COMPUTER-ASSISTED APPROACHES TO SUPPORT RADIOFREQUENCY ABLATIONS OF SPINAL METASTASES

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik

der Otto-von-Guericke-Universität Magdeburg

von M.Sc. Georg Hille

geb. am 20.03.1989 in Erfurt

Gutachterinnen/Gutachter

Prof. Dr. Klaus-Dietz Tönnies

Prof. Dr. Alexander Schlaefer

Prof. Dr. Dorit Merhof

Magdeburg, den 14.12.2020

ABSTRACT

The spine is the most common body part to develop bone metastases from various primary tumours with increasing case numbers over the last decades. The drastic effects on the quality of life evoked by spinal metastases, such as severe pain symptoms or neurological deficits due to nerve root and spinal cord compressions, demand a fast-acting, yet gentle therapeutic solution, as enabled by minimally invasive interventions like radiofrequency ablations. For this purpose, one or multiple applicators with electrode needle tips are placed within the tumour volume and necrotise the cancerous tissue by high frequency-induced tissue heating. The entire clinical workflow of such a minimally invasive intervention is based on medical imaging, starting from the initial diagnosis over image-guidance during the intervention to therapy control via follow-up scans. Computer-assisted strategies can support the radiologists to obtain more relevant information from the acquired images and to transfer these to subsequent processing steps. This enables a more sophisticated workflow, while reducing the required time and workload of the radiologists. However, the specific image processing aspects to achieve this goal are challenging with regard to the required expert-like performance, high level of automatisation, and short computational times.

This thesis focuses on various aspects throughout the clinical workflow of radiofrequency ablations of spinal metastases. For this purpose, the thesis is structured following the chronological sequence of the clinical process and contains approaches to support the radiologist during the pre-, intra- and post-interventional phase. Limitations and gaps in the existing state of the art of each aspect led to the development and implementation of novel strategies to provide suitable and applicable solutions. In detail, segmentation approaches of involved anatomical structures like vertebral bodies, metastases and resulting necrosis zones have been developed - the latter two being, to the best of the author's knowledge, the first of their kind. Furthermore, an image registration method is presented, which is able to cope with the poor image quality of interventional imaging and the specific issue of spinal deformations due to different patient positioning. Finally, a novel framework is proposed to automatically visualise and quantify the treatment outcome of spinal metastasis interventions. Each of the above-mentioned methods has been evaluated on a wide range of patient data in order to demonstrate robustness, reliability, accuracy, and speed to meet the clinical objectives.

ZUSAMMENFASSUNG

Die Wirbelsäule ist die muskuloskelettale Struktur, in der sich am häufigsten Knochenmetastasen verschiedenster Primärtumore entwickeln und dies mit stetig steigenden Fallzahlen in den letzten Jahrzehnten. Wirbelsäulenmetastasen verursachen eine drastische Beeinträchtigung der Lebensqualität der meisten Patienten, bedingt durch eine ausgeprägte Schmerzsymptomatik sowie teilweise durch neurologische Ausfallerscheinungen aufgrund von Nervenwurzel- und Rückenmarkkompressionen. Dies wiederum erfordert eine unverzögert wirksame, aber im Hinblick auf das fortgeschrittene Alter der meisten Patienten trotz allem schonende therapeutische Lösung, wie sie insbesondere minimal-invasive Eingriffe, beispielsweise die Radiofrequenzablation, versprechen. Hierbei werden ein oder mehrere Applikatoren mit Elektroden an deren Spitze in dem Tumorvolumen platziert, welche anschließend das metastasierte Gewebe mittels hochfrequenzinduzierter Gewebeerhitzung nekrotisieren. Die medizinische Bildgebung spielt während des gesamten klinisch-therapeutischen Prozesses einer minimal-invasiven Intervention eine entscheidende Rolle; angefangen bei der initialen Diagnose, über die bildgestützte Durchführung des Eingriffs, bis hin zur abschließenden Therapiekontrolle mittels Bildgebung. Computergestützte Strategien können Radiologen gezielt dabei helfen, aus den generierten Bildern zusätzlich relevante Informationen zu gewinnen und diese auch auf nachfolgende Prozessschritte zu übertragen. Dies eröffnet die Perspektive auf einen insgesamt fortschrittlicheren klinischen Arbeitsablauf und reduziert darüber hinaus die zeitliche und mentale Arbeitsbelastung der beteiligten Radiologen. Die spezifischen Bildverarbeitungsaspekte zur Erreichung dieses Ziels sind jedoch im Hinblick auf ihre Komplexität sowie die erforderliche methodische Genauigkeit und die kurzen Berechnungszeiten, eine Herausforderung.

Diese Dissertation thematisiert verschiedene Aspekte des klinischen Arbeitsablaufs bei der Radiofrequenzablation von Wirbelsäulenmetastasen. Hierzu folgt die Arbeit dem chronologischen Ablauf des klinischen Prozesses und beinhaltet Strategien für die zielgerichtete Unterstützung der Radiologen während der prä-, intra- und post-interventionellen Phase. Bestehende Limitationen oder Lücken im derzeitigen Stand der Technik jedes einzelnen Aspektes erforderten die Entwicklung und Umsetzung neuer Lösungsstrategien, welche auf geeignete Art und Weise die klinischen und technischen Anforderungen erfüllen. Im Detail wurden Verfahren zur Segmentierung von relevanten anatomischen Strukturen, wie bspw. der Wirbelkörper, der Metastasen und der resultierenden Nekrosezonen entwickelt, sowie ein Bildregistrierungsverfahren, das der schlechteren Bildqualität interventioneller Bildgebung und dem spezifischen Problem der Wirbelsäulendeformationen aufgrund unterschiedlicