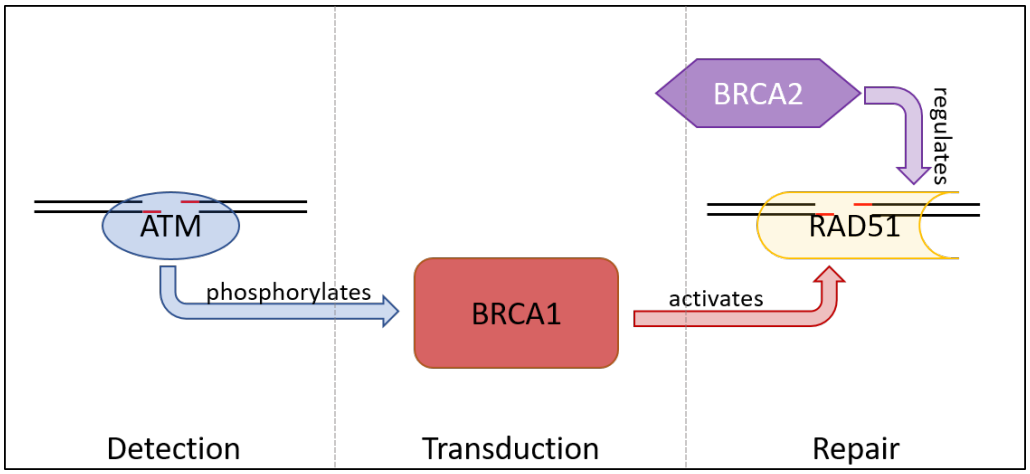


Figure 1: A sketch of four factors used in the repair of double strand breaks, which are related to breast cancer.

2.1.3 Cancer Nomenclature and Staging

As mentioned earlier, over 200 different types of cancer can be distinguished. The nomenclature for cancer types is based on the cancer site, the cell type it is originating from, the shape of the cancer cell or the name of the physician, who first identified it. Starting with the originating cell type, malignant cells developing from epithelial cells are called carcinoma, whereas cells from connective tissue are called sarcoma. A further classification can be made, by using a prefix. A carcinoma originating from the glandular epithelium is called adenocarcinoma, whereas a carcinoma stemming from the squamous epithelium is called squamous carcinoma.

For sarcoma, the tissue type further identifies the cancer. Again using prefixes, the sarcoma can be categorized with 'osteo-' for bone tissue, 'chandro-' for cartilage, 'lipo-' for fat tissue, 'rhabdo-' for skeletal muscle and 'leiomyo-' for smooth muscle.

Hematologic cancers are called lymphoma, when having a lymphoid origin, and myeloma, when stemming from myeloid bone marrow tissue. The lymphoma is further categorized into Hodgkin and Non-Hodgkin lymphoma, where the first is originating from lymphocytes. If the hematologic cancer is originating from plasma cells, it is called multiple myeloma.

To assess the severity of the cancer, a scoring system provided by American Joint Committee on Cancer (AJCC) is used:

- G1: Well differentiated
- G2: Moderately differentiated
- G3: Poorly differentiated
- G4: Undifferentiated

The basis for the grade is usually a pathologic report, where the cancer tissue is examined under a microscope. One measure for staging is the comparison of the

cancerous tissue with the normal tissue. A term used is differentiation. It "describes [the process] by which immature cells become mature cells with specific functions."[1] A tumor cell that is well-differentiated closely resembles the surrounding tissue and is thought to be less aggressive. On the other hand, an undifferentiated tumor cell differs from the surrounding tissue and is represents a more aggressive cancer type.

The pathologist can consider three different measures in classifying tissue: The tubule formation, the anaplasia and the mitosis rate. Based on those three measures a score for the pathologic tissue is given. It has been shown, that this score correlates closely with the survival rate of breast cancer patients [21].

In addition to the pathologic report, the staging takes into account if metastases exist and whether the lymph nodes are involved. The detailed staging for each cancer type can be found in the 'AJCC Cancer Staging Manual' [22].

2.1.4 Treatment Modalities

If a patient is diagnosed with cancer, different treatment options are available dependent on the patient's physical condition, the cancer type and the stage of the cancer. The most common ones are:

- Active surveillance
- Surgery
- Chemotherapy
- Endocrine therapy
- High-intensity focused ultrasound
- Radiation therapy

The basics of each one of them, the benefits and occurring side effects are described briefly in the following:

Active surveillance is used for non-aggressive cancer types, when the risk of the cancer is low, e.g. a patient, who is diagnosed with prostate cancer using the prostate-specific antigen (PSA). The cancer might not be critical or affect the quality of life for the patient, therefore a treatment of the tumor might worsen the state of the patient without benefit. However, active surveillance ensures that in the case of the cancer becoming more aggressive, appropriate actions can be chosen.

Surgery is the oldest option for cancer treatment. Bower and Waxman [23] define six different purposes of surgical oncology:

1. Cancer prophylaxis
2. Cancer classification

[1] Definition of "Differentiation" by NCI Dictionary of Cancer Terms, 2018 [20]

3. Cancer treatment
4. Surgery for oncological emergency
5. Palliative cancer care
6. Reconstructive surgery post cancer treatment

In case of an inherited cancer gem line, parts of the body can be removed as a cancer prophylaxis. An example is the familial adenomatous polyposis (FAP), which can lead to colon cancer. Here, it is recommended to remove the colon before the age of 25. Another example is a BRCA mutation, which can be an indication for a prophylactic mastectomy, meaning the removal of breasts or part of it.

As mentioned in the previous chapter, the classification of cancer is dependent on the pathologic evidence. For this, part of the tumor tissue is needed, which is gathered through biopsy. This might require a surgery.

For cancer treatment, in case of a non-metastatic cancer, the tumor can often be removed using surgery. However, during the surgery, tumor cells might get distributed in the tissue, leading to metastasis. This can be avoided by accompanying surgery with either chemotherapy or radiation therapy.

An oncological emergency occurs, for instance, when a tumor in the back bone is in risk of inhibiting the function of the spinal cord. In this case, a surgery is needed to avoid further damage.

Concerning point 5. in the above list, in palliative care, a surgery can reduce the pain by decreasing the size of the tumor. However, the goal here is not the curation of the cancer but only the improvement of the quality of life of the patient.

A surgical excision of tissue during cancer treatment can severely change the appearance of the subject and, thus, a reconstructive surgery either by an oncologic surgeon or a plastic surgeon might be indicated.

Surgery is an invasive treatment, which can cause severe side effects. For aggressive types of cancer, such as pancreatic cancer it is de facto the only option to save the patient's life. However, surgery can cause severe side effects and is only applicable when the patient is able to recover from it. A surgery can furthermore be accompanied with radiation therapy or chemotherapy to ensure necrosis of metastasis or excess tumor cells, which have not been removed surgically.

In **chemotherapy**, a drug is used to decrease the number of cancer cells. There are different types of agents: Alkylating agents aim at fast dividing cells by inducing DNA damage, whereas intercalating agents change the shape of the DNA. Anti-metabolites or topoisomerase inhibitors, on the other hand, inhibit the DNA multiplication. Other agents inhibit polymerization or depolymerization of tubulin, which inhibits the mitosis of a cell. Therefore, chemotherapy not only damages cancer cells but also other fast dividing cells in the body, such as the bone marrow, the gastrointestinal tract epithelium, gonadal cells and the hair follicles. Therefore, side effects include myelosuppression resulting in low concentration of red and white blood cells due to the damage of the bone marrow, inflammation or ulceration of the

gastrointestinal tract and infertility due to damage of the gonadal cells. Further, side effects range from mental and cognitive problems, over nausea, vomiting and fatigue, to gynecological and urinary problems [24]. In addition to that, different cancer types have differing likelihoods of developing resistances against chemotherapeutic agents. Occurring cell mutations might for instance bypass certain metabolic pathways to overcome anti-metabolites, increase the activity of efflux pumps or decrease the influx to lower the concentration of toxic agents, or develop an increased DNA repair ability [23].

Endocrine therapy or hormonal therapy is used in treatment of breast and prostate cancer. Breast cancer cells need oestrogen in order to grow. Thus, a decrease in oestrogen or a inhibiting of oestrogen receptors can be used to treat breast cancer. Similarly, prostate cancer needs testosterone to be able to grow. Therefore, the decrease in testosterone or inhibiting its receptors is able to reduce the prostate tumor. However, side effects are e.g. impotence, bone and joint pain, and hot flashes.

High-intensity focused ultrasound (HIFU) is a rather new technique, in which ultrasound-induced hyperthermia is used for cancer treatment. This technique is non-invasive and has been studied for brain tumors [25], breast cancer [26], bone metastasis [27], uterine fibroids [28] and prostate cancer [29]. The ablation is guided by either ultrasound imaging or MRI. However, this technique is not widely available and limited to a few cases of cancer for clinical practice.

In **radiation therapy**, different types of radiation are used to induce DNA damage and cause necrosis of tumor cells. The main goal is destroying the tumor, whilst preserving healthy tissue. A more in-depth introduction into radiation therapy will be given in Chapter 2.6.

2.2 Most Common Cancer Types and Treatment Options

In the following, the most common cancer types and the applicable treatment options are explained. This will give a brief overview and is by far not exhaustive. The goal is to give a motivation for the improvement in radiation therapy through this work.

The four most common cancer types in females are breast cancer (30%), lung cancer (13%), colorectal cancer (7%) and uterine cancer (7%). For men, these are prostate cancer (19%), lung cancer (14%), colorectal cancer (9%) and bladder cancer (7%). Due to the higher mortality in lung and pancreatic cancer, those two are within the top four causes for cancer related deaths. The four most prevalent death causes for females are lung cancer (25%), breast cancer (14%), colorectal cancer (8%) and pancreatic cancer (7%). For men, these are lung cancer (26%), prostate cancer (9%), colorectal cancer (8%) and pancreatic cancer (7%) [30]. The estimated new cancer cases and cancer deaths for the most common cancer types in the USA in 2018 are shown in Figure 2. Due to early screening and improvements in treatment, the mortality rate for breast cancer has decreased by 39% from 1975 till 2015 [31]. This example shows, that improvements in medical care can improve the survival rate significantly and, thus, provides a strong motivation for the research done in this work.

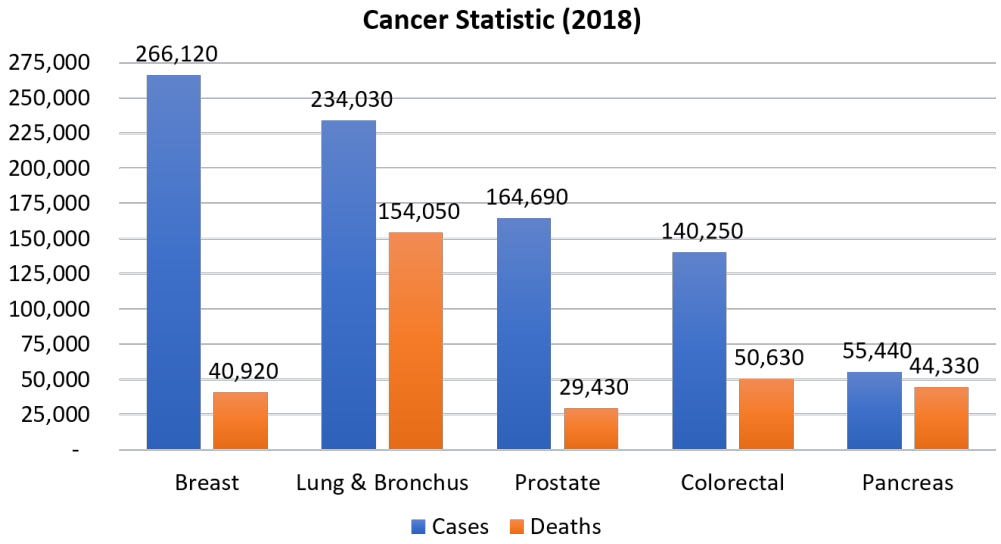


Figure 2: The estimated new cancer cases and cancer deaths for the four most deadly cancers (Breast/Prostate, Lung & Bronchus, Colorectal, Pancreas) in the USA for 2018 for both genders combined. Data from Cancer Statistics, 2018 [30].

2.2.1 Breast Cancer

Breast cancer is the most common cancer in women with an estimation of 266,120 new cases in the US for 2018, but, with a 5-year survival rate of 90% (as of 2018), poses a good life expectancy [30]. The survival rate, however, is highly dependent on the stage in which the cancer is detected. Stage I is here defined for a tumor with a

size of less than 2 cm and no nodes. Stage II has a tumor size of 2 to 5 cm and might have moveabale axillary nodes. Stage III breast cancer involves either the chest wall or the skin and might have fixed axillary nodes. Breast cancer is classified in stage IV once it develops metastases. The 5-year survival rate for stage I is 88%, 69% for stage II, 43% for stage III and 12% for stage IV (as of 2010) [23]. In a newer study from 2017, breast cancer was merely classified in local, regional and distant. Here, the five year survival rates were found to be 93–97% for local breast cancer, 78–90% for regional and 26–40% for distant depending on the race of the women [32]. The relative prevalence in the stage, in which the cancer is detected, and the 5-year survival rate for breast cancer in US in 2018 can be seen in Figure 3.

As the survival rate strongly decreases for later cancer detection, the American Cancer Society recommends a annual mammographic screening for breast cancer for women between 45 and 55, and biannually for women above 55 until their life expectancy is less than ten years [33].

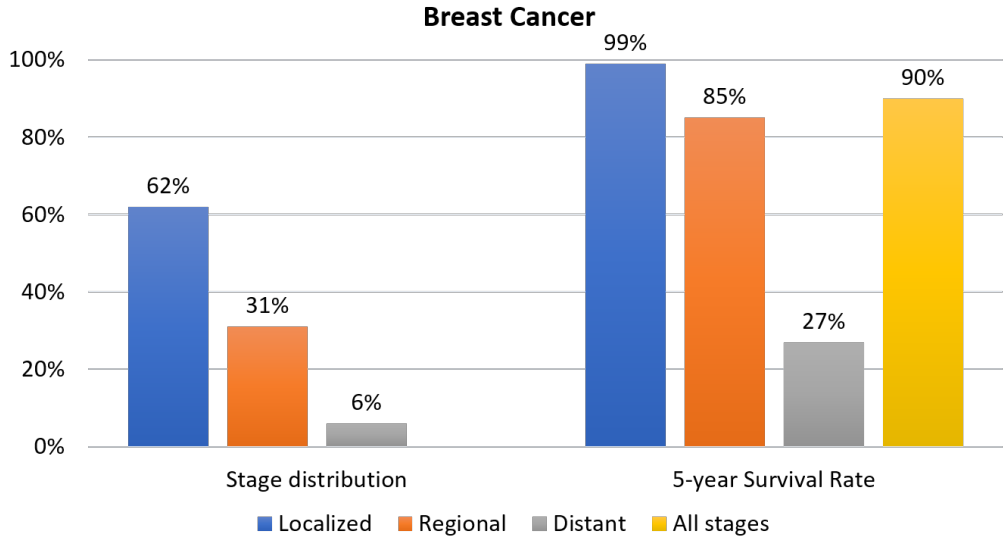


Figure 3: The relative prevalence of the stage in which the cancer is detected and the 5-year survival rate for breast cancer in US in 2018. The data are taken from the Cancer Statistics, 2018 [30].

Once a suspicious finding is seen in a mammographic image, a pathological examination is needed to confirm the finding. For this purpose, a fine or core needle biopsy can be undertaken, which can be guided using ultrasound. A risk of the procedure is the contamination of surrounded tissue with cancerous cells leading to metastases. If the finding is confirmed and the tumor is localized, the tumor can be removed surgically. If the tumor is regionally spread and auxiliary nodes are involved, a mastectomy might be necessary. The surgery can be accompanied with either pre- or post-surgical endocrine therapy, chemotherapy or radiation therapy. Pre-surgical therapy can be used to reduce the tumor size before it is removed, whereas post-surgical therapy aims at necrosis of metastatic cancer cells and at ensuring the cell death of residual cancer cells. Adjuvant radiation therapy leads to

a decrease of recurring local breast cancer from 40–60% to 4–6% [23] and is able to improve the cancer free survival compared to groups, which only receive post-operative chemotherapy [34] [35]. Radiation therapy might also be only targeted at the regional lymph nodes after breast preserving surgery. This significantly improves the survival rate [36].

For metastatic breast cancer, the median survival rate is between 18 and 24 months and is hardly curable. In postmenopausal women, a hormonal treatment is feasible with a response rate of 70%. For premenopausal women, the natural hormon production needs to be mitigated, for instance, through radiative castration. Alternatively, a chemotherapy can be applied. Radiation therapy, on the other side, can be used to reduce pain from bone metastases. High-dose chemotherapy, using bone marrow transplants or stem cell replacement, was used during 1980s but showed to not increase the survival rate significantly [23].

2.2.2 Prostate Cancer

Prostate Cancer is the most common cancer in men with an estimation of 164,690 new cases in the USA in 2018. It is also the second most common cause for cancer death with an estimation of 29,430 death in 2018 [30]. The survival rate is dependent on the stage of the cancer. The 5-year survival rate for localized prostate cancer is nearly 100 %, but only 30 % for distant prostate cancer [30]. The former one might be due to overdetection caused by the low specificity of the PSA test. The relative prevalence in the stage, in which the cancer is detected and the 5-year survival rate for prostate cancer in US in 2018 can be seen in Figure 4. The current recommendation is to discuss the applicability of a screening for the prostate speicific antigen (PSA) with patients aged 55 to 69 and stop screening for men older than 70 years [37]. A positive PSA does not necessarily implicate a tumor finding but is an indication for a further investigation using MRI, transrectal ultrasound and core needle biopsy. The treatment is dependent on whether the tumor has spread or not and on the pathological finding. 92 % of prostate cancers are detected in localized stage [38]. In addition to pathologic investigation, a genetic examination of the tumor cells can give insight weather the cancer is of an aggressive type or not [39].

The treatment option vary from active surveillance, over radiation therapy to radical prostatectomy. In the age group younger than 64 years, radical prostatecotomy is the most prevalent with 51 %, followed by radiation therapy and no treatment with each 23 %. For the age group 65 to 74 years, radiation therapy alone is the most common with 36 %, followed by radical postatectomy (30 %) and no therapy 29 %. For patients older than 75 year, most often no therapy is applied (48 %), followed by radiation therapy alone (33 %) [38].

The radical prostatectomy might be accompanied with radiation therapy and hormonal therapy. In radical prostatectomy, the whole prostate is removed. Additionally, the seminal vesicles and lymph nodes in proximity might be removed. This surgery can either be done through an open retropubic approach, an open perineal approach or a laparoscopic approach. In the first case, the surgery occurs through an open cut in the lower belly. In the second case, through an open incision in the

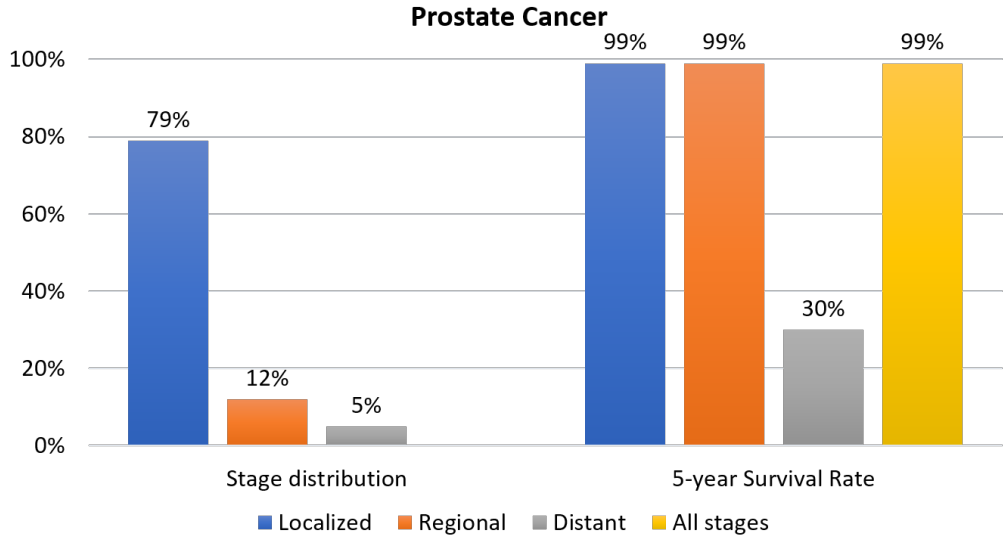


Figure 4: The relative prevalence of the stage when the cancer is detected and the 5-year survival rate for prostate cancer in US in 2018. The data are taken from the Cancer Statistics, 2018 [30].

perineum, which lays between rectum and the scrotum. In a laparoscopic surgery, the surgical instruments are inserted through three holes and the surgery is performed without the need of an open incision. The main complications of these procedures are incontinence and impotence. These complications also occur in radiation therapy as not only the tumor is targeted but also the healthy tissue of the prostate. Furthermore, a higher risk for osteoporoses exists due to the radiation pathway.

A new treatment option is high-intensity focused ultrasound, where either through the rectum or the urethra ultrasound is applied. The induced hyperthermia of the tumor tissue then leads to necrosis. It is a minimal invasive procedure, which preserves the prostate.

2.2.3 Lung Cancer

Lung cancer is the cancer with the second highest number of cancer deaths both in male and female in the USA in 2018 [30]. The 5-year survival rate in localized detection is at 56 %, whereas it is at only 5 % in a metastatic stage. The relative prevalence in detection stage and the 5-year survival rates for the different stages can be seen in Figure 5. 57 % of the lung cancer is detected in a distant stage, which is due its asymptomatic behavior [38].

The treatment options depend on the stage in which the cancer is detected. For localized or regional non-small lung cancer, a lobectomy, a sublobal excision or pneumectomy might be undertaken. Furthermore, in localized lung cancer a radiation therapeutic approach can be undertaken. The surgery can be improved through a VATS surgery, instead of an open incision. Here, through several small incisions the lobe is removed. A newer approach called unipolar lung resection, removes the lobe

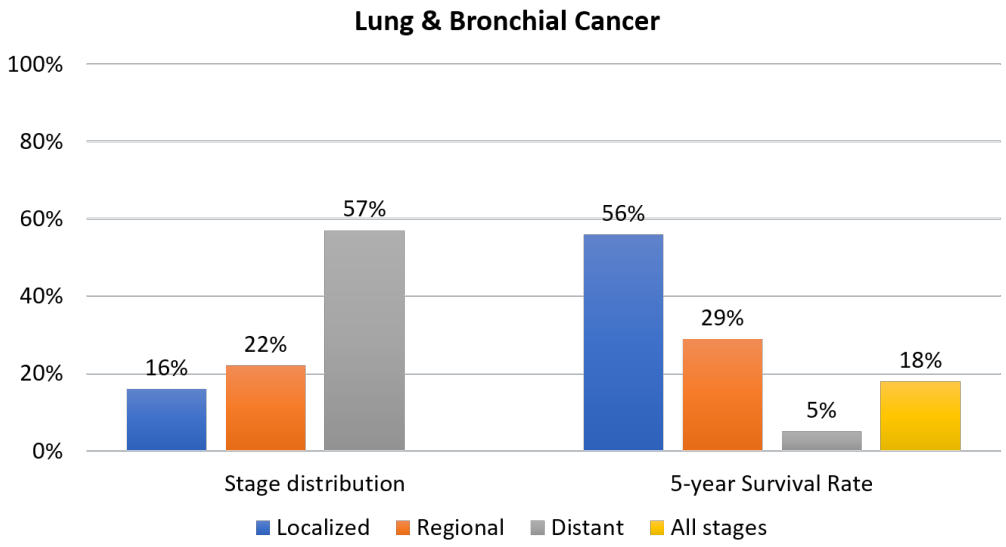


Figure 5: The relative prevalence of the stage when the cancer is detected and the 5-year survival rate for lung and bronchial cancer in US in 2018. The data are taken from the Cancer Statistics, 2018 [30].

through only one incision. The less invasive treatments seem to improve the pain management and lead to shorter hospital stays [40].

A localized lung tumor can also be treated using radiation therapy as a primary treatment. In addition to that, radiation therapy can be used before a surgery to reduce the tumor size or after the surgery to reduce the risk of a recurring tumor. Furthermore, metastatic lung cancer and small cell lung cancer are managed through chemotherapy and palliative care, such as radiation therapy for brain and bone metastases.

As smoking increases the risk of having lung cancer, a CT screening for these high-risk groups is under discussion [41–43].

2.2.4 Colorectal Cancer

Colorectal Cancer is the third most common cancer both in male (9 %) and female (7 %). The overall 5-year survival rate is 65 %, with 90 % for localized, 71 % for regional and 14 % for metastatic colorectal cancer [30]. The distribution over the different stages at presentation as well as the 5-year survival rates are shown in Figure 6.

Symptoms for colorectal cancer include constipation, diarrhoea, change in bowel habit, abdominal pain and rectal bleeding. Furthermore, anemia and low red cell volume can occur. However, these symptoms are not cancer specific and there predictive value is relatively low. For instance, the positive predictive value for rectal bleeding for men over 60 years ranges from 2.4 % to 4.5 % [44]. This is partially caused by the similarities of these symptoms to other diseases such as irritable bowl syndrome and inflammatory bowel disease [45]. Therefore, the detection of colorectal

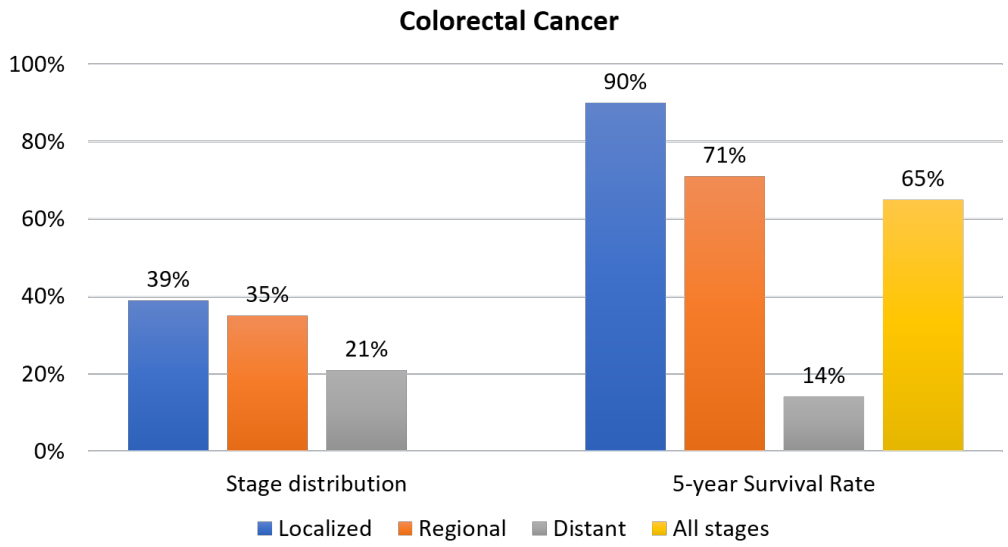


Figure 6: The relative prevalence in stage when the cancer is detected and the 5-year survival rate for colorectal cancer in US in 2018. The data are taken from the Cancer Statistics, 2018 [30].

cancer is difficult and, especially for low-risk symptoms, the detection might be delayed. There are discussions about screening for colorectal cancer, which can include a fecal blood test, endoscopy or CT colonography, but the cost efficiency and sensitivity of those tests seem to be not sufficient [46].

Colorectal cancer can be inherited through familial adenomatous polyposis. Here, multiple polyps are developed in the colon with the potential of forming a tumor. As it leads to colonic cancer in most of the patients, a prophylactic colectomy should be done before the age of 25 [23].

The management for stage I and II cancer is mostly done using partial or total colectomy. For stage III colorectal cancer, chemotherapy is additionally applied to colectomy. For stage IV cancer, chemotherapy is the most prevalent treatment. For stage I to III colorectal cancer adjuvant radiation or chemotherapy can be applied [38].

2.2.5 Pancreatic Cancer

Even though pancreatic cancer is a fairly rare cancer, with a prevalence of 3.2 %, it is the fourth most common cause for cancer death. This is due to its low survival rate, with a 5-year survival rate of 8 %. The survival rate is higher for localized pancreatic cancer with 32 % and regional with 12 %. However, metastatic pancreatic cancer has a low 5-year survival rate of 3 % [30]. The relative prevalence in detection stage and the 5-year survival rates for the different stages can be seen in Figure 7.

The detection of pancreatic cancer is difficult as it might be asymptomatic. Possible symptoms are pain in the abdomen and back, as well as weight loss, anorexia and fatigue [23]. In addition to that, patients can have jaundice. Pancreatic cancer

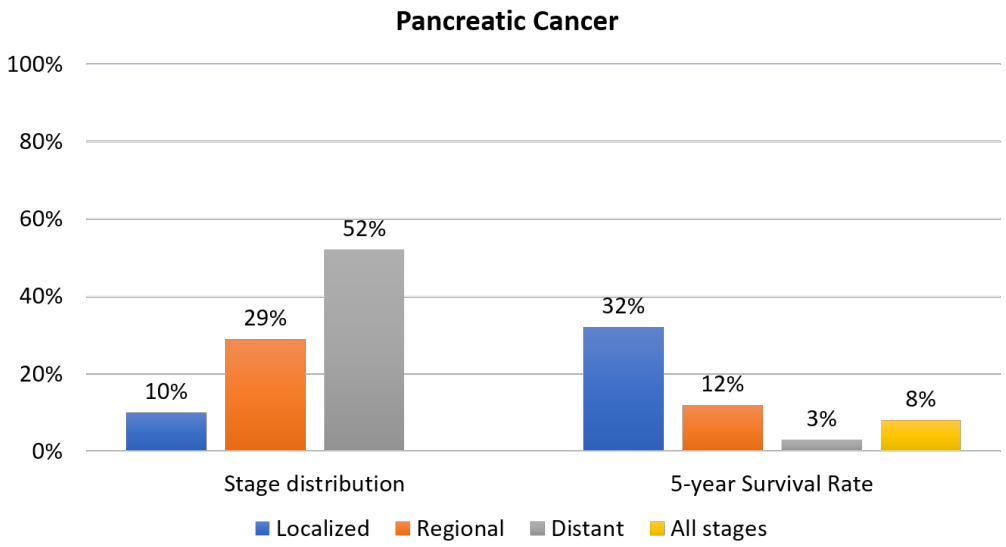


Figure 7: The relative prevalence of the stage when the cancer is detected and the 5-year survival rate for pancreatic cancer in US in 2018. The data is taken from the Cancer Statistics, 2018 [30].

tends to be an aggressive cancer, which leads to a fast growth and spreading of the tumor. One reason for its deadliness is its location: Placed in the posterior abdomen, it is surrounded by several critical organs and blood vessels. Therefore, in less than 20 % of the cases it is operable [23]. The alternative treatments are radiation therapy, chemotherapy and immunotherapy. However, as the survival rates are low, the focus is mostly on quality of life instead of curability.

2.3 Basics of Photon–Electron Interactions

Radiation therapy and X-ray imaging as well as CT imaging are based on the same physics of photon–electron interactions. In radiation therapy, ionizing radiation is used to induce double-strand breaks in cancer cells, whereas in the two mentioned imaging modalities, non-ionizing radiation is used to measure the electron density within the human body. Thus, a basis for photon-electron interactions is needed in order to provide an understanding for both the imaging modalities used and the radiation therapy.

2.3.1 Atomic Structure

The electrons are surrounding the core on discrete orbits, which are differentiable through their different energy levels. The energy levels can be found through solutions of the Schrödinger equation [47]:

$$\frac{-\hbar^2}{2m}\Delta\phi(\mathbf{r},t) + E_{\text{pot}}(\mathbf{r},t)\phi(\mathbf{r},t) = i\hbar\frac{\partial\phi(\mathbf{r},t)}{\partial t} \tag{1}$$

This equation leads, similar to the swinging motion of a string to discrete eigenmodes, which characterize the discrete energies of the electrons. This was first discoverd by Schrödinger and the solution for the hydrogen atom can be found in his article "Quantisierung als Eigenwertproblem" from 1926 [48].

2.3.2 The Photon

First described by Albert Einstein in 1905, the photon is a particle, which is electromagnetic energy. Thus, electromagnetic energy does not only exhibit wave features but also has particle characteristics. This was first discovered through the photoelectric effect, in which a certain wavelength is necessary, independent from its intensity, to loosen electrons from a metal plate. Einstein explained this effect by a quantization of the energy of a electromagnetic wave [49]. This smallest energy quantization needs to be higher than the potential energy of the electron with the highest potential energy. The relationship between the frequency and the energy of the photon was already discovered by Planck before Einstein's interpretation as particles [50] and is found to be [51]:

$$E = h\nu = \frac{hc}{\lambda} \tag{2}$$

where E is the energy of the photon, h: the Planck constant, c: the speed of light, ν: the frequency and λ: the wavelength.

As the photon can also be seen as a particle, it possesses a momentum, which was first described by de Broglie in his PhD thesis 'Recherches sur la théorie des quanta' [52]. This lead to the de Broglie equation for the momentum to be [53]:

$$p = \frac{h\nu}{c} = \frac{h}{\lambda} \tag{3}$$

2.3.3 Scattering

When an electron is exposed to a radiation field, the photons can interact with the electrons. When the direction of motion of the photon is changed, it is called scattering. A good derivation for the different scattering types can be found in 'X-Ray and Neutron Reflectivity' by François de Bergevin [54]. Its approach shall here be sketched to give an overview of the different scatter types. The assumption is that energy and momentum are conserved. Thus, the following equation holds [54]:

$$\frac{1}{2m}\left(\mathbf{p}-\frac{e\mathbf{A}}{c}\right)^2=\frac{\mathbf{p}^2}{2m}+\frac{e^2}{2mc^2}\mathbf{A}^2-\frac{e}{mc}\mathbf{Ap} \tag{4}$$

with $\mathbf{p}$ being the momentum operator of the electron and $\mathbf{A}$ of the radiation field, m the mass of the electron, c the speed of light and e the charge of the electron.

Now, when using a pertubation of the interaction terms, one can find one term associated with $\mathbf{A}^2$ and two with $\mathbf{Ap}$, which are [54]:

$$\begin{aligned} b = &\underbrace{r_\mathrm{e}\left\langle s\right|\mathbf{e}_\mathrm{sc}^T e^{ik_\mathrm{sc}r}\mathbf{e}_\mathrm{in}e^{-ik_\mathrm{in}r}\left|i\right\rangle}_{\text{Thompson Scattering}} \\ &\underbrace{-r_\mathrm{e}\sum_i\frac{\left\langle s\right|\mathbf{e}_\mathrm{sc}^T\mathbf{p}e^{+i\mathbf{k}_\mathrm{sc}\mathbf{r}}\left|c\right\rangle\left\langle c\right|\mathbf{e}_\mathrm{in}\mathbf{p}e^{-i\mathbf{k}_\mathrm{in}\mathbf{r}}\left|i\right\rangle}{m(E_c-E_i-\hbar\omega_\mathrm{in}+i\Gamma_c/2)}-r_\mathrm{e}\sum_i\frac{\left\langle s\right|\mathbf{e}_\mathrm{sc}^T\mathbf{p}e^{-i\mathbf{k}_\mathrm{sc}\mathbf{r}}\left|c\right\rangle\left\langle c\right|\mathbf{e}_\mathrm{in}\mathbf{p}e^{+i\mathbf{k}_\mathrm{in}\mathbf{r}}\left|i\right\rangle}{m(E_c-E_i-\hbar\omega_\mathrm{sc})}}_{\text{Dispersive Scattering}} \end{aligned} \tag{5}$$

where r_e is the electron radius, $\langle s|$: the state of the scattered photon, $|i\rangle$: the state of the initial photon, $\langle c|$: the state of the electron, $\mathbf{e}_\mathrm{sc}$: the unit vector in direction of the scattered photon, $\mathbf{e}_\mathrm{sc}$: the unit vector in direction of the incoming photon, $\mathbf{k}_\mathrm{sc}$: the wave number of the scattered photon, $\mathbf{k}_\mathrm{in}$: the wave number of the incoming photon p: the momentum of the photon, m: the mass of the electron, E_c: the energy of the electron after scattering, E_i: the energy of the electron before scattering, ω_in: the frequency of the incoming photon, Γ_c: the half life of the electron state after scattering and ω_sc: the frequency of the scattered photon.

The total scatter cross section of a medium can be defined through the attenuation of a through passing beam with:

$$\frac{dI}{dz}=-n\sigma I \tag{6}$$

with I: the beam intensity, z: the length of the path through the medium, n: the density of scattering centers and σ: the total scatter cross section.

When keeping only the first term, and thus assuming that the electron state stays unchanged, one can easily derive the total scattering cross section by squaring it. In the following, only the case with one electron shall be considered for simplicity and later the generalization to a number of Z atoms will be made. The total cross section can be found to be:

$$b_{th}^2=r_e^2\left(\mathbf{e}_\mathrm{sc}^T\mathbf{e}_{in}\right)^2\sum_s\left\langle i\right|e^{i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}}\left|s\right\rangle\left\langle s\right|e^{i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}}\left|i\right\rangle, \tag{7}$$

which can be simplified to the following, using that it is the sum over all possible electron states:

$$b_{th}^2 = r_e^2 \left(\mathbf{e}_{\mathrm{sc}}^T \mathbf{e}_{in}\right)^2 \sum_s \langle i|\, e^{i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}} e^{-i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}}\, |i\rangle = \left(r_e \mathbf{e}_{\mathrm{sc}}^T \mathbf{e}_{in}\right)^2 \tag{8}$$

When only considering elastic scattering, which is scattering that leaves the photon energy unchanged, the initial and final electron states are the same. Therefore, we can write [54]:

$$b_{th} = r_e \mathbf{e}_{\mathrm{sc}}^T \mathbf{e}_{in} \underbrace{\langle i|\, e^{i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}}\, |i\rangle}_{f_{ii}} \tag{9}$$

$$f_{ii} = \int \phi^*(\mathbf{r})\phi^(\mathbf{r}) e^{i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}} d\mathbf{r} = \int \rho(\mathbf{r}) e^{i(\mathbf{k}_{sc}-\mathbf{k}_{in})\mathbf{r}} d\mathbf{r} \tag{10}$$

and with this the elastic scattering cross section can be found to be [54]:

$$\frac{d\sigma}{d\Omega}_{\mathrm{elastic}} = \left(r_e \mathbf{e}_{\mathrm{sc}}^T \mathbf{e}_{in}\right)^2 \mid f_{ii} \mid^2 \tag{11}$$

Here, we see that the elastic scattering cross section is correlated to the electron density of the atom. When more than one electron is present, the resulting scattering can be written as the sum of the form factors [54]:

$$\frac{d\sigma}{d\Omega}_{\mathrm{elastic}} = \left(r_e \mathbf{e}_{\mathrm{sc}}^T \mathbf{e}_{in}\right)^2 \mid \sum_{1\le j\le Z} f_{jj} \mid^2 \tag{12}$$

and the total scattering cross section [54]:

$$\frac{d\sigma}{d\Omega}_{\mathrm{elastic}} = \left(r_e \mathbf{e}_{\mathrm{sc}}^T \mathbf{e}_{in}\right)^2 \left(Z + \sum_{1\le j\ne l\le Z} f_{jj}^* f_{ll} - \mid \sum_{1\le j\ne \le Z} f_{jl} \mid^2\right) \tag{13}$$

The elastic scattering cross section is now the Fourier transform of the electron density [54].

2.3.4 Absorption

In addition to the above described scattering, the radiation can be absorbed. The most preponderant mechanism here is the photoelectric effect, in which the energy of the photon is used to expel the electron from the atom. For this, the energy of the photon needs to be higher than the bonding energy of the electron. Thus, there are absorption edges in the spectrum: Once the energy of the photons is higher than the bonding energy of the electron in a certain shell, considerably more absorption occurs. The absorption cross section can be approximated with [53]:

$$\sigma = \frac{2\pi}{3} \alpha r_n^2 \left(\frac{\omega_n}{\omega}\right)^3 \approx \frac{\pi}{12} \frac{\alpha^3 Z^4}{n^2} r_0^2 \left(\frac{\omega_0}{\omega}\right)^3 \tag{14}$$

with α: finestructure constant, r_n: the track radius of the electron, ω_n: the frequency of the electron rotation, ω: the frequency of the incoming wave, Z: the number

of electrons, n: energy level, r_0: classical electron radius and $\omega_0 = mc^2/\hbar$: the relativistic maximum frequency.

In addition to that, radiation with a lower photon energy can excite an electron into an empty state. The electron decays then with a certain half-life to energetic lower states, while it might emit radiation. The radiative effect is called fluorescence.

2.3.5 Electron-Positron Pair Production

For radiation with an energy higher than the rest energy of an electron and a positron, meaning an energy higher than 1.022 MeV, this radiation can create an electron-positron pair. Due to the momentum conservation, part of the momentum of the photon needs to be transfered to an surrounding atom. Thus, the probability of a pair production is dependent on the surrounding medium and on the frequency of the photon. It can be approximated with [53]:

$$\sigma_p = \alpha Z^2 r^2 \tag{15}$$

The dominating mechanism for absorption and scattering is dependent on the energy of the photon. For low energies the photoelectric effect dominates, then for higher energies the Compton effect starts dominating and then for even higher energies the pair production takes the most prevalent role.

2.4 Imaging Modalities

For both diagnosis and treatment of cancer, it is important to obtain accurate anatomical information. For this purpose, different imaging modalities exist, which shall be described briefly in the following chapter and more extensively for computed tomography as this is the imaging modality used in this work.

2.4.1 Ultrasound

Ultrasound is a non-invasive imaging technique, which has the advantage of being non-radiative. Therefore, it is a low risk imaging, which can even be used for fetal imaging. The basis here is the difference in acoustic impedance inside the human body. Once an acoustic wave encounters an interface of media with different acoustic impedances, it is partially reflected with the reflected energy ratio being [55]:

$$E_r/E_i = \Gamma^2 = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2 \tag{16}$$

with E_r: the energy of the reflected wave, E_i: the energy of the incident wave, Γ: the reflection coefficient and Z_i: the acoustic impedances.

Now, using the time or phase differences in the reflected wave, the position of the reflecting boundary can be reconstructed, and the amplitude gives information about the acoustic properties of the medium. Therefore, an image can be constructed, which is able to differentiate between different media.

In cancer care, ultrasound can be used to image tumors, which are accessible through ultrasound. Some application cases are ultrasonic imaging of breast cancer, transrectal ultrasound for prostate cancer imaging and esophagogastroduodenoscopy for endoscopic imaging of pancreatic cancer. A disadvantage of ultrasound imaging is that it can only be applied to tissues, which are not covered by bone. In addition to that, it gives mostly 2D information about a small region of the patient.

2.4.2 MRI

Magnetic resonance imaging (MRI) is based on the nuclear spin resonance, which is different for varying molecules and atoms. The spin is an inherent property of every atom. When an external magnetic field is applied, the two different spin states become energetically differentiable. The energy difference is given by:

$$\Delta E = \hbar \gamma B_0 \tag{17}$$

with γ: the gyromagnetic moment of the electron and B_0: the applied magnetic field. The motion of the spin can, here, be thought similar to the precessing movement of a gyroscope. The precession frequency is called Lamor frequency and is found to be:

$$\omega_0 = \mid \gamma B_0 \mid \tag{18}$$

The orientation of the spins in thermal equilibrium can be described by a Boltzman distribution. When a electromagnetic pulse is applied, the spin distribution can be

brought out of equilibrium. The relaxation of the resulting magnetization is two fold. First, due to the inhomogenity of the external field and the slightly varying microscopic environments of the particles, the spin movements become out of phase. Second, the spins align back into their equilibrium state. Both decay times can be used in MRI imaging, defining the T1 and T2 contrast, which is named after the two different decay times.

MRI imaging provides a good spatial resolution of down to 0.25 mm in an MRI device with a 3-Tesla magnet in a medical setting [56]. Furthermore, the use of contrast agents can give additional anatomical information, such as the specific imaging of blood vessels in angiography, or metabolic information. As no external radiation is necessary, the imaging technique has a low risk. However, with high-field MRI, the risk of burns due to closed body loops increases. Additionally, patients with magnetic implants are not able to use MRI. Even though, MRI gives a good anatomical image, it does not provide the electron density, which is needed for radiation therapy planning. However, there are current approaches by companies to calculate the electron density from an MRI image.

2.4.3 Positron-Emission Tomography

The usability of PET for oncology comes from the possibility of attaching radioactive markers to metabolic molecules. Therefore, tissues with a high metabolic rate can be found. As cancer tissue is mostly fast growing compared to other tissue, its metabolism rate is high. Thus, PET imaging can give an additional indication for cancer tissue.

As a product of the β^+-decay of isotopes like fluorine-18, a positron is created. This positron creates a positronium atom together with an electron. Positronium is a highly unstable system as it is formed out of an electron and its anti-particle the positron. It decays with a half life of approximately 124 ns [57], creating an even number of photons. The probability of the decay decreases strongly with the number of photons emitted, such that in practice only two-photon decays are used. Within the life time of the positron, it is able to travel a couple of millimeters dependent on the isotope. For instance, for the above mentioned isotope fluorine-18, its average distance is 0.35 mm [58].

One property of the positron-electron annihilation is that due to the momentum conservation, the two photons are emitted in opposite directions. Therefore, by using coincidence measurements, the photons from the same location can be detected and the position of the annihilation reconstructed. Therefore, the location of the metabolic markers can be detected inside the patient's body. This leads to a more accurate spatial detection of the cancer.

2.4.4 X-Ray

The first published X-ray image was taken by Wilhelm Röntgen in 1895 [59]. The principle is still the same, even though the measurement time and the radiation dose are drastically decreased. In principle, electrons are accelerated within a vacuum tube and when hitting a metal plate emit bremsstrahlung, which is the x-ray radiation.

The wavelength of the emitted photons is dependent on the voltage difference with which the electrons are accelerated. In medical x-ray imaging, voltages between 10 kV and 100 kV are typically used. Furthermore, not only uncharacteristic radiation is produced but also material specific radiation, which is dependent on the metal used as a target.

The radiation then is guided through the human tissue, which shall be imaged, and captured afterwards either by photographic film or nowadays mostly by electronic photo detectors. As different tissues differ in their attenuations, the x-ray is attenuated depending on which tissue it is passing through. The attenuation is caused by the in section 2.3 mechanisms; mainly absorption and Compton scattering. From the cross section the attenuation coefficient can be easily derived with [53]:

$$\mu = N N_{at} \sigma \tag{19}$$

with N: number of atoms, N_{at}: atom number density and σ the total cross section. The attenuation within the tissue can be calculated with [60]:

$$I = I_0 e^{-\mu x} \tag{20}$$

This can be rewritten for spatially changing attenuation coefficients using the integral [61]:

$$I = I_0 e^{\int -\mu(x) dx} \tag{21}$$

The x-ray image is, thus, only a 2D projection of the 3D attenuation profile of the tissue.

2.5 Computed Tomography

In the following, computed tomography (CT) shall be explained in a bit more detail as this image modality is used in this work. The principle of CT is similar to X-ray imaging. Also here X-rays are used to gather information about the attenuation profile of the human tissue. However, the goal is to achieve 3D information in contrast to the 2D projection produced by standard X-ray imaging.

2.5.1 The Sinogram

In CT imaging, a form of X-ray images are taken from different angles in order to reconstruct 3D information. The simplest idea of viewing it is by first considering one slice of the 3D image and using several parallel x-ray beams. The parallel x-ray beams travel through the transverse plane of the human body and each beam is attenuated according to the tissue it is passing through. Now, in the next step, the beams are passing through the body from a different direction, creating a slightly different image. This procedure is done for many different angles. At each angle, a one dimensional image is created. Stacking all those one dimensional images together, the so called sinogram is created. Mathematically the sinogram can be described as the radon transform of the attenuation profile:

$$p(\theta, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - s)dxdy \tag{22}$$

with θ: being the angle at which the x-ray projection is taken, s: the position on the one dimensional x-ray projection and $f(x, y)$: the 2-dimensional attenuation profile of the tissue.

2.5.2 Reconstruction

The simples version of reconstruction is performing the inverse radon transform on the sinogram. This is equivalent to a 2D Fourier transform and can, thus, be calculated quickly on a computer. The inverse radon transform is defined as [62]:

$$f(x, y) = \int_{0}^{\pi} \int_{\infty}^{\infty} \frac{\partial p(\theta, r)}{\partial r} \frac{1}{x\cos\theta + y\sin\theta - r} drd\theta \tag{23}$$

However, this method leads to a blurred image, as the intensity due to scattering decays with $1/r$. Therefore, deconvolving the sinogram with the opposite convolutional filter, improves the reconstructed image. This is called filtered back-propagation. The deconvolution can be fastened up by multiplying the Fourier transformed sinogram with the deconvolution function instead of performing the convolution operation [63]. In addition to that, several other filter types can be applied such as Shepp-Logan, Hammak or Hamming.

Another approach to reconstruction is the use of iterative image reconstruction. The goal of all reconstructions is to find the image vector f, which satisfies the equation:

$$p = \mathbf{A}f \tag{24}$$

when p is the projected vector and $\mathbf{A}$: the matrix of the forward solution. This equation can be solved by finding the inverse of $\mathbf{A}$. This is, however, impractical as the number of elements of the matrix is of order 10^8 to 10^{12} [62]. One approach is to optimize each equation in the equation system separately and iterate over the process. This is called Algebraic Reconstruction Technique (ART). As ART calculates each ray after each other, it is fairly slow. It can be improved by updating all equation simultaneously, which is then called Simultaneous Algebraic Reconstruction Techniques (SART). The update rule is [64]:

$$f_j^{(k+1)} = f_j(k) + \alpha \frac{\sum_i \left(\frac{p_i - \sum_{n=1}^{N} A_{in} f_n^{(k)}}{\sum_{n=1}^{N} A_{in}} \right)}{\sum_i A_{ij}} \tag{25}$$

Here, the forward solution is applied to an image guess f^k and is compared to the actual sinogram p. The image guess is then improved until the error between the suggested sinogram and the actual sinogram is small enough.

A more robust approach can be achieved using a Bayesian method. For this, an additional noise term is assumed. Therefore, the measured projection can be described as:

$$p = \mathbf{A}f + n \tag{26}$$

When assuming a gaussian noise with zero center, finding the maximum a posteriori likelihood becomes equal to finding the following minimum [64]:

$$f = \arg\min_f \left(\mid p - \mathbf{A}f \mid^2 + \lambda \mid f \mid^2 \right) \tag{27}$$

with λ: the regularization parameter. λ needs to be larger for a higher noise level. Finding this minimum can be done, for instance, using gradient descent algorithms.

2.5.3 Cone Beam Computed Tomography

In cone beam computed tomography (CBCT), a cone-like x-ray beam is used. A sketch can be seen in Figure 8. Therefore, in comparison to conventional single-slice CT, the patient is not imaged slice by slice but several slices at once. This gives an advantage in speed but the image quality suffers from it, as only approximate reconstruction mechanisms exist for circular CBCT. However, for helical CBCT an exact reconstruction mechanism exists [65]. But it is more prone to error for noisy images.

As CBCT is often used with low radiation dose, the measurement contain more noise. A solution to the problem can be iterative reconstruction, but commercially often filtered back propagation (FBP) is still used. The CBCT is more susceptible for scattering artefacts and thus the image quality can be poor. However, it is nowadays often used in radiation therapy to correct the patient position prior to treatment. In comparison to fan-beam CT (FBCT), it was found that not only the image quality in CBCT is worse but also twice to three times as much radiation is needed in CBCT compared to FBCT [66]. In comparison to multi-slice CT, it was found that CBCT uses less radiation whilst being comparable in image quality [67].

Additionally, for CT-guided ablation of metastatic lung cancer, CBCT is able to improve the treatment speed in comparison to CT imaging [68].

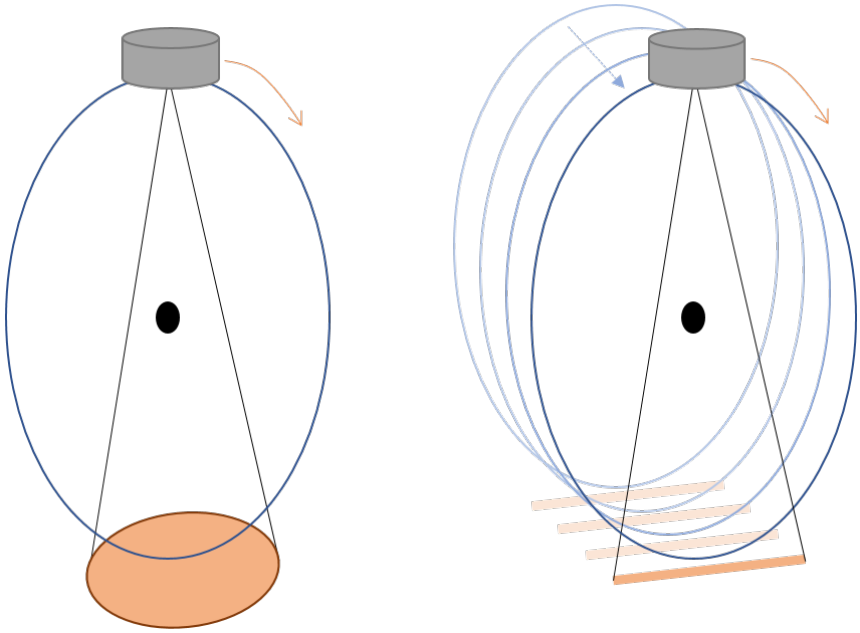


Figure 8: A sketch of a CBCT (left) and a single-slice circular CT device (right).

2.5.4 Artifacts

When a X-ray beam passes through a strongly absorbing medium, such as bones, its low energy photons are more likely to be absorbed or scattered. Thus, after the object mainly high energy photons remain, which experience a reduced attenuation profile. Therefore, the grey values in a CT image after a bone structure are lower than in parts where the X-ray beam has not passed through bone structure before. This effect is called **beam hardening**. As CBCT has a lower average photon energy, it is more susceptible to beam hardening artifacts [69].

Another problem are **motion artifacts**, which are generated, if the patient or part of his tissue moves. In CBCT, this can cause artifacts, which look similar to scattering artifacts. In imaging the abdomen this can become a problem, when bowel movement occurs.

Because of the geometry of the CBCT imaging, tissue close to the X-ray source are scanned with more beams than those at the detector side. This can lead to **aliasing** artifacts, which can be seen as stripes in the image.

In order to reduce the artifacts, different filters can be tried. Furthermore, iterative reconstruction algorithms seem to be able to reduce artifacts and different artifact removal algorithms have been proposed [70].

2.6 Radiation Therapy

In radiation therapy, radiation is used to damage cancer cells and ultimately cause necrosis in the cancer cells. Similarly to chemotherapy, there is a trade off between successful cancer treatment and toxicity. In the following, a brief introduction to the mechanisms of radiation therapy, its different types and radiation therapy planning will be given.

2.6.1 Principle of Radiation Therapy

The goal of radiation therapy is to cause necrosis in cancer tissue, whilst preserving healthy tissue. Through ionizing radiation, double-strand breaks (DBS) are introduced in the DNA of the cell. As the tumor cells are growing fast in comparison to healthy tissue, the genetic damage leads to problems in the duplication earlier. In addition to that, most cancer types stem from a mutation in the DNA repair mechanisms. Hence, cancer cells are more susceptible to radiation therapy than healthy cells. This said, ionizing radiation causes mutations also in healthy tissue, leading to a risk of secondary tumors due to the radiation therapy. Thus, it is crucial to spare healthy tissue around the tumor.

The type of radiation therapy used is determined by the desired clinical outcome. For instance, in adjuvant radiation therapy to localized breast cancer surgery, it might be the goal to necrotize potential metastatic tumor cells. Therefore, the whole breast is radiated. On the other side, in brachytherapy of localized prostate cancer, the goal is to only treat the tumor and spare the surrounding healthy prostate tissue. Even more crucial, in stereotactic radiation therapy, or radiosurgery, it is of utterly importance to spare the surrounding tissue. An application is the removal of brain tumors, where the surrouding brain needs to be preserved.Therefore, a good spatial accuracy is needed for radiation therapy.

2.6.2 Dose Distribution

Generally, the absorbed dose is dependent on the depth in the tissue, the energy of the incoming photons, the tissue composition, the intensity and the shape of the beam. A phantom is used to estimate the delivered dose for new machines and create a models for the delivered dose in the patient. However, the dose in the patient is estimated by considering the radiation transport inside the tissue.

One effect, which needs to be consider, is the **dose build-up**. When the photons arrive at the skin, secondary electrons are emitted through the photo electric effect. These electrons are high in energy and are therefore also considered as ionizing radiation and add up to the delivered dose. These secondary electrons are created within the first few centimeters of the tissue; creating the so called 'build-up zone'. In this zone, the dose increases with depth as more and more secondary electrons are contributing to the total ionizing radiation. However, these electrons are stopped easily through the tissue, such that after the build-up zone, they experience a nearly exponential decay. For treatment in deeper laying body regions, this prevents over-radiation of the skin. However, when treating skin cancer, it poses a hindrance, as

the skin is barely radiated. This problem can be overcome by placing a bolus on the skin prior to the treatment. The bolus then covers the build-up zone, such that the skin is treated with the maximum intensity. The depth of the build-up zone is dependent on both the energy of the photons and the tissue composition, leading to a larger build-up zone for higher energetic photons. This phenomenon can be used to treat deeper laying structures without over-radiating outer anatomical layers [60]. A sketch of the dose build-up is shown in Figure 9.

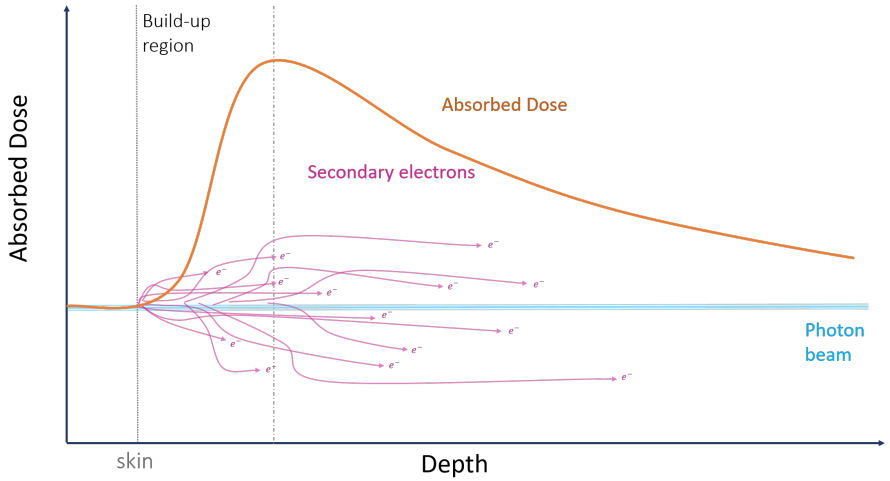


Figure 9: A sketch of the build-up zone with the approximated delivered dose.

The above mentioned effect is only one of those, which needs to be taken into consideration. Generally the absorption is dependent on the electron density of the medium itself and the intensity of the beam at this location. As bones are dense tissue with a Houndsfield Unit (HU) of 1700 for the outer bone and 301 for the inner bone [71], tissue, which is radiated with a beam passing through a bone first, receives less radiation than tissue treated without a bone in the pathway. Thus, it is necessary to model the dose distribution prior to the treatment, in order to ensure the right dose at the target volume without harming surrounding tissue.

2.6.3 Treatment Planning

As described in Section 2.6.2, the delivered dose is dependent on the body constitution. The target volume is typically defined in the following way: First, the tumor itself needs to be identified and segmented in the image from the surrounding tissue. The tumor volume is called the **gross tumor volume**, short GTV. Secondly, a margin is drawn around the tumor accounting for non-imageable tumor spread. This volume is called the **clinical target volume** (CTV). Thirdly, due to uncertainties in treatment planning, such as movement of the tumor between the planning and the treatment, another margin needs to be drawn around the CTV, to ensure the treatment of the tumor. This last volume is then called **planned target volume** (PTV) [72].

During treatment planning, another type of volume needs to be segmented and taken into account. These are the organs at risk (OAR). Goal of the treatment planning is to minimize the radiation exposed to the OAR, whilst maintaining the planned dose in the PTV. The organs usually need to be segmented manually from CT images prior to the treatment. An example is the treatment of prostate cancer, where the OAR are the bladder and the rectum. An over-radiation of the OAR can lead to side effects such as (fecal) incontinence or inner bleeding.

The time needed for this segmentation varies between different tumor sites, clinical protocol and clinician. For example, for head and neck cancer, the segmentation takes between 25 and 145 minutes [73]. This does not only reduce the cost-efficiency of the treatment, but also worsens the clinical outcome due to the delay of the treatment [74]. In addition to that, on-couch adaptation to the anatomy of the day becomes unfeasible and, in the case of a significantly changed anatomy, the treatment needs to be rescheduled. The segmentation time poses, therefore, a hindrance in clinical outcome and cost-efficiency.

In the past, conventional image segmentation methods have been used to automate the segmentation process, such as thresholding and smoothening [75], multi-atlas approaches [76] and more recently Laplacian meshes [77]. However, for some anatomical regions, such as the pelvis, the structure can change significantly, such that the classical approaches are unable to provide a sufficiently accurate segmentation.

2.7 Deep Learning

Deep Learning is a part of the more broad subject of artificial intelligence. Generally spoken, deep learning uses an artificial neural network to conquer a task. It is deep in comparison to classical machine-learning approaches like support vector machines (SVM), which can be thought to use only one layer of neurons. In contrast to this, deep learning uses several to several hundreds of those neural layers.

2.7.1 The Neuron

Each neuron can be seen as a connection between a lower and a higher layer. The general structure is multiplying of the input by weight, adding a bias and executing an activation function. Thus, its mathematical operation can be written as:

$$f(x) = A(wx + b) \tag{28}$$

with A: the activation function, w: the weights and b: the bias. Through the learning process, both the weights and the bias are usually learned. In some cases, also parameters of the activation function need to be learned.

Each neuron then forwards its output to another neuron and creates, thus, a neural network.

2.7.2 Loss Functions

In order to measure how well the model performs, the output is mapped to a single number. This loss function or error function can, therefore, be used to adjust the network to an desirable result. In the following, the loss functions used in this work will be defined.

For pixel-wise output, such as image segmentations, a simple loss function is the **binary categorical accuracy**. It is defined as the percentage of the pixels, which are correctly classified.

For medical segmentations, as large parts of the image are belong to the background, the binary categorical accuracy can lead to false classification of the whole image to the background class. Therefore, a definition of the loss using the intersection and union of the segmentation with the ground truth is helpful. One loss function is the **jaccard loss**, also called jaccard distance, which is based on the jaccard index defined as [78]:

$$r = \frac{\mid X \cap Y \mid}{\mid X \cup Y \mid} \tag{29}$$

The jaccard loss can then be defined as:

$$J = 1 - r \tag{30}$$

Similarly, the Sørensen-Dice coefficient is defined by intersections:

$$DSC = \frac{2 \mid X \cap Y \mid}{\mid X \mid + \mid Y \mid} \tag{31}$$

This can be rewritten as:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{32}$$

with TP: true positives, FP: false positives and FN: false negatives. The dice loss can then be defined as $1 - DSC$.

2.7.3 Back Propagation

Back propagation is the key part of the training of the neural network. Mathematically each parameter, weight and bias of the network can be seen as a dimension in the multi-dimensional solution space. The goal of the training is to find the global or at least local minimum of a loss function, which can be seen as the error of the current state. Therefore, the network can be trained by decreasing the loss iteratively. This can be done through different learning algorithms, which are mostly based on gradient descend. Here, the gradient at a certain point is calculated and, following the gradient, the parameters are updated and the loss function decreased. The calculation of the gradient is based on the chain rule:

$$\frac{\partial f(g(x))}{\partial x} = f'(g(x)) \cdot g'(x) \tag{33}$$

This can be iteratively done for as many functions as necessary, which are inserted into each other. In back propagation, first the innermost gradient is calculated at the output side of the network. Then using the chain rule, every other gradient is calculated using the gradient of the connecting neurons. The update rule for each parameter is here defined as:

$$w \leftarrow w - \alpha \frac{\partial L(w)}{\partial w} \tag{34}$$

with L: loss function, α: learning rate and w: parameter to be updated.

The learning rate is a hyper parameter, which needs to be defined by the developer. A hyper parameter is a user defined parameters, which is not learned by the model. A too high learning rate hinders conversion by jumping in the space of the loss function, whereas a too low learning rate converges only slowly and might find a local minimum far from the global one.

2.7.4 Learning Algorithms

As mentioned above, most learning algorithms are based on the gradient descend. However, for a large data size, it is inefficient to sample through the whole dataset in order to calculate the gradient. It can be shown that the accuracy of the gradient increases with $\sqrt{N}$, when N is the number of samples. Therefore, the sample size needs to be four times larger to double the accuracy. The learning speed can be improved, by stochastically sampling through the dataset while training. This is called **stochastic gradient descend** (SGD). The update rule is here the same

as in gradient descend with the difference that the loss function and its respective gradients are only calculated using a subset of the dataset.

In order to achieve a smoother convergence, the preceding gradients can be taken into account. This results in a smoother movement inside the loss space, as a decaying average of the before calculated gradients is taken [79]. An update rule for this is the so called **RMS-Prop**:

$$w_i \leftarrow w_{i-1} - \alpha \frac{\partial L_i(w)}{\partial w} - \beta \frac{\partial L_{i-1}(w)}{\partial w} \tag{35}$$

For this the gradients of each iteration is saved.

Another option to improve the convergence, is the use of the second moment of the gradient. This approach, called **Adam**, has been suggested by Kingma and Ba [80]. The update algorithm is here defined as [80]:

1. calculate gradient: $g_t \leftarrow \nabla_w f_t(w_{t-1})$
2. update first moment: $m_t \leftarrow (\beta_1 m_{t-1} + (1-\beta_1) \cdot g_t) \,/\, (1-\beta_1^t)$
3. update second moment: $v_t \leftarrow (\beta_2 v_{t-1} + (1-\beta_v) \cdot g_t^2) \,/\, (1-\beta_2^t)$
4. update parameters: $w_t \leftarrow w_{t-1} - \alpha m_t / (\sqrt{v_t} + \epsilon)$

2.7.5 Activation Functions

Activation functions play an essential role in deep learning, as they introduce the non-linearity. Hornik and colleagues [81] showed that a multi-layer neural network with a squashing activation function is able to approximate any Borel measurable function[2] given that it has sufficiently many hidden units. Therefore, a neural network can be seen as an estimator for any function. Based on this result, deep learning is a method to estimate an unknown functional relationship between a certain input and output.

Several activation function exist, which can be used for training neural networks. The idea behind the first activation functions was to imitate neurons, which have a binary output. Thus, one possible activation function is a step function. However, with such a function, the gradient is either 0 or infinite, and therefore no back propagation is possible. To overcome this problem, the **sigmoid** function can be introduced. The shape is similar to a step function with the difference that it is smooth. The sigmoid function is defined as:

$$\sigma(x) = \frac{e^x}{e^x + 1} \tag{36}$$

Another choice is the **hyperbolic tangent**, which in fact is a scaled and biased sigmoid function [83]. It is defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{37}$$

[2]Every continous numerical function on a Borel measurable subspace of $\mathbb{R}^d$ is a Borel measurable function. However, for an exact definition the reader is referred to "Measure and Integration Theory" [82]

Both of those functions are differentiable and therefore the gradients for the back propagation can be calculated. However, it has turned out, that both of those functions are prone to the vanishing gradient. This problem mainly occurs in deep networks. The gradients, as shown in Equation 33, are multiplied through every layer. Therefore, the gradient decreases exponentially with each layer. If the gradient is smaller than one, it approximates zero after a sufficient number of layers. Therefore, layers close to the input are not iteratively improved.

Because of the vanishing gradient problem, the **rectified linear unit**(ReLU) has been suggested as an activation function [84]. They are defined as the following:

$$f(x) = \max(0, x) \tag{38}$$

Dahl and colleagues [84] reported that partially due to the ReLU activation better results compared to state of the art, which was sigmoid and tanh at the time, could be achieved.

A further development of ReLU are leaky ReLUs. Here, even at a value less than zero, the output is not fully inhibited. One advantage is that the gradient is not vanishing once the input is less than zero. It is defined as:

$$f(x) = \max(\alpha x, x) \tag{39}$$

with $\alpha \leq 1$. This parameter can be treated as a hyper parameter provided by the user, or trained as a parameter of the neural network. In the later case, it is called **parameterized ReLU** (PReLU).

Another activation function is the **exponential linear unit**(ELU). Preserving the shape of the ReLU and PReLU for positive x, it replaces the linearity for negative x by an exponential decay. Thus, similarly to ReLU and PReLU it diminishes the vanishing gradient problem, but is able to approach better a zero average activation, which improves the learning rate [85]. It is defined by:

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(\exp(x) - 1) & \text{if } x \leq 0 \end{cases}$$

The different activation functions are sketched in Figure 10.

2.7.6 Overfitting and Regularization

If the capacity of a model is higher than the data size of the training set, the model tends to learn specific cases instead of a general description of the task. This means that the model performs well on the training data set but poorly on the test set, a phenomenom known as overfitting. A simple example is when the true distribution is linear with gaussian noise but the model learns each training point instead of the linear relationship. This problem can even occur for more complex tasks. For instance using Google's inception_v3 network on a mammographic database with 2,500 images for classification into two classes already leads to overfitting, as the model is capable of remembering each of the images in the training set.

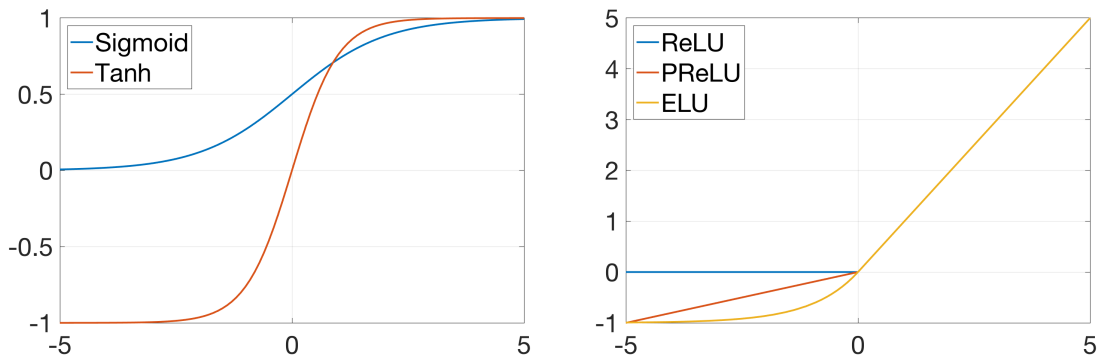


Figure 10: Five activation functions: sigmoid, hyperbolic tangent, ReLU, PReLU with $\alpha = 0.1$ and ELU.

One regularization method is **dropout**. Here, each neuron is assigned a probability of being turned on or off during training. Thus, the capacity of the model is reduced. It has also shown that the accuracy of a model is improved using dropout [86]. A more interesting proof has been provided by Gal and Ghahramani [87], showing that dropout is equal to Bayesian approximation of the posterior likelihood. Therefore, by keeping the dropout active during validation, a Bayesian approximation can be sampled, which then provides not only a binary output but also an uncertainty measure for the prediction.

Another regularization method is **early stopping**. For this, the training is stopped prior to full convergence on the training set. The validation error usually starts increasing after a certain point in training, when overfitting occurs. At this point, the model starts remembering single training data points instead of a general relationship. Therefore, by stopping early enough, this can be prevented.

Furthermore, **data augmentation** can be used as a preprocessing. With this technique more training data can be artificially generated. Options for it are random cropping, flipping, mirror, rotation, adding of noise and warping. The problem here is the trade-off between creating too similar samples and samples, which are not representative anymore. For instance, warping of CT scans can lead to unreasonable anatomic structures, such that it prevents the model from learning the true anatomical structure.

2.7.7 Batch Normalization

In batch normalization, the input to a chosen layer is normalized and often whitened. This is another method to avoid the vanishing gradient problem. Batch normalization is here done for each batch separately. It has been shown that batch normalization is able to improve the accuracy and enhance the training speed. For each layer the input to this layer is normalized and whitened along each input dimension using the following equation [88]:

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{\text{Var}\left[x^{(k)}\right]}} \tag{40}$$

2.7.8 Layers

The layers are the 'lego blocks' out of which a network is usually constructed. The conceptionally simplest layer is a **fully connected layer** (FC), where every neuron of the previous layer is connected to every neuron of the following layer. The input is for each neuron multiplied with a neuron and input specific weight. After this a bias is added. Therefore, the number of trainable parameters is $(N+1)M$ with N: begin the number of incoming neurons and M: being the number of outgoing neurons. Due to this squared dependency of the parameter count on the size of the network, it easily becomes inefficient, when considering large input data, such as (3D) images.

An alternative, which is especially useful for images, are **convolutional layers** (Conv). A convolution is a mathematical operation, which can be defined in the following way [89]:

$$f \circledast g(x) = \int f(t)g(x-t)dt \tag{41}$$

In classical image processing, this can be used for smoothening and edge detection filters. Instead of continuous functions, a discrete convolution is used. The image matrix is then convolved with a smaller matrix, leading to a transformed image. A common convolution matrix size is here 3x3, 5x5 or 7x7. The output of this discrete convolution is then calculated by [90]:

$$O_{ij} = \left[\sum_{p=-n/2}^{n/2} \sum_{q=-m/2}^{m/2} I_{p+i,q+j} f_{n/2+p,m/2+q} \right] /mn \tag{42}$$

with O: output image, I: the input image and m, n: filter width and height. A moving average filter and a Laplacian filter as examples are shown in Figure 11.

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

0	1	0
1	-4	1
0	1	0

Figure 11: The convolution matrices for a moving average (left) and a Laplacian filter (right).

In deep learning, the weights of the matrix are learned during training. Typically, the convolutional filter for the first layer represent different types of edge and frequency filters [91], whereas features in higher layers represent more complex shapes, such as car wheels or face shapes [92]. Therefore, by making the model deeper, as to say by stacking convolutional layers after each other, more higher level representations can be learned. This knowledge can then be used for example for image classification or image segmentation. A commonly used term is the **receptive field** of a neuron. As the network gets deeper, the field of data a neuron can see from the input through the stacked convolutions increases. Therefore, the receptive field is a term to describe what kind of features a neuron could potentially learn. While in the early layers of a neural network it is fairly small, only allowing simple edge detection, in deeper layers

the receptive field becomes large, such that larger spatial context such as faces can be detected. An illustration of the receptive field is shown in Figure 12. Here, the grid represents the matrix of values in each layer. In the second layer, the neuron is able to see a field of size 3x3 in the previous layer. In the third layer, the neuron sees a grid of size 3x3 of the second layer but already a grid of size 5x5 of the first layer.

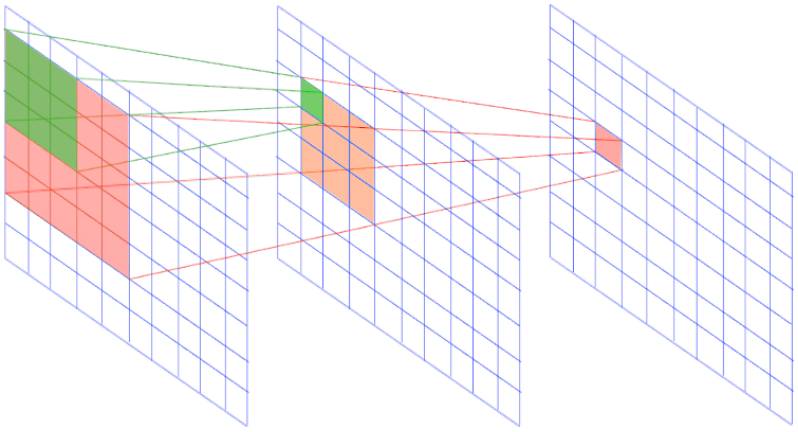


Figure 12: An illustration of the receptive field of a neuron in the third layer using a convolutional filter of size 3x3.

Sometimes in literature an **activation layer** is defined. This means that an activation function, such as the ones described in Section 2.7.5, are applied to each neuron. Furthermore, there are **dropout layers**, which apply the in Section 2.7.6 explained dropout to each neuron.

In addition to that, **pooling layers** are commonly used. The goal is here to reduce the resolution of the previous layer. This can work as a regularization but is mainly necessary in order to construct deeper models with limited memory. In addition to that, the pooling layer increases the receptive field by a factor equal to size of its grid. The most common used pooling operation is max pooling. Here, the maximum value of a grid is propagated to the next layer. The grid size then specifies the downsampling factor. Other pooling strategies, are average pooling, where the average of each element inside the grid is taken, mixed max pooling and gated max pooling [93]. However, the most prevalent pooling layer due to its simplicity and compute cost efficiency is the max pooling layer. In the following, unless otherwise specified, the max pooling layer will be meant by pooling layer.

2.7.9 Model Architectures

The architecture of the model largely depend on the application case. However, this is not strongly correlated to the field of science. For instance, a segmentation model used for segmenting objects on a street can be good in segmenting different organs in a CT scan. In the following, the main architecture types shall briefly be described.

Sequential Neural Networks (SNN) have one input and one output. There, every layer comes after another. These networks can be build using fully connected layers, convolutions or similar.

A subtype of SNNs are **convolutional neural networks** (CNN), where the layers of the SNN are convolutional layers. These are especially good for applications, which have spatial correlation, such as images and movies. An example for a CNN is the Alex-Net by Alex Krizhevsky [91] or the inception network by Google.

Another idea for increasing the performance of CNNs are **residual neural networks**, where residual connections are introduced. In these networks in addition to the convolutional layers, skip connections are introduced. Through the skip connections, the information obtained by previous convolutions can be used in deeper parts of the network. This is especially important, when the network becomes deeper. This has first been implemented by He and collegues (2015) in their ResNet [94]. It is also possible to scale those skip connections, which can make the training more stable. This has been done, for example, in the Inception-ResNet network [95]. A sketch of a shallow CNN and its residual counterpart is shown in Figure 13.

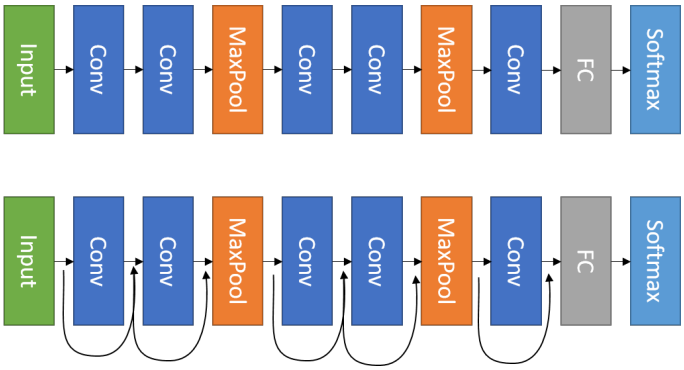


Figure 13: A sketch of a shallow CNN and its residual counterpart.

Another type of neural networks are **recurrent neural networks** (RNN). Here, there are several inputs, which are fed after each other to the network. The network itself forwards the hidden unit from one iteration of the network to the next. Thus, there are two inputs to the network at each stage: information from the previous stage and the new input. An illustration of the simplest RNN with one hidden unit is shown in Figure 14. RNNs are especially good for sequential information, such as natural language processing or artificial audio generation. One prominent building block of a RNN is the **long-short term memory**(LSTM) [96].

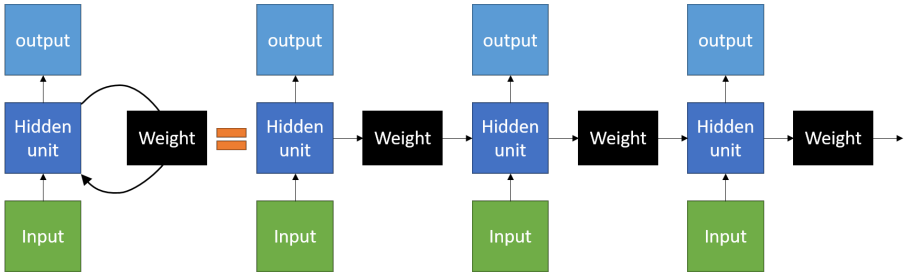


Figure 14: An illustration of a simple RNN. This is a slightly modified version from Goodfellow and colleagues) [96].

Deep auto-encoders (DAE) are a type of SNNs, which have the goal of finding a lower dimensional representation of the input and being able to restore the input from this low-dimensional representation. This can be used for example for denoising of speech [97] or images [98]. An illustration for the shape of an DAE is given in Figure 15.

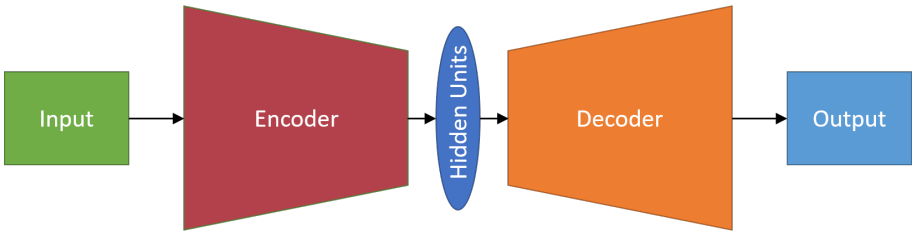


Figure 15: An illustration of an DAE.

Neural Segmentation Networks (NSN) typically use a structure similar to auto-encoders, in the sense that a dimensionality reduction is performed followed by a reconstruction network. Differences exist however in whether the spatial information is completely omitted such as in the anatomically constraint neural network (ACNN) [99], or if the spatial resolution is only reduced, as for example in the UNet [100]. The later one has been used for segmentation of CT images of pancreatic tumor [101], liver [102] and the Xenopus kidney [103]. However, those approaches either use a 2D-UNet or are in need of another neural network on top of the UNet. One important improvement to former CNNs are the skip connections, which preserve more detailed information for the segmentation at a later stage in the network [104]. The structure for a 3-level UNet is illustrated in Figure 16.

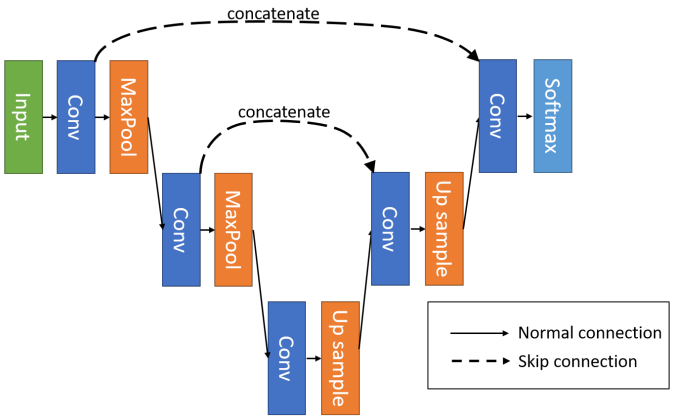


Figure 16: An illustration of a UNet3 similar to the one in Hänsch and colleagues (2018) [2].

In the approach by Hänsch and colleagues [2], a 3-level UNet was used to segment parts of the female pelvis. In this approach, the 3D images is processed in patches, meaning that the network is applied to a subset of the actual image, which then outputs a even smaller subset. Then the receptive field of the network is shifted and

the next output patch generated. As the network is translational invariant due to its fully convolutional structure, the network performs similar for the shifted input. Through overlapping receptive fields, as to say as smaller output size than input size, consistency in between the classification is tried to ensure.

3 Research Material and Methods

3.1 Model Architectures

The architecture of a model has several requirements. First, it needs to have enough capacity to be able to perform the task at hand. However, its capacity cannot be too high, such that overfitting becomes a problem. Second, the hardware limitations need to be considered, which include the memory available and the training time. Third, the size of the model can be constraint by the final application, meaning that there can be a time constraint in the inference.

In this work, an architecture was developed for different segmentation tasks and will be compared to a model created by Varian Medical Systems for research purposes and to the 3-level UNet model developed by Frauenhofer MEVIS.

3.1.1 5-level UNet

One problem of the 3-level UNet proposed by Frauenhofer [2] is the misclassification of muscle tissue as breast tissue, when applying it to CT scans of the upper female body. One possible reason for this is the fairly small receptive field of 44x44x44 compared to an image size of approximately 256x256x120. Therefore, the network is not able to comprehend the anatomical context of the image patch well.

In order to mitigate this problem, a 5-level UNet was developed similar to Ronneberger and colleagues [100]. The main difference to the Frauenhofer's approach is the input and output size of the network. In the 3-level UNet from Frauenhofer, the input size is 44x44x44 and the output size approximately 12x12x12. A smaller output size than input size is needed to achieve smoother contours. In the 5-level UNet, the input size is the full volume of interest (VOI); in the breast case 256x256x112. The output size can therefore be equal to the input size as no smoothening through sliding is needed when performing the inference on the whole volume.

In comparison to Ronneberger and colleagues [100], this 5-level UNet, in the following called UNet5, uses 3D images and replaces up-sampling layers with transpose convolution layers. The depth of the convolution layers also differ slightly, in order to stay within the memory boundaries. The filter size is constantly 3x3x3. The depth of the filters for the first level is 16 and increased by a factor of 2, when moving to a lower resolution level, to avoid bottlenecking. The transpose convolution use the same filter depth as the resolution level they are convolving from. Skip connections are implemented between each level in the downwards path to the corresponding level in the upwards path.

A sketch of the implemented architecture is shown in Figure 17. At each Conv block, there is a convolutional, a dropout, a batch normalization and an activation layer. The order of these layers is described as layer order and is hyper searched. The architecture uses 6,645,843 trainable parameters.

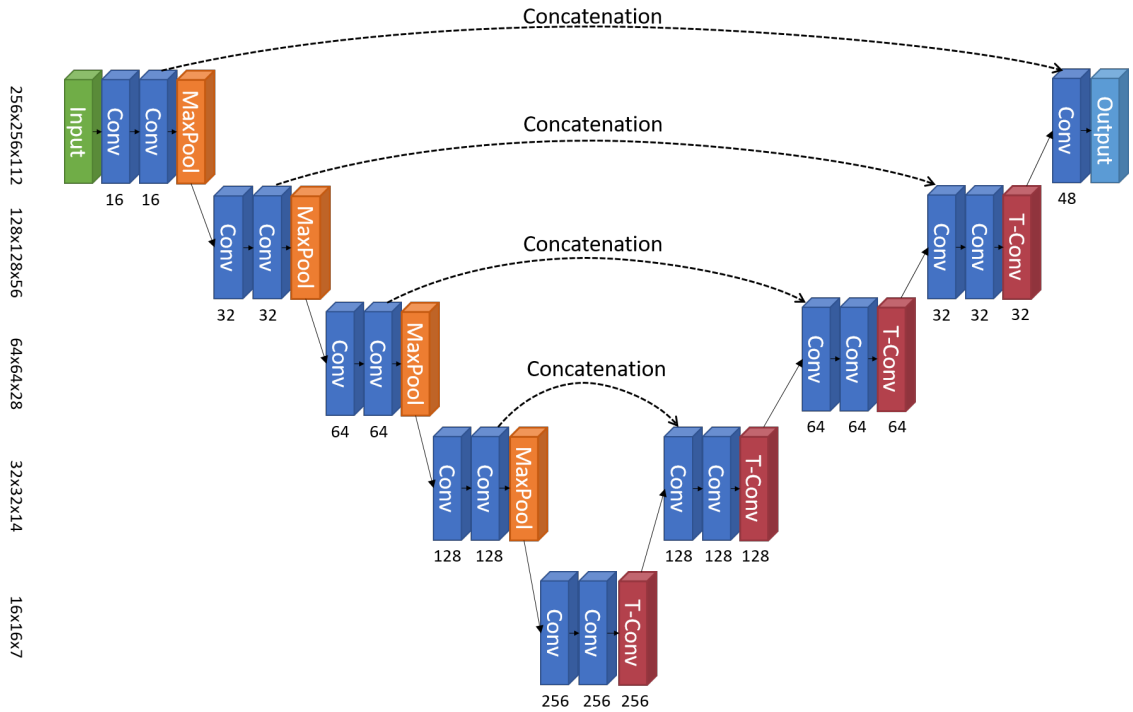


Figure 17: The UNet5 architecture with the different sizes of the image in the different levels and the number of filters for each (transpose) convolution.

3.1.2 BibNet

As an attempt to improve the performance of the UNet5, the BibNet was invented. Its structure is similar to the UNet5 but it tries to process image information at different resolution levels. The idea is to preserve high-resolution features and feed them to lower resolution levels. Therefore, a mesh like architecture is made. The structure can be seen in Figure 18. The convolutional filter have a size of 3x3x3, and a depth of 16 in the first resolution level, which is increased by a factor of 2 for each resolution level change to prevent bottlenecking. The parameter count is due to the extra intermediate layers increased in comparison to the UNet5 to 7,679,187 trainable parameters.

3.1.3 Slim BibNet

In an attempt to increase the inference speed and reduce the training time, the number of filters used per layer was reduced by a factor of 4 in comparison to the above presented BibNet. With this approach, the trainable parameters reduce significantly from roughly 7.7 million to 482,247.

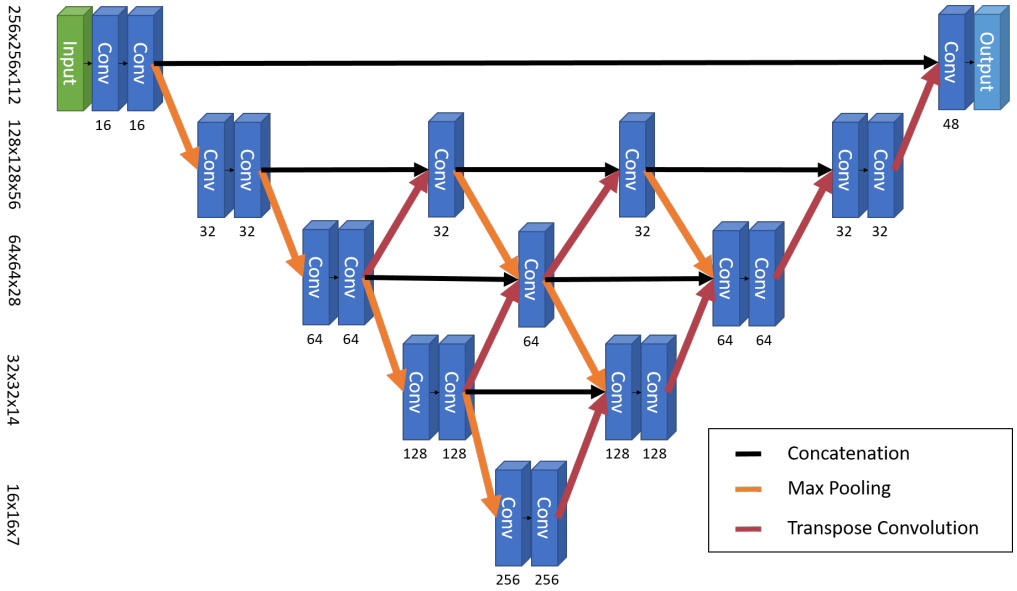


Figure 18: The BibNet architecture with the different sizes of the image in the different levels and the number of filters for each convolution.

3.1.4 BibNet2

Residual networks have shown that by making a network deeper its performance can be increased. This concept is applied in the further development of the BibNet. Herefore, the filter size per convolutional layer is kept from the slim BibNet, but each convolutional block in the downward path is made deeper using residual blocks. The architecture is sketched in Figure 19. In comparison to the original BibNet, the trainable parameters are decreased to 4,349,232. One reason for this is the reduction in the number of filters per convolutional layer. This helps to increase the training and inferring speed. However, the increased complexity compared to the slim BibNet should enable the network to retain its segmentation accuracy.

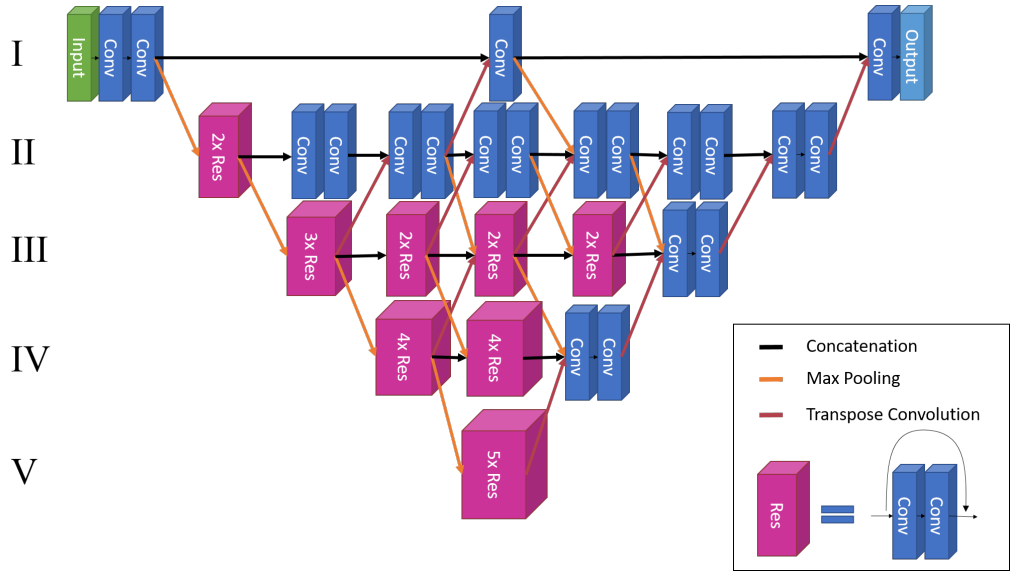


Figure 19: The BibNet2 architecture.

3.2 Dataset

The images together with the ground truth segmentation are from different hospitals. These segmentations were clinically used to create treatment plans for cancer patients.

3.2.1 Breast

The breast dataset contains 97 patients with contours for left breast, right breast and the heart. From this dataset, 8 patients are randomly chosen as a validation set. The data was reviewed during the course of the thesis, due to inconsistencies in the segmentation, such as missing slices or clearly wrong segmentations. This was done by Varian experts. In the following, it will be mentioned whether the revised or the original data has been used.

3.2.2 Female Pelvis

The initial dataset included 412 patients with ground truths for bladder, uterus and rectum. Out of this data set, 97 patients are taken for validation. The extended dataset has 565 patients with ground truths for bladder, uterus and rectum. Out of this, 92 patients are taken for validation.

3.2.3 Male Pelvis

The male pelvis dataset contains 150 CT images from 3 different clinical sites. From those images 150 pseudo CBCT (pCBCT) images are generated; one for each CT image. Thus, the final dataset consists of 150 CT images and 150 pCBCT images. The dataset is then randomly divided into 240 training images and 60 validation images. Each image has ground truth segmentations for the bladder, the prostate,

the rectum and the seminal vesicles. The pCBCT scans contain 67 slices, whereas the CT scans contain between 100 and 250 slices.

3.3 Computing platform

The training was performed on different local workstations with Nvidia Quadro graphic cards K80, P4000 and P5000. As a high level interface, a deep learning framework developed by Varian was used. It builds on top of Keras, which uses Tensorflow as a backend. Tensorflow itself then uses CUDA 9.0 and the corresponding cuDNN library.

3.4 Training Strategies

The training of the network is a crucial part for the outcoming performance. As the inference of a model uses less memory than the training, a different network size can be chosen for training and inference. The idea behind this is that fully convolutional networks share parameters for different image sizes. Therefore, by first using smaller patches of images for training, a larger network parameter-wise can be trained. The patch size is then increased during inference to match the actual image size. The limitation of this approach is that a high level representation, in the form of hidden units, can only be achieved upon the grade in which the inference task is similar to the training task. This means that when the input patch is small in training compared to inference, the network cannot perform well. This is because the task in the inference then differs greatly from the trained task. Therefore, the training in general is better if the patch size can also be increased during training. However, as the hardware capacities are limited, using smaller patches can improve the results by allowing the creation of higher capacity models.

In the training in this work, the patch size is maximized for the available graphics card. In the breast case, this means that for the training on a Nvidia Quadro P4000 with 8 GB GPU memory, the patch size is set to 256x256x32, whereas on a Nvidia Quadro P5000 with 16 GB memory, it is set to 256x256x64. In the inference, the patch size is increased to 256x256x80 on the P4000 and to 256x256x112 on the P5000. For the female pelvis, a patch size of 192x192x48 is used during training, due to the differing image size in the training set. During inference, the patch size is increased to 192x192x64. For the male pelvis dataset, the patch size during training is 256x256x32 on the P4000 and 256x256x64 on the P5000. During inference, the patch size is increased to 256x256x64 for the UNet5 and BibNet, and to 256x256x192 for the BibNet2. The large axial size in this case, is due to some CT images in the training set, which reach from the tigh to the chin.

Furthermore, a decaying learning rate is applied with an decrease from 10^{-3} to 10^{-5}. Additionally, a hyper search is performed for PReLU and ELU activations and different random seeds, which influence the weight initialization. As an optimizer, Adam is used.

The models are created using the python implementation of keras with tensorflow as a backend. The framework for data augmentation, 3D batch processing and model

evaluation is based on python.

For most of the models, a so-called hyper search is used. Here, several models are trained up until a specified number of epochs. These models differ in their hyper parameters such as activation functions and layer orders. Additionally, the random seed for the weight initialization is sampled for each model separately. In this approach, the best model after a few epochs is chosen and then trained until convergence. The underlying assumption here is that a model that performs well in the beginning will also perform well at the end of the training. However, this only holds true if the number of epochs needed for training are similar for the different models in the hyper search.

3.5 Training - Breast

3.5.1 UNet5 - Breast unrevised

The UNet5 was trained on the unrevised breast dataset using the breasts and the heart as segmentation targets. The images were downsampled by a factor of two. The loss function was the jaccard loss with an additional surface loss with weight 1. The learning rate during the hyper search is changed from 10^{-3} to 10^{-4} and afterwards from $5 \cdot 10^{-3}$ to $5 \cdot 10^{-5}$. During the training of UNet5, the graphic card was changed from a P4000 to a P5000 after 1000 epochs, which allowed increased the patch size from 256x256x32 to 256x256x64.

3.5.2 UNet5 - Breast revised

Similarly to Section 3.5.1, the UNet5 was trained. However this time on the reviewed breast dataset. The loss function was the jaccard loss without an additional surface loss. The learning rate during the hyper search is changed from 10^{-3} to 10^{-4} and during the auto train phase from $5 \cdot 10^{-3}$ to $5 \cdot 10^{-5}$. The hyper search searched through 6 different UNet5 with layer orders either dropout-convolution-activation-normalization or dropout-normalization-activation-convolution, activation either PReLU or ELU and varying random seed for the weight initialization.

During the following training period of the UNet5, the graphic card was changed from a P4000 to a P5000 after 465 epochs, which allowed increasing the patch size from 256x256x32 to 256x256x64.

3.5.3 UNet3

A 3-level UNet similar to Hänsch and colleagues [2] has been trained on both the unrevised and the revised breast data set, using a hyper search for activation functions and layer order. The model was trained until convergence.

3.5.4 BibNet2

In a similar fashion, the BibNet2 has been trained on the revised breast dataset. However, the layer order was kept fixed to dropout-convolution-activation-normalization and the activation to ELU. The model was then trained until convergence.

3.6 Training - Female Pelvis

The task for the female pelvis is to segment the rectum, bladder and uterus. Three different models have been trained for this: the UNet3, the UNet5 and a Varian in-house model. The training has been done twice: once for the initial dataset and once for the extended dataset.

3.6.1 UNet5 - initial Dataset

The training of the UNet5 on the initial dataset used a hyper search for activation functions, choosing between PReLU and ELU, layer order and random seed. 5 different models have started the training and the worst models have been dropped after 10 and 20 epochs. The best model was then trained until convergence, which was after 182 epochs. The training was performed on a P4000 with a patch size of 192x192x48.

3.6.2 UNet5 - extended Dataset

Similarly, the UNet5 has been trained on the extended pelvis data set with the difference of dropping the first two models at 10 epochs, one at 25 epoch and the fourth one after 50 epochs. The best model was then continued to train until convergence, which occurred after 237 epochs.

3.6.3 Other Models

In a similar way, a UNet3, a BibNet, a BibNet2 and the slim BibNet have been trained on the extended female pelvis dataset. For each model a hyper search was used. However, only for the UNet3 the layer order and activation function have been hyper searched. For the other ones, these have been fixed during the hyper search and, thus, only the random seed was modified. All of the models have been trained until convergence with changing the learning rate during training from $5 \cdot 10^{-3}$ to $5 \cdot 10^{-5}$.

3.7 Training - Male Pelvis

The task for the female pelvis is to segment the bladder, the prostate, the rectum and the seminal vesicles. Four different models have been trained on this anatomical site: the UNet3, the in-house model, the UNet5 and the BibNet2. For all of the models, the images were down sampled by a factor of 2 in the axial slices. The final resolution is therefore 256x256 for the CT scans and 197x197 for the pCBCT scans. All models used a hyper search in the beginning and were then trained until convergence.

3.7.1 BibNet2

For the BibNet2, the learning rate during the hyper search was changed from $5 \cdot 10^{-3}$ to $1 \cdot 10^{-4}$, and during the following training from $1 \cdot 10^{-4}$ to $1 \cdot 10^{-5}$. The total

training contained 300 epochs for the best model, which took less than one day.

3.7.2 UNet3

For the UNet3, the learning rate during the hyper search was changed from $1 \cdot 10^{-3}$ to $1 \cdot 10^{-4}$, and during the following training from $1 \cdot 10^{-4}$ to $1 \cdot 10^{-5}$. The total training contained 1050 epochs for the best model, which took less than one day.

3.7.3 Other Models

Similarly, a UNet5 and the Varian in-house model have been trained. Again, a hyper search was used for changing the random seed and the models have been trained until convergence.

3.8 Visualizing Filters

The convolutional layers are thought to represent higher level features the deeper in the network they are located. To verify this thought, the filters of the UNet5 trained on the revised breast data set, were visualized in the following way:

An image with random noise is fed into the network. Then, the output of the layer to be visualized is taken and its mean value calculated. This mean value is maximized by back propagating to the image using gradient ascent. Thus, the image is iteratively modified to maximize the output of the filter. In our case, a step size of 20,000 and 50 iterations have shown to be reasonable.

3.9 Visualizing Segmentation

For the visualization of the segmentation for the female pelvis, CT images are taken from the public dataset *The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma* collection (TCGA-UCEC) [105]. These are CT images from female subjects containing both the abdomen and pelvis, as well as the thorax. The resolution and window levels are slightly different from the training dataset and, thus, this visualization also provides a robustness test. In addition to the segmentation of the neural network, smoothening, hole filling and selection of the largest connected component as a post-processing (PP) is applied where indicated.

For the visualization of the segmentation for the male pelvis, CT images are taken from the public dataset *The Cancer Genome Atlas Colon Adenocarcinoma* collection (TCGA-COAD) [106]. The images used are from male subjects and contain either only the pelvis region or are full thorax images. Again the resolution and window levels are slightly different to the training data and the image quality compared to the CT images of the training data is worse. This is due to the age of the CT images: The training images were taken between 2014 and 2017, whereas the CT images used from the TCGA-COAD are from 1999.

3.10 Inference Time

The inference time was measured on a Nvidia P5000 GPU or, for the CPU test, two Intel Xeon CPU E5-2640 v4 with 2.4 GHz have been used. Here, only the pure inference time was taking into account, not including, for example, the loading of modules such as Tensorflow. The inference time was then divided by the number of voxels.

3.11 Distance Measurement

The dice score can be used easily as a loss function during training, as it calculates the error for each voxel. However, the score itself is not necessarily easy to understand. As the score is calculated as a percentage of voxel, which are outside of the union of the ground truth and the segmentation, large organs generally are able to perform better. A more meaningful measure is the average surface distance. Here, the distance between the surface of the two volumes is calculated. To account for both over and under contouring, in this work the root-mean-square (RMS) of the distance for each surface voxel of the segmentation is taken.

4 Results

4.1 Breast Models

In the following, the results for the segmentation in the breast case are presented and discussed.

4.1.1 UNet5 - Breast unrevised

The validation dice loss for the training of the UNet5 on the breast dataset is shown in Figure 20a. It can be seen that after approximately 150 epochs the validation loss is stagnating and does not improve until the patch size is increase at the change to a better graphics card. The two points from here are that the model is able to train within a couple of hundreds of epochs and that more training does not improve the validation loss. The second point is that the patch size during training changes the validation loss significantly. The first saturation was with a validation dice loss at 0.21 and the second at 0.10.

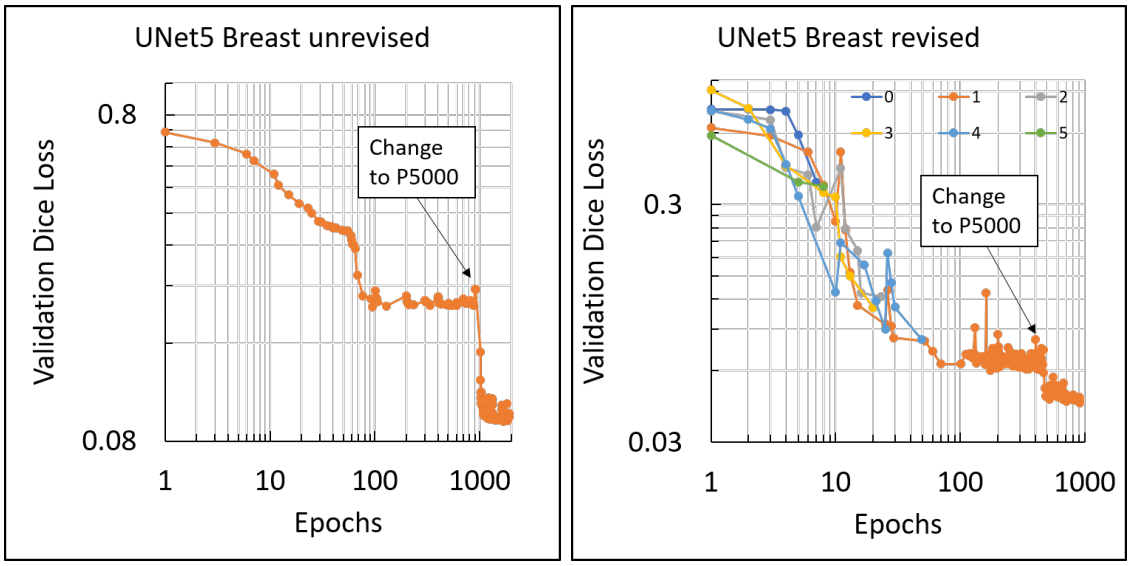
4.1.2 UNet5 - Breast revised

The validation dice loss for the training of the UNet5 on the revised breast dataset is shown in Figure 20b. The graphics card was changed after 465 epochs. Similar to the training on the unrevised training set, the validation loss decreases significantly after increasing the patch size. The model seems to be nearly unchanged after 100 epochs, leading to only minor improvements. The increased performance in respect to the validation dice loss after changing to a larger patch size is not necessarily an indicator for a better performance during inference, because the patch size is automatically increased to 256x256x112 during inference. Thus, the initial improvement may not be fully reflected during evaluation of the model. In final evaluation, the dice scores for the model trained only with the smaller patch size is less than 1% different to the model trained with the larger patch size for the two breasts. However, the score improved from a dice score of 0.821 to 0.941 for the heart. In visual inspection, it could be seen that the inferior heart boundary was not appropriate for the training with the smaller patch size. Here, the network tend to contour parts of the superior liver as part of the heart. For the larger patch size, this problem is mitigated.

4.1.3 Model Comparison

For both data sets, a UNet5, a UNet3 and a model developed previously for research purposes by Varian, from now on in-house model, have been trained on the revised breast data set. The evaluation is done on 8 patients.

The results on the revised data set are shown in Table 1. It can be seen that the Varian in-house model performs better for the right breast, whereas the UNet5 performs slightly better on the heart and right breast. On the other side, the time to compute the inference on one patient is significantly slower in the in-house model compared to the UNet5. This is due to the architectural design. The UNet5 processes



(a) Initial breast dataset (b) Revised breast dataset

Figure 20: The validation dice loss during the training of the UNet5 on the initial and revised female breast dataset, using the breasts and the heart as segmentation targets. The graphic card and with it the patch size was changed after 1000 epochs for the initial and 465 epochs for the revised dataset.

the whole image at once, whereas the in-house model has a patch-wise approach. Therefore, the in-house model needs to perform the inference several times for each patient, whereas it is needed only once for the UNet5. The training of the model on the revised dataset did improve the performance of the UNet5. This can be due to a more consistent dataset, making it easier for the model to learn the anatomical structures. The improvement is for all structures significant with a p-value of less than 0.005.

The BibNet2 performs equally well as the UNet5 on the heart, but is worse on the left breast and better on the right breast. Therefore, its overall performance is comparable to the UNet5 and the in-house model. However, its inference speed is the fastest of all on the GPU, reducing the inference time in comparison to the UNet5 by 25% and in comparison to the in-house model by a factor of 46.5. Therefore, as the scores show no clear difference, this model could be thought as being the best for this anatomical site.

The inference time per voxel and for a 256x256x112 sized image for both inference on GPU and on CPU can be seen in Table 2.

The results from the evaluation of the models on the unrevised data set are shown in Table 3. The models were trained and evaluated on the unrevised dataset, except for the 'UNet5 retrain', which was trained on the revised data set and evaluated on the unrevised set. Interesting to note is that the UNet5 improved for the breasts by training on the revised data set. This means that the model did improve not only for the revised data set but also for the non-revised one. Therefore, it can be concluded that the model was able to learn the anatomical structures better with

model	heart	left breast	right breast
BibNet2	0.955	0.926	**0.955**
In-house	0.953	0.939	0.944
UNet3	0.944	0.927	0.922
UNet5 old dataset	0.923	0.878	0.865
UNet5 P4000	0.917	0.821	0.939
UNet5 P5000	**0.958**	**0.942**	0.937

Table 1: The median dice scores for the different models evaluated on the revised breast data set

model	GPU voxel (ns)	GPU image (s)	CPU voxel (μs)	CPU image (s)
BibNet2	**76.3**	**0.6**	15.3	112.4
In-house	3555.4	26.1	167.8	1231.3
UNet3	1154.6	8.5	90.5	664.0
UNet5	105.6	0.8	**13.2**	**96.8**

Table 2: The median dice scores for the different models evaluated on the revised breast data set

model	heart	left breast	right breast
Varian in-house	0.954	0.906	**0.918**
UNet3	0.950	**0.922**	0.906
UNet5	0.951	0.893	0.882
UNet5 retrain	**0.956**	0.908	0.915

Table 3: Median dice scores for the different models evaluated on the unrevised breast data set and the median inference time.

the improved dataset. This indicates that the model did not only learn a certain contouring practice but learned the underlying anatomy.

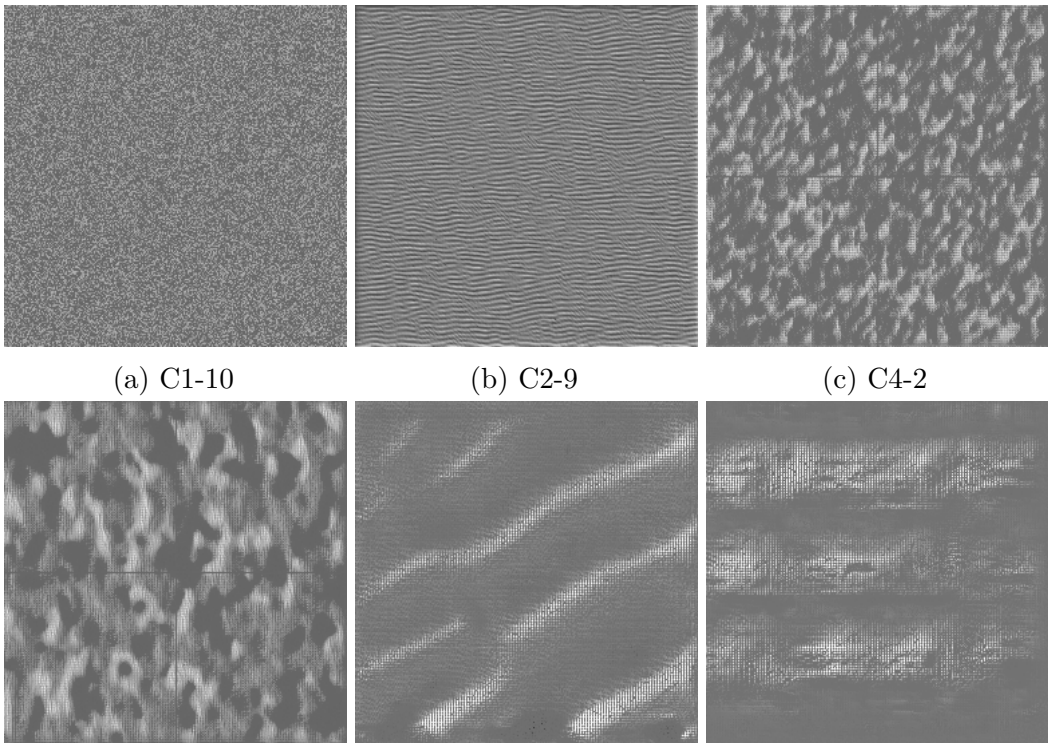
Score-wise, the UNet5 is for some anatomically structures worse than the UNet3 and, compared to the Varian in-house model, only better for the right breast. However, the inference time is smaller by a factor of 4 compared to the ResLane and the UNet3.

4.1.4 Visualizing Filters

Axial slices of some of the filters from the different layers are shown in Figure 21 and Figure 22. The naming of the layers is here C1–C14 for every block shown in Figure 17 on the downward path starting from the first convolutional layer, and E1–E12 for the layers in the upward path starting from the first transpose convolutional layer.

In the downward path, it can be seen that the first layer has no macroscopic shapes, whereas the second layer seems to have some pattern recognition. The deeper the layer is the more complex structures are detected by the layer in the downward path.

In the upward path, the layers respond to similar features as in the later convolu-

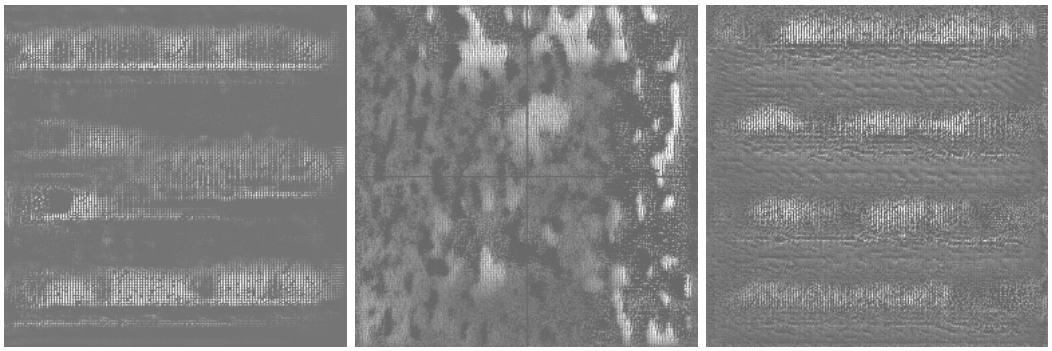


(a) C1-10 (b) C2-9 (c) C4-2

(d) C5-5 (e) C11-30 (f) C14-61

Figure 21: Axial slices from some of the filters of the convolutional layers of the downward path in the UNet5 trained on the breast data set.

tional layers of the downward path. Some of the layers, however, seem to represent more detailed patterns, for example filter 15 of layer E6. This is reasonable, as the upward path tries to reproduce the segmentation from the low resolution representation of the downward path. Therefore, some layers have the same macroscopic information as at the end of the downward path, whereas other layers are more receptive to the finer textures.

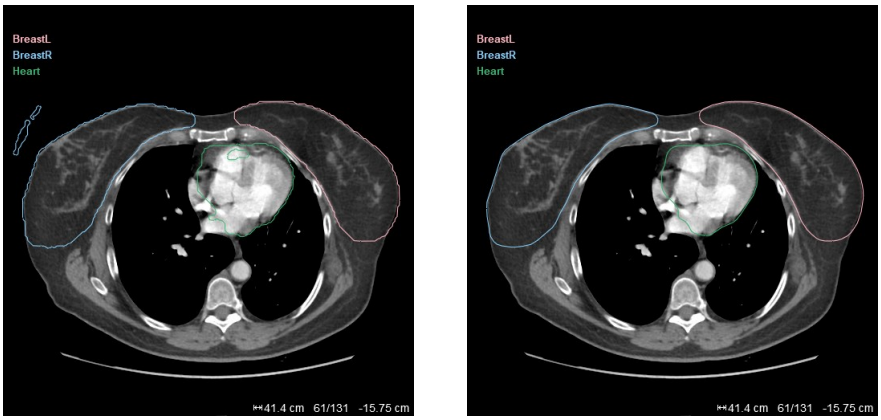


(a) E2-10 (b) E6-15 (c) E9-10

Figure 22: Axial slices from some of the filters of the convolutional layers in the upward path of the UNet5 trained on the breast data set.

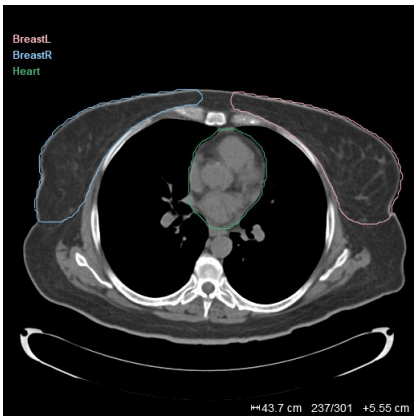
4.1.5 Visualizing Segmentation

The segmentation of the female breast and heart for three subjects using the UNet5 with and without PP is shown in Figure 23.

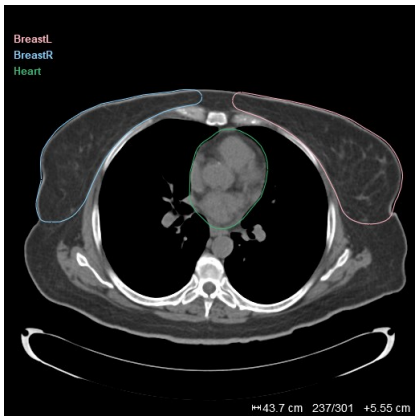


(a) Patient 1 without PP

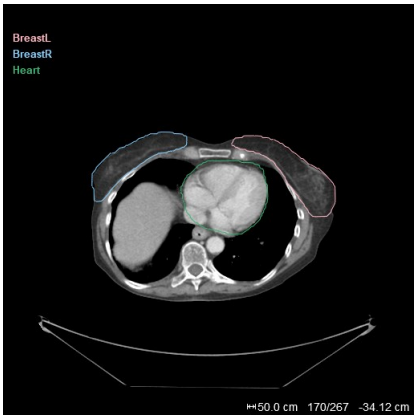
(b) Patient 1 with PP



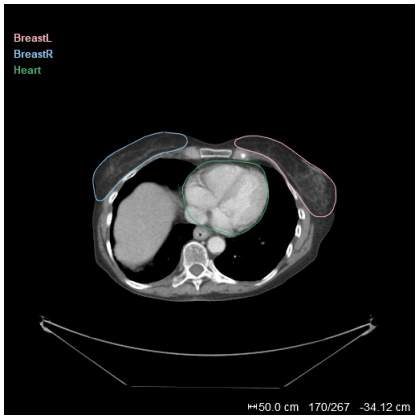
(c) Patient 2 without PP



(d) Patient 2 with PP



(e) Patient 3 without PP



(f) Patient 3 with PP

Figure 23: The segmentation of breasts and heart of three patients of the TCGA-UCEC with and without post-processing.

4.1.6 Contouring Problems

The UNet3 in the breast case has a tendency of misclassifying muscle tissue in the back and in the arm as breast tissue. The reason for this is that the tissues' appearance is similar and can only be distinguished by its context. As the receptive field of the UNet3 is small, it is unable to distinguish between the different areas. This problem can be mitigated through post-processing by taking the biggest connected component. It could be seen that this is able to reduce the misclassification as long as the misclassified tissue is separated and forms a smaller connected component as the actual tissue.

The UNet5 and the BibNet have similar challenges. These are mostly contouring a few slices too many in the superior breast or a few slices too many in the inferior or superior heart.

4.2 Female Pelvis

In the following, the results for the segmentation of the female pelvis are shown. The goal here was to segment the bladder, rectum and uterus from CT and CBCT scans.

4.2.1 Training UNet5 - initial Dataset

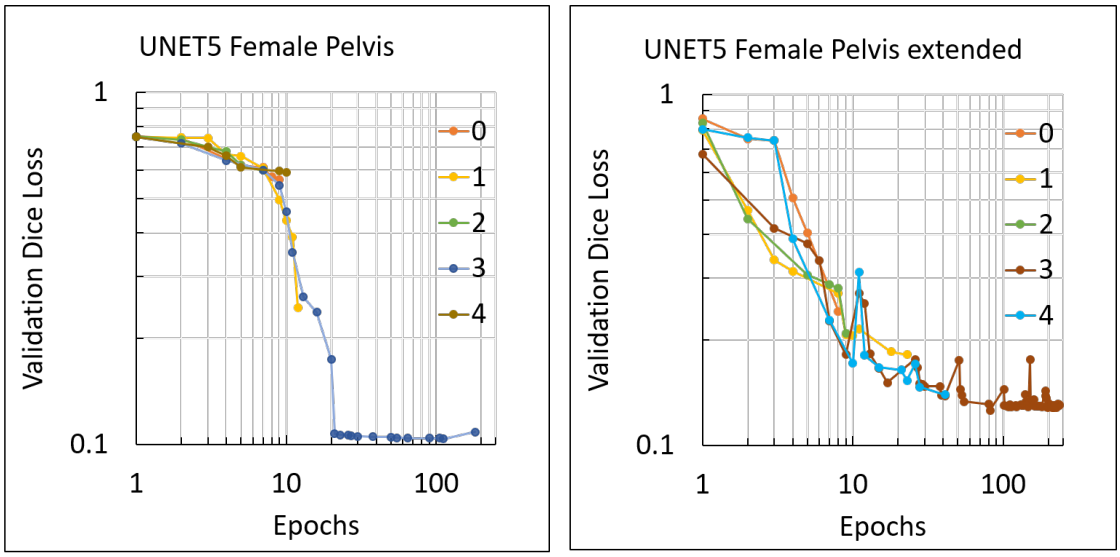
The validation dice loss of the UNet5 on the initial dataset during the training is shown in Figure 24a.

The validation dice loss decreases for the first 20 epochs strongly but then remains between 0.10 and 0.11 for the rest of the training. In this training, the different models, which are initially created, seem to perform equally, making the value of the hyper search questionable.

4.2.2 Training UNet5 - extended Dataset

The validation dice loss of the UNet5 on the extended dataset during the training is shown in Figure 24b.

In this training, the different models, which are initially created, have different validation losses during the first couple of epochs but seem to be fairly equivalent at 10 epochs. The final validation dice score is around 0.13, which is higher than on the non-extended dataset. This might be due to different validation data, which could be harder for the model to learn.



(a) initial female pelvis dataset (b) extended female pelvis dataset

Figure 24: Shown are the validation dice losses during the training of the UNet5 on the initial and extended female pelvis dataset, using rectum, bladder and uterus as segmentation targets.

4.2.3 Model Comparison

The median dice scores for the different models on the initial female pelvis data set are shown in Table 4.

model	Bladder	Rectum	Uterus
BibNet	**0.920**	0.816	**0.876**
Varian in-house	0.90	0.80	0.85
UNet3	0.916	0.806	0.843
UNet5	**0.920**	**0.829**	0.862

Table 4: Shown are the median dice scores for the different models evaluated on the initial female pelvis data set and the median inference time.

The inference times per voxel and for a full pelvic CBCT (192x192x64) on both GPU and CPU are given in Table 5.

model	GPU voxel (ns)	GPU image (s)	CPU voxel (μs)	CPU image (s)
In-house	2193.25	5.17	225.3	513.55
UNet3	1272.7	3.00	101.6	240
BibNet	126.0	0.30	15.6	36.8
UNet5	105.6	0.25	13.0	30.8
BibNet2	76.3	0.18	15.3	36.12
slim BibNet	**39.6**	**0.09**	**5.6**	**13.16**

Table 5: Shown are the inference times per voxel and for a full image both for GPU and CPU.

It can be seen that the UNet5 and the BibNet outperform both the Varian in-house model as well as the UNet3. Furthermore, their inference times are considerably shorter. One downside to both models is that the training of these models takes longer as each epoch takes between 500s and 1000s compared to 4s-30s for the Varian in-house model and the UNet3. However, this difference needs to be in perspective of the epochs needed for training. The Varian in-house model and the UNet3 usually require 2000 to 3000 epochs, compared to the approximately 200 epochs for the BibNet and UNet5. The reason for this is, that the later two train on the whole image, whereas the former two train only on a small patch. Therefore, each epoch for UNet5 and BibNet takes longer but allows the model to learn more. The general training time is in the order of one day to a couple of days.

On the extended dataset, the UNet3 improved, whereas the BibNet and UNet5 stayed approximately the same. The results can be seen in Figure 25 and the median dice scores are shown in Table 6. Here, the BibNet is as good as the Varian in-house model, but is outperformed by the UNet3. Furthermore, the UNet5 performs in this test worse on the rectum compared to the non-extended dataset. This shows one problem in training DNNs: The training process itself is stochastic and, thus, it cannot be assured that two trainings converge to the same minimum or have the exact

same performance. By training the model on three different organs simultaneously, it forces the model to have a trade-off between performance on each one of them. The trade-off itself is, hereby, statistically learned through the applied loss functions.

model	Bladder	Rectum	Uterus	overall
BibNet	0.895	0.809	0.878	0.848
slim BibNet	0.871	0.787	0.850	0.804
BibNet2	0.890	0.779	0.864	0.843
Varian in-house	0.897	0.803	0.879	0.855
UNet3	**0.918**	**0.812**	**0.892**	**0.873**
UNet5	0.903	0.782	0.854	0.851

Table 6: The median dice scores for the different models evaluated on the extended female pelvis data set.

When performing a two-tailed paired t-test, the differences in the overall mean values presented in Table 6 are statistically significant with a p-value of less than 5% , except for the overall means between UNet5, BibNet and BibNet2. This is reasonable as those architectures are related to each other. In addition, when using Bonferroni's correction, the differences between the in-house model and the UNet5, the UNet3 and the BibNet become statistically not significant. However, when comparing single structures, several comparisons are statistically not significant, e.g. for the uterus between UNet5 and BibNet, and between in-house and UNet3. Additionally, for the rectum between in-house and UNet3, as well as between in-house and BibNet. For the bladder, the difference is not significant between in-house and BibNet. This shows how similar the performances of the individual models are compared with the inter-individual variety.

The slim BibNet performs worse than the other models on average but is still comparable to the UNet5 for rectum and uterus. This result is not surprising, as the slim BibNet is a significantly reduced model. On the other hand, the inference time is reduced by a factor of 2 to 3 in comparison to the UNet5 and BibNet on both the CPU and GPU. The use case for this model is, hence, when the compute power poses a limitation for the other models. This can be the case if the inference needs to be performed on a normal CPU. Here the difference between the in-house model and the slim BibNet is significant: The in-house model takes 8.5 minutes, whereas the slim BibNet only takes 13.2 seconds. This can be a strong advantage, when the model is used in an adaptive treatment setting, where time is a crucial factor. It also needs to be considered, that here 2 Intel Xeon CPU E5-2640 v4 have been used. This might be a faster configuration than might be found in a hospital environment. The inference time there might even be slower, such that the time consumption poses an argument for the BibNet-4, even though its accuracy is slightly worse than for the other models.

The root-mean-square of the surface distance has been calculated for the different models and organs. A boxplot of the distances is shown in Figure 26. The results correlate with the dice scores, when it comes to the comparison of the different models. The median rms surface distance is here between 2 and 5 mm for the different models.

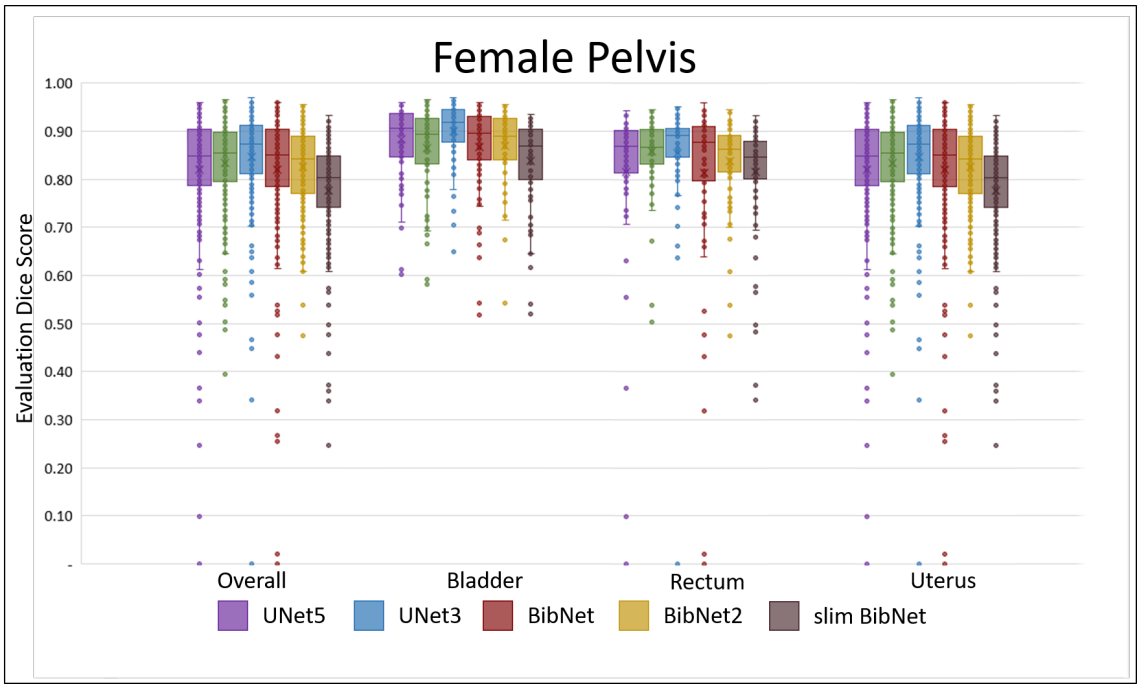


Figure 25: Box plots for each of five models both for all organs combined and for each organ separately for the extended female pelvis data set.

Here, the BibNet performs best on the bladder, whereas, the UNet3 is best on the Rectum and Uterus.

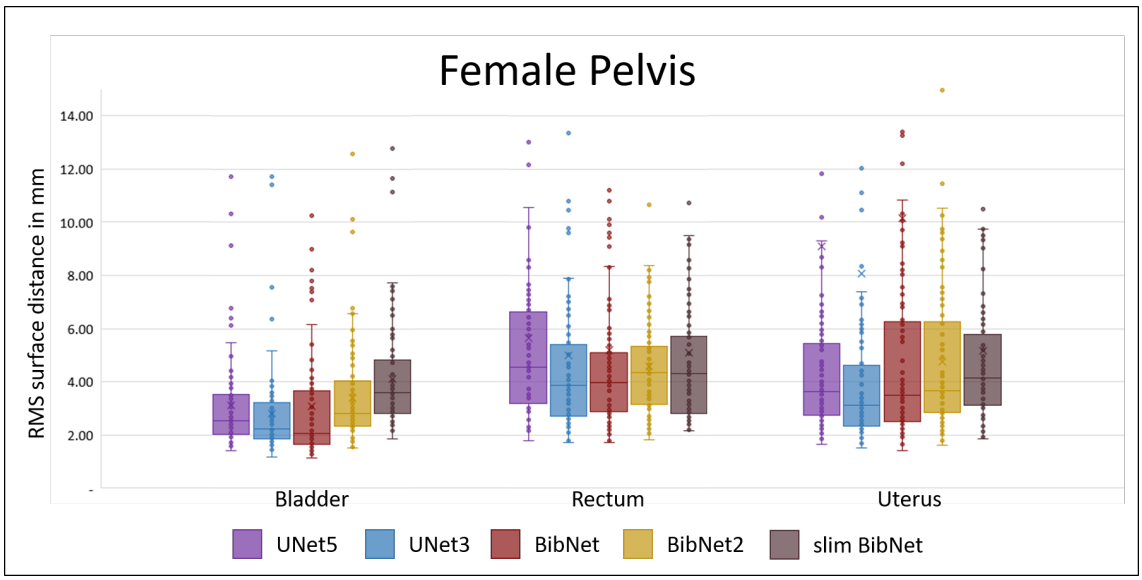


Figure 26: Box plots of the RMS surface distance for each of the six models both for all organs combined and for each organ separately for the extended female pelvis data set.

The inference of the UNet3 and BibNet trained on the extended data set and

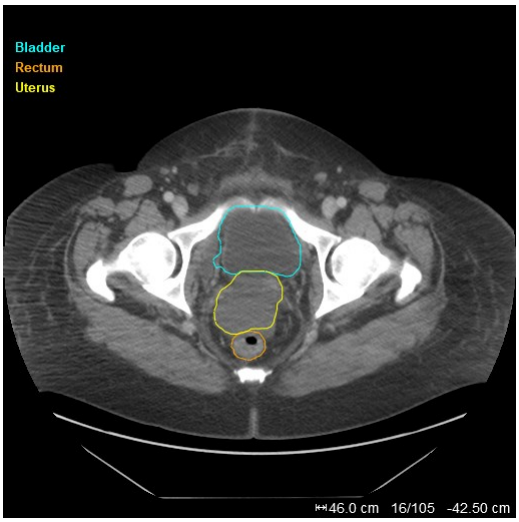
of the UNet5 trained on the non-extended data set is shown for an axial slice in Figure 27. For each model, the raw output of it and with applied post-processing is shown. The post-processing takes the largest connected component for each organ and applies gaussian smoothening. For UNet3, the uterus is not segmented after the post-processing because the largest connected component was found in the heart. This could be avoided by selecting an appropriate volume of interest.

4.2.4 Contouring Problems

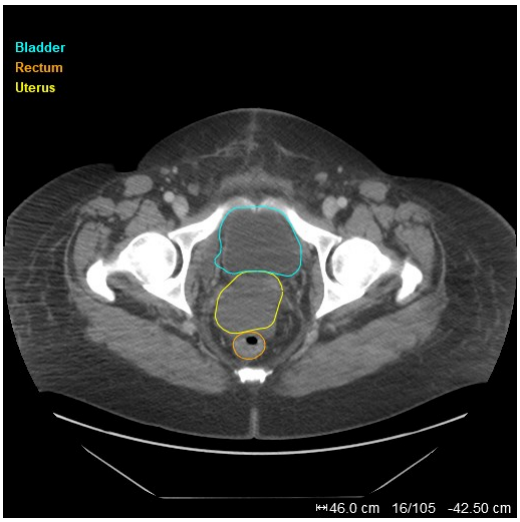
The contouring of the female pelvis is mainly affected by the artifacts stemming from the bowel movement. These occur mainly in the superior part of the abdomen and, thus, influence the superior uterus. All of the tested networks tend to stop contouring once the artifacts become too strong.

Another problem is that the contouring practices of the rectum are varying in the ground truth. Especially, the point at which the rectum starts or ends is not uniform in the training set. Therefore, the models cannot learn a specific contouring practice. This is a reason for the low scores of the rectum in comparison to bladder and uterus, which are fairly accurately described.

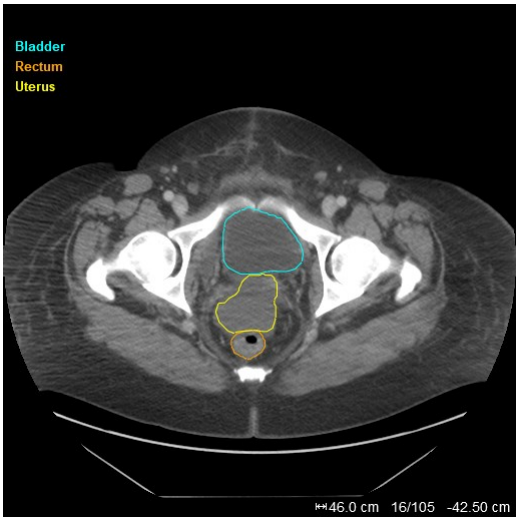
For the uterus, the main problem are artifacts, which occur in the superior part. Due to the bowel movement, there are movement artifacts at the superior uterus. This can make the uterus indistinguishable from surrounding tissue. The contours from all models seem to avoid the artifacts and, thus, end the contours to early. Reducing the artifacts could possibly be integrated at different levels. The reconstruction algorithm could be improved, e.g. using iterative reconstruction methods, which are more robust to motion artifacts. Furthermore, filtering might be able to reduce the artifacts. The effect of filters is questionable, however, due to the shape of the artifacts. Another option is the implementation of an anatomically constrained model, which is able to produce the contours despite the artifacts. This has been done for the heart by Oktay and colleagues [99], but requires a significantly more difficult construction and training of the model. Furthermore, in comparison to the heart, the pelvis region is more prone to anatomical changes, e.g. due to a more or less filled bladder and bowel movement. Therefore, it is not clear whether an ACNN would be able to generate good contours. In a first attempt to create the auto-encoder part of the ACNN with the same architecture as presented by Oktay and colleagues [99], the model was not able to learn a representation with a dice loss of less than 0.28, even though the parameter size of the hidden unit was searched between 64 and 4098. This might be an indication that the model is unable to learn the strongly changing anatomical structure of the female pelvis. However, further investigations are needed to confirm this preliminary result.



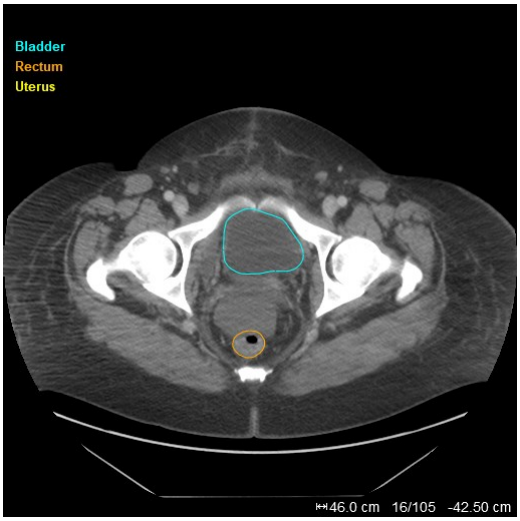
(a) BibNet without PP



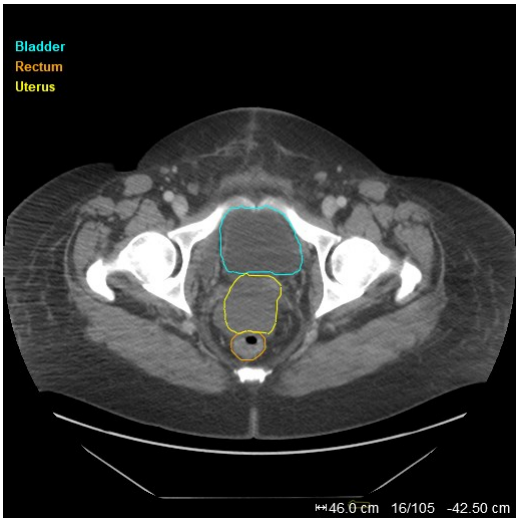
(b) BibNet with PP



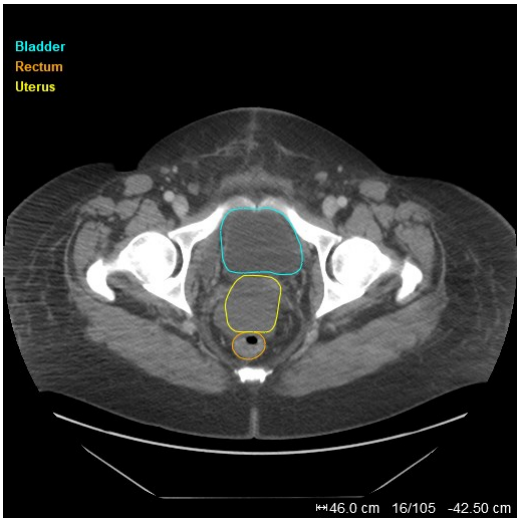
(c) UNet3 without PP



(d) UNet3 with PP



(e) UNet5 without PP



(f) UNet5 with PP

Figure 27: The segmentation of bladder, rectum and uterus from BibNet, UNet3 and UNet5 on a patient from the TCGA-UCEC with and without post-processing.

4.3 Male Pelvis

In the following, the results from the different models on the male pelvis are shown. The median values can be seen in Table 7.

model	Bladder	Prostate	Rectum	Seminal Vesicles
BibNet2	0.956	**0.886**	0.878	0.716
in-house	**0.962**	0.873	0.856	0.745
UNet3	0.961	0.879	0.866	**0.776**
UNet5	0.952	0.871	**0.880**	0.747

Table 7: The median dice scores for the different models evaluated on the male pelvis data set.

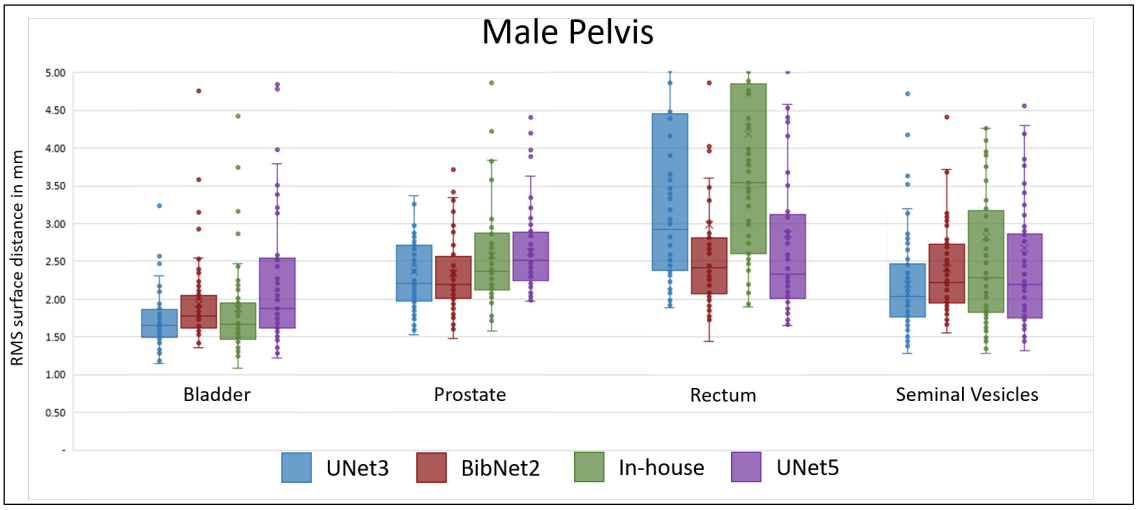


Figure 28: The average surface distances for each of the four model both for each organ separately for the male pelvis in millimeter.

A paired two sample t-test for means was performed between the BibNet2 and the UNet3, showing no statistical difference for Bladder and Prostate, but statistical significant difference for rectum and seminal vesicles. Thus, it can be concluded that the models perform with similar accuracy for the first two organs but have different strengths for the rectum and the seminal vesicles. In the evaluation of the UNet3, in two CT cases each, the bladder was contoured in the liver and the rectum was contoured in the bowel. The reason for this is the receptive field of the UNet3, which is small compared to the image size. Therefore, the context of each patch is unclear to the network. In postprocessing, the largest connected component is taken, such that smaller misclassifications are unnoticed. However, in whole torso CT scans, which are partly used for evaluation, the largest connected component is not necessarily the desired organ anymore. This problem could be mitigated by choosing an appropriate volume of interest (VOI). Due to this behavior, the mean value of the dice score for the bladder and the rectum are comparably large for the

UNet3. The receptive field of the BibNet2 is with 256x256x192 large enough, such that the whole image is processed in either one or two patches. Therefore, enough context is available to the network, such that these misclassifications do not occur.

In addition to the dice score, the root-mean-square (rms) of the surface distance between the segmentation and the ground truth for each organ was calculated. The results per patient are displayed in Figure 28. Here, the UNet3 is able to perform best for bladder and seminal vesicles. Whereas, the 20 times faster BibNet 2 is better on the Prostate and the Rectum. However, only the difference in the seminal vesicles is statistically significant (p-value < 0.05) in a two-way t-test for the mean value.

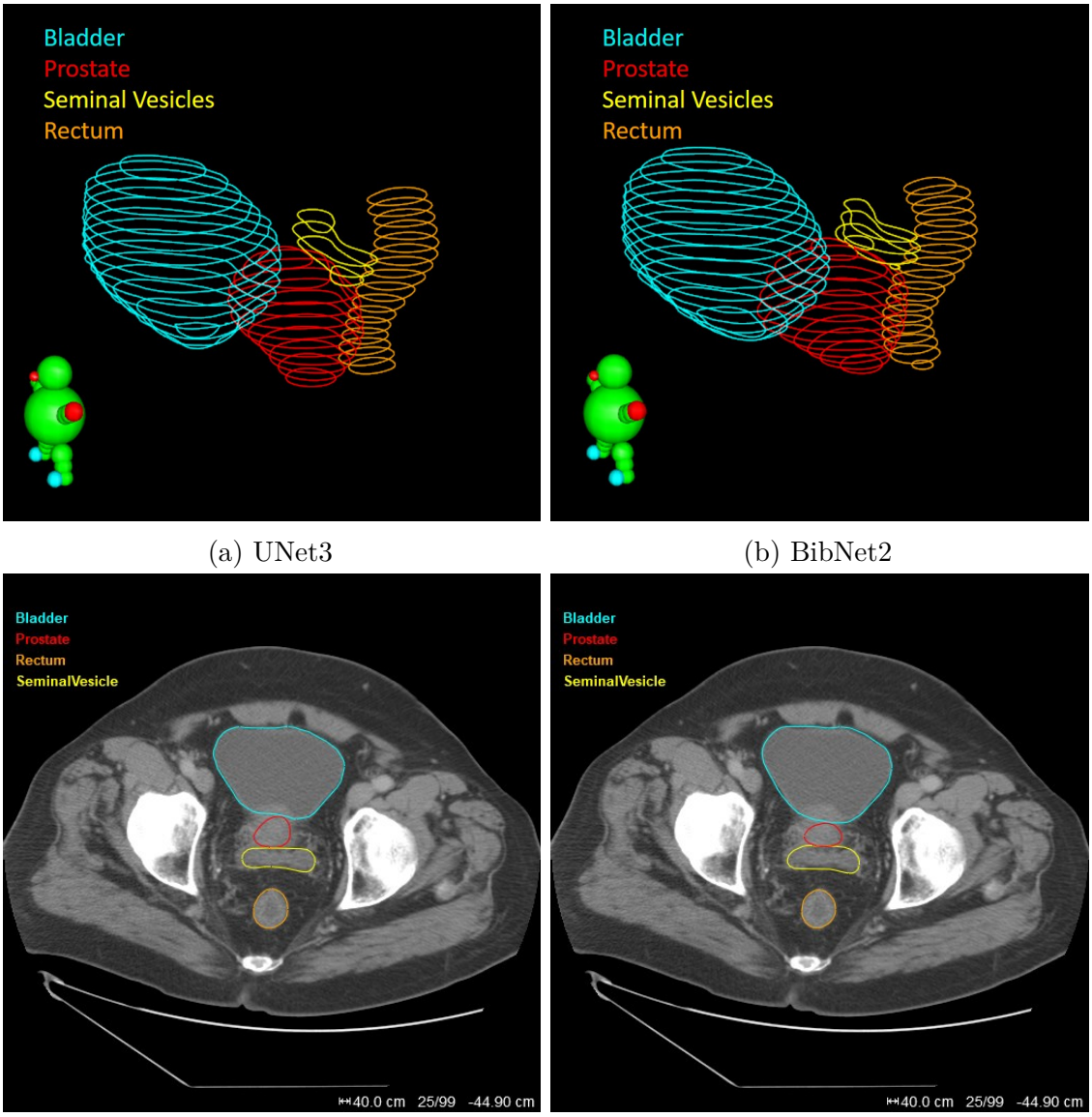
All models are able to segement the structures with an average error of less than 5 mm. The voxel spacing of the input image is 2.34 mm in the axial plane and 2.5 mm along the body axis. Thus, the median segmentation accuracy is between 1 and 2 voxels. For the bladder and the seminal vesicles, the median segmenatation accuracy is for some of the models even below the size of one voxel.

4.4 Contouring Problems

The contouring accuracy in terms of the root-mean-square of the surface distance is better for the male pelvis compared to the female pelvis. The main reason for this is the type of imaging data used. While the female pelvis data contains both CT and CBCT images, the male pelvis dataset is constituted of CT images and pCBCT images from these CT image. These pCBCT images do not contain the same artifacts from bowel movement as the real CBCT images. Therefore, the contouring problem, seen in with the female pelvis, in the lower abdomen due to these artifacts does not occur in the male pelvis case. Hence, the models are able to contour the superior part of the organs more accurately than in the female cases.

4.4.1 Visualizing Segmentation

In Figure 29, the segmentation from the UNet3 and BibNet2 with simple PP is shown for one axial slice in the male pelvis. The slice was chosen in such a way that all four structures are visible. In addition to that, the three dimensional structures are shown. As can be seen, the segmentations only differ slightly. For instance, the prostate is contoured more anterior in the UNet3 segmentation compared to the BibNet2. This coincides well with the similarity in scores.



(a) UNet3 (b) BibNet2

(c) UNet3 (d) BibNet2

Figure 29: The segmentation of a male pelvis by the UNet3 and the BibNet2. Simple PP has been applied.

5 Discussion

In this work, a 5-level UNet and modified versions of it have been constructed and applied for segmenting the female breast and pelvis, and the male pelvis. Herein, the segmentation accuracy is similar or better than the Varian in-house solution for all the sites and better in the breast case in comparison to the 3-level UNet proposed by Hänsch and colleagues (2018). The inference time is for both UNet5 and BibNet 10 times faster than the 3-level UNet and 20 times faster than the Varian in-house solution. The improved version BibNet2 is a factor of 15 faster than the 3-level Unet and a factor of 45 faster than the Varian in-house solution.

It has been shown for the breast case that improvements in the human drawn image contours translate to better accuracy of the model. In the pelvis case, the increase in the amount of training data, did not seem to improve the training results, which might be due to inconsistencies within the training data, such as different contouring guidelines and the difference in CT and CBCT images. Hence, it seems that good quality data are able to improve models, whereas an increase in mediocre quality data might not improve the outcome. On the other side, one consideration might be to generate slightly different models for different contouring practices or image modalities. This would increase the coherence within the train data and, thus, improve the training.

The segmenting process in the female pelvis is deteriorated due to the artifacts caused by bowel movements. This especially influences the accuracy in the superior uterus. For improving the results, either an anatomically constraint neural network could be tried or an iterative CBCT reconstruction might be able to mitigate the movement artifacts. The downside of the ACNN is its complex training process, which makes fast model development unfeasible. Furthermore, due to its fully connected layer, it is unable to scale to different input sizes.

Another common mistake for the model is to extend the organ for a few slices too long or a too short. An improvement could be seen in the breast case by training the model for more epochs, even though the validation loss within the training did not change. However, this effect cannot be guaranteed and is also affected by differing contouring standards from different clinics. Thus, either models adapted to each clinical practice are needed, or the correction effort for the clinician must be small enough to justify the use of the neural network in comparison to segmenting by hand.

In future work, the accuracy could possibly be improved through using larger patch sizes during training on graphic cards which provide 24 GB or 36 GB of memory, such as the Nvidia P40 or V100. In addition to that, refactoring of the training data or extending the training data could improve the results.

The results of this work seem to generalize over the three anatomical sites, which have been used. Thus, it gives an indication that the model architectures can be applied also to different anatomical sites. This might generate a great impact in the work flow of the radiation therapy planning, as the manual segmenting part might be replaced by an automatic segmentation. Hence, the time from diagnosis to treatment for each patient can be reduced, which improves the therapy outcome. Furthermore, the reduced work time for the segmentation reduces costs for the hospitals. In

addition to that, a possible application is in regions where less radiologists are available, such as India and Africa. Through automation of the segmentation, the hospitals will be able to treat more patients with the same number of radiologist and, thus, improving the quality of life for people in the developing world.

References

[1] "Causes of death statistics," *eurostat*, 2018.

[2] A. Hänsch, V. Dicken, T. Grass, T. Morgas, J. Klein, H. Meine, and H. K. Hahn, "Deep learning based segmentation of organs of the female pelvis in cbct scans for adaptive radiotherapy using ct and cbct data," *Computer Assisted Radiology and Surgery - CARS 2018, Berlin*, p. 133, 2018.

[3] A. J. F. Griffiths, S. R. Wessler, S. B. Carroll, and J. Doebley, *Introduction to Genetic Analysis*. W. H. Freeman & Company, 11 ed., 2015.

[4] G. Almouzni, S. Khochbin, S. Dimitrov, and A. P. Wolffe, "Histone acetylation influences both gene expression and development of xenopus laevis," *Developmental Biology*, vol. 165, no. 2, pp. 654–669, 1994.

[5] F. Bernardi and J. Ninio, "The accuracy of dna replication," *Biochimie*, vol. 60, no. 10, pp. 1083–1095, 1979.

[6] D. N. Cooper and H. Youssoufian, "The cpg dinucleotide and human genetic disease," *Human Genetics*, vol. 78, pp. 151–155, 1988.

[7] D. N. Cooper, M. Mort, P. D. Stenson, E. V. Ball, and N. A. Chuzhanova, "Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in cpnpg trinucleotides, as well as in cpg dinucleotides," *Human Genomics*, vol. 4, no. 6, p. 406, 2010.

[8] W. C. Burhans and M. Weinberger, "Dna replication stress, genome instability and aging," *Nucleic Acids Research*, vol. 35, no. 22, pp. 7545–7556, 2007.

[9] M. R. Lieber, "The mechanism of double-strand dna break repair by the nonhomologous dna end-joining pathway," *Annual Review of Biochemistry*, vol. 79, no. 1, pp. 181–211, 2010.

[10] H. Maacke, K. Jost, S. Opitz, S. Miska, Y. Yuan, L. Hasselbach, J. Lüttges, H. Kalthoff, and H.-W. Stürzbecher, "Dna repair and recombination factor rad51 is over-expressed in human pancreatic adenocarcinoma," *Nature Oncogene*, vol. 19, pp. 2791–2795, 2000.

[11] P. P. Connell, K. Jayathilaka, D. J. Haraf, R. R. Weichselbaum, E. E. Vokes, and M. W. Lingen, "Pilot study examining tumor expression of rad51 and clinical outcomes in human head cancers," *International Journal of Oncology*, pp. 1113–1119, 2006.

[12] A. Mitra, C. Jameson, Y. Barbachano, L. Sanchez, Z. Kote-Jarai, S. Peock, N. Sodha, E. Bancroft, A. Fletcher, C. Cooper, D. Easton, R. Eeles, and C. S. Foster, "Overexpression of rad51 occurs in aggressive prostatic cancer," *Histopathology*, vol. 55, no. 6, pp. 696–704, 2009.

[13] G.-B. Qiao, Y.-L. Wu, X.-N. Yang, W.-Z. Zhong, D. Xie, X.-Y. Guan, D. Fischer, H.-C. Kolberg, S. Kruger, and H.-W. Stuerzbecher, "High-level expression of rad51 is an independent prognostic marker of survival in non-small-cell lung cancer patients," *British Journal Of Cancer*, vol. 93, pp. 137–143, 2005.

[14] L. Yong, Y. Hui, L. Rong-Zhen, Z. Ying, Z. Mei-Fang, W. Xin, and J. Wei-Hua, "Elevated expression of rad51 is correlated with decreased survival in resectable esophageal squamous cell carcinoma," *Journal of Surgical Oncology*, vol. 104, no. 6, pp. 617–622, 2011.

[15] R. Barbano, M. Copetti, G. Perrone, V. Pazienza, L. A. Muscarella, T. Balsamo, C. T. Storlazzi, M. Ripoli, M. Rinaldi, V. M. Valori, T. P. Latiano, E. Maiello, P. Stanziale, M. Carella, A. Mangia, F. Pellegrini, M. Bisceglia, A. O. Muda, V. Altomare, R. Murgo, V. M. Fazio, and P. Parrella, "High rad51 mrna expression characterize estrogen receptor-positive/progesteron receptor-negative breast cancer and is associated with patient's outcome," *International Journal of Cancer*, vol. 129, no. 3, pp. 536–545, 2010.

[16] H. Maacke, S. Opitz, K. Jost, W. Hamdorf, W. Henning, S. Krüger, A. C. Feller, A. Lopens, K. Diedrich, E. Schwinger, and H.-W. Stürzbecher, "Over-expression of wild-type rad51 correlates with histological grading of invasive ductal breast cancer," *International Journal of Cancer*, vol. 88, no. 6, pp. 907–913, 2000.

[17] F. Couch, H. Shimelis, S. Hart, R. Moore, A. Thomas, G. Lipton, and E. Iversen, "Cancer risks and response to targeted therapy associated with brca2 variants of uncertain significance," *American Association for Cancer Research*, p. 78, 2018.

[18] J. L. Wood and J. Chen, "Dna damage sensing and signaling," in *DNA Repair, Genetic Instability, and Cancer*, ch. 1, pp. 1–22, World Scientific, 2007.

[19] A. Renwick, D. Thompson, S. Seal, P. Kelly, T. Chagtai, M. Ahmed, B. North, H. Jayatilake, R. Barfoot, K. Spanova, L. McGuffog, D. G. Evans, D. Eccles, D. F. Easton, M. R. Stratton, T. B. C. S. C. (UK), and N. Rahman, "Atm mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles," *Nature Genetics*, vol. 38, pp. 873–875, 2006.

[20] National Cancer Institute, "Differentiation," *NCI Dictionary of Cancer Terms*, 2018.

[21] R. V. Hutter, "The role of the pathologist in the management of breast cancer," *CA: A Cancer Journal for Clinicians*, vol. 41, no. 5, pp. 283–299, 1991.

[22] M. Amin, S. Edge, F. Greene, D. Byrd, R. Brookland, M. Washington, J. Gershenwald, C. Compton, K. Hess, D. Sullivan, J. Jessup, J. Brierley, L. Gaspar, R. Schilsky, C. Balch, D. Winchester, E. Asare, M. Madera, D. Gress, and L. Meyer, *AJCC Cancer Staging Manual*. Springer, 2017.

[23] M. Bower and J. Waxman, *Oncology - Lecture Notes*. John Wiley & Sons, 2011.

[24] R. Wagland, A. Richardson, J. Armes, M. Hankins, E. Lennan, and P. Griffiths, "Treatment-related problems experienced by cancer patients undergoing chemotherapy: a scoping review," *European Journal of Cancer Care*, vol. 24, no. 5, pp. 605–617, 2015.

[25] D. Coluccia, J. Fandino, L. Schwyzer, R. O'Gorman, L. Remonda, J. Anon, E. Martin, and B. Werner, "First noninvasive thermal ablation of a brain tumor with mr-guided focusedultrasound," *Journal of Therapeutic Ultrasound*, vol. 2, no. 1, p. 17, 2014.

[26] L. G. Merckel, F. M. Knuttel, R. Deckers, T. van Dalen, G. Schubert, N. H. G. M. Peters, T. Weits, P. J. van Diest, W. P. T. M. Mali, P. H. H. B. Vaessen, J. M. H. H. van Gorp, C. T. W. Moonen, L. W. Bartels, and M. A. A. J. van den Bosch, "First clinical experience with a dedicated mri-guided high-intensity focused ultrasound system for breast cancer ablation," *European Radiology*, vol. 26, no. 11, pp. 4037–4046, 2016.

[27] D. Harding, S. L. Giles, M. R. D. Brown, G. R. ter Haar, M. van den Bosch, L. W. Bartels, Y.-S. Kim, M. Deppe, and N. M. deSouza, "Evaluation of quality of life outcomes following palliative treatment of bone metastases with magnetic resonance-guided high intensity focused ultrasound: An international multicentre study," *Clinical Oncology*, vol. 30, no. 4, pp. 233–242, 2018.

[28] J. Chen, Y. Li, Z. Wang, P. McCulloch, L. Hu, W. Chen, G. Liu, J. Li, and J. Lang, "Evaluation of high-intensity focused ultrasound ablation for uterine fibroids: an ideal prospective exploration study," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 125, no. 3, pp. 354–364, 2017.

[29] S. Crouzet, A. Blana, F. J. Murat, G. Pasticier, S. C. W. Brown, G. N. Conti, R. Ganzer, O. Chapet, A. Gelet, C. G. Chaussy, C. N. Robertson, S. Thuroff, and J. F. Ward, "Salvage high-intensity focused ultrasound (hifu) for locally recurrent prostate cancer after failed radiation therapy: Multi-institutional analysis of 418 patients," *BJU International*, vol. 119, no. 6, pp. 896–904, 2017.

[30] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[31] American Cancer Society, "Breast cancer facts & figures 2017-2018," *Atlanta: American Cancer Society, Inc.*, 2017.

[32] C. E. DeSantis, J. Ma, A. Goding Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 6, pp. 439–448, 2017.

[33] K. Oeffinger, E. Fontham, R. Etzioni, and et al, "Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society," *JAMA*, vol. 314, no. 15, pp. 1599–1614, 2015.

[34] M. Overgaard, P. S. Hansen, J. Overgaard, C. Rose, M. Andersson, F. Bach, M. Kjaer, C. C. Gadeberg, H. T. Mouridsen, M.-B. Jensen, and K. Zedeler, "Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy," *New England Journal of Medicine*, vol. 337, no. 14, pp. 949–955, 1997.

[35] J. Ragaz, S. M. Jackson, N. Le, I. H. Plenderleith, J. J. Spinelli, V. E. Basco, K. S. Wilson, M. A. Knowling, C. M. Coppin, M. Paradis, A. J. Coldman, and I. A. Olivotto, "Adjuvant radiotherapy and chemotherapy in node-positive premenopausal women with breast cancer," *New England Journal of Medicine*, vol. 337, no. 14, pp. 956–962, 1997.

[36] W. Budach, E. Bölke, K. Kammers, P. A. Gerber, C. Nestle-Krämling, and C. Matuschek, "Adjuvant radiation therapy of regional lymph nodes in breast cancer - a meta-analysis of randomized trials- an update," *Radiation Oncology*, vol. 10, no. 1, p. 258, 2015.

[37] J. Jin, "Screening for prostate cancer," *JAMA*, vol. 319, no. 18, p. 1946, 2018.

[38] K. D. Miller, R. L. Siegel, C. C. Lin, A. B. Mariotto, J. L. Kramer, J. H. Rowland, K. D. Stein, R. Alteri, and A. Jemal, "Cancer treatment and survivorship statistics, 2016," *CA: A Cancer Journal for Clinicians*, vol. 66, no. 4, pp. 271–289, 2016.

[39] C. A. Reichard and E. A. Klein, "Current approaches to prostate cancer staging and risk stratification: In the midst of a paradigm shift," in *Prostate Cancer* (A. P. Dicker, W. K. Kelly, N. G. Zaorsky, and E. J. Trabulsi, eds.), pp. 67–76, Demos Medical Publishing, LLC., 2015.

[40] N. Veeramachaneni, "Management of lung cancer in 2017," *Freeman Professional Development*, 2017.

[41] H. Katki, S. Kovalchik, C. Berg, L. Cheung, and A. Chaturvedi, "Development and validation of risk models to select ever-smokers for ct lung cancer screening," *JAMA*, vol. 315, no. 21, pp. 2300–2311, 2016.

[42] M. VA and on behalf of the U.S. Preventive Services Task Force, "Screening for lung cancer: U.s. preventive services task force recommendation statement," *Annals of Internal Medicine*, vol. 160, no. 5, pp. 330–338, 2014.

[43] M. C. Tammemägi, T. R. Church, W. G. Hocking, G. A. Silvestri, P. A. Kvale, T. L. Riley, J. Commins, and C. D. Berg, "Evaluation of the lung cancer risks at which to screen ever- and never-smokers: Screening rules applied to the plco and nlst cohorts," *PLOS Medicine*, vol. 11, no. 12, pp. 1–13, 2014.

[44] W. Hamilton, R. Lancashire, D. Sharp, T. J. Peters, K. Cheng, and T. Marshall, "The risk of colorectal cancer with symptoms at different ages and between the sexes: a case-control study," *BMC Medicine*, vol. 7, no. 1, p. 17, 2009.

[45] S. E. Fisher and I. R. Daniels, "The clinical presentation of colorectal cancer," in *Colorectal Cancer* (G. Brown, ed.), pp. 1–14, Cambridge University Press, 2007.

[46] S. Halligan, "Screening for colorectal cancer," in *Colorectal Cancer* (G. Brown, ed.), pp. 34–58, Cambridge University Press, 2007.

[47] W. Demtröder, *Experimentalphysik 3*. Springer-Verlag Berlin Heidelberg, 3 ed., 2005.

[48] E. Schrödinger, "Quantisierung als eigenwertproblem," *Annalen der Physik*, vol. 384, no. 4, pp. 361–376, 1926.

[49] E. Albert, "über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt," *Annalen der Physik*, vol. 17, pp. 132–148, 1905.

[50] P. Max, "über irreversible strahlungsvorgänge," *Verlag der Königlichen Akademie der Wissenschaft*, pp. 440–480, 1899.

[51] M. Eichhorn, "Quantum-mechanical fundamentals of lasers," in *Laser Physics: From Principles to Practical Work in the Lab*, pp. 1–21, Cham: Springer International Publishing, 2014.

[52] L. De Broglie, *Recherches sur la théorie des quanta*. PhD thesis, Migration-université en cours d'affectation, 1924.

[53] D. Meschede, *Gerthsen Physik*. Springer, 24 ed., 2010.

[54] F. de Bergevin, "The interaction of x-rays (and neutrons) with matter," in *X-Ray and neutron reflectivity* (J. Daillant and A. Gibaud, eds.), pp. 1–59, Springer, 1999.

[55] H. Azhari, *Basics of biomedical ultrasound for engineers*. Wiley, 2010.

[56] D. F. van Wijk, A. C. Strang, R. Duivenvoorden, D.-J. F. Enklaar, A. H. Zwinderman, R. J. van der Geest, J. J. P. Kastelein, E. de Groot, E. S. G. Stroes, and A. J. Nederveen, "Increasing the spatial resolution of 3t carotid mri has no beneficial effect for plaque component measurement reproducibility," *PLOS ONE*, vol. 10, no. 7, pp. 1–15, 2015.

[57] S. G. Karshenboim, "Precision study of positronium: Testing bound state qed theory," *International Journal of Modern Physics A*, vol. 19, no. 23, pp. 3879–3896, 2004.

[58] V. Dilsizian and G. M. Pohost, *Cardiac CT, PET and MR*. John Wiley & Sons, Incorporated, 2010.

[59] H. Waters, "The first x-ray, 1895," *The Scientist*, 2011.

[60] F. M. Khan and J. P. Gibbons, *The physics of radiation therapy.* Wolters Kluwer Health, 5 ed., 2014.

[61] S. Ulzheimer and T. Flohr, "Computed tomography," in *Biomedical imaging* (R. Salzer, ed.), John Wiley & Sons, 2012.

[62] J. Jiri, *Medical Image Processing, Reconstruction and Restoration.* Taylor & Francis CRC, 2006.

[63] J. T. Bushberg, *The essential physics of medical imaging.* Wolters Kluwer Health, 2011.

[64] Z. Messali, N. Chetih, A. Serir, and A. Boudjelal, "A quantitative comparative study of back projection, filtered back projection, gradient and bayesian reconstruction algorithms in computed tomography (ct)," *International Journal of Probability and Statistics*, vol. 4, no. 1, pp. 12–31, 2015.

[65] A. Katsevich, "Analysis of an exact inversion algorithm for spiral cone-beam ct," *Physics in Medicine & Biology*, vol. 47, no. 15, p. 2583, 2002.

[66] L. Lechuga and G. A. Weidlich, "Cone beam ct vs. fan beam ct: a comparison of image quality and dose delivered between two differing ct imaging modalities," *Cureus*, vol. 8, no. 9, 2016.

[67] V. Razafindranaly, E. Truy, J.-B. Pialat, A. Martinon, M. Bourhis, N. Boublay, F. Faure, and A. Lataief-Boudrigua, "Cone beam ct versus multislice ct: Radiologic diagnostic agreement in the postoperative assessment of cochlear implantation," *Otology & Neurotology*, vol. 37, no. 9, pp. 1246–1254, 2016.

[68] R. L. Cazzato, J.-B. Battistuzzi, V. Catena, R. F. Grasso, B. B. Zobel, E. Schena, X. Buy, and J. Palussiere, "Cone-beam computed tomography (cbct) versus ct in lung ablation procedure: Which is faster?," *CardioVascular and Interventional Radiology*, vol. 38, no. 5, pp. 1231–1236, 2015.

[69] A. K. Nagarajappa, N. Dwivedi, and R. Tiwari, "Artifacts: The downturn of cbct image," *J Int Soc Prev Community Dent*, vol. 5, no. 6, pp. 440–445, 2015.

[70] S. Leng, J. Zambelli, R. Tolakanahalli, B. Nett, P. Munro, J. Star-Lack, B. Paliwal, and G.-H. Chen, "Streaking artifacts reduction in four-dimensional cone-beam computed tomography," *Medical physics*, vol. 35, no. 10, pp. 4649–4659, 2008.

[71] P. Mah, T. E. Reeves, and W. D. McDavid, "Deriving hounsfield units using grey levels in cone beam computed tomography," *Dentomaxillofacial Radiology*, vol. 39, no. 6, pp. 323–335, 2010.

[72] N. G. Burnet, S. J. Thomas, K. E. Burton, and S. J. Jefferies, "Defining the tumour and target volumes for radiotherapy," *Cancer Imaging*, vol. 4, no. 2, pp. 153–161, 2004.

[73] P. M. Harari, S. Song, and W. A. Tom, "Emphasizing conformal avoidance vs. target definition for imrt treatment planning in head and neck cancer," *Int J Radiat Oncol Biol Phys*, vol. 77, no. 3, pp. 950–958, 2010.

[74] Z. Chen, W. King, R. Pearcey, M. Kerba, and W. J. Mackillop, "The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature," *Radiotherapy and Oncology*, vol. 87, no. 1, pp. 3–16, 2008.

[75] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 490–498, 2001.

[76] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion x2014;application to cardiac and aortic segmentation in ct scans," *IEEE Transactions on Medical Imaging*, vol. 28, no. 7, pp. 1000–1010, 2009.

[77] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. A. D. Guise, "Liver segmentation on ct and mr using laplacian mesh optimization," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2110–2121, 2017.

[78] M. Levandowsky and D. Winter, "Distance between sets," *Nature*, vol. 234, pp. 34–35, 1971.

[79] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[80] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference for Learning Representations, San Diego*, 2015.

[81] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[82] H. Bauer and R. B. Burckel, *Measure and Integration Theory*. De Gruyter, Inc., 2011.

[83] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *From Natural to Artificial Neural Computation* (J. Mira and F. Sandoval, eds.), (Berlin, Heidelberg), pp. 195–201, Springer Berlin Heidelberg, 1995.

[84] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for lvcsr using rectified linear units and dropout," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8609–8613, 2013.

[85] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[86] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, 2015.

[87] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, pp. 1050–1059, 2016.

[88] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[89] G. van Dijk, *Distribution Theory : Convolution, Fourier Transform, and Laplace Transform*. De Gruyter, Inc., 2013.

[90] P. M. Mather and M. Koch, *Computer Processing of Remotely-Sensed Images : An Introduction*. John Wiley & Sons, Incorporated, 2010.

[91] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[92] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), pp. 818–833, Springer International Publishing, 2014.

[93] C.-Y. Lee, P. W. Gallagher, and Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," in *Artificial Intelligence and Statistics*, pp. 464–472, 2016.

[94] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[95] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning.," in *AAAI*, vol. 4, p. 12, 2017.

[96] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[97] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder.," in *Interspeech*, pp. 436–440, 2013.

[98] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[99] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan, *et al.*, "Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation," *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 384–395, 2018.

[100] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[101] Z. Quo, L. Zhang, L. Lu, M. Bagheri, R. M. Summers, M. Sonka, and J. Yao, "Deep logismos: Deep learning graph-based 3d segmentation of pancreatic tumors on ct scans," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1230–1233, 2018.

[102] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmadi, and B. H. Menze, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds.), (Cham), pp. 415–423, Springer International Publishing, 2016.

[103] Ö. Çiçek, A. Abdulkadir, S. 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 2016* (S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds.), (Cham), pp. 424–432, Springer International Publishing, 2016.

[104] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications* (G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and J. Cornebise, eds.), (Cham), pp. 179–187, Springer International Publishing, 2016.

[105] B. J. Erickson, D. Mutch, L. Lippmann, and R. Jarosz, "Radiology data from the cancer genome atlas uterine corpus endometrial carcinoma (tcga-ucec) collection," *The Cancer Imaging Archive*, 2016.

[106] S. Kirk, Y. Lee, C. A. Sadow, S. Levine, C. Roche, E. Bonaccio, and J. Filiippini, "Radiology data from the cancer genome atlas colon adenocarcinoma (tcga-coad) collection," *The Cancer Imaging Archive*, 2016.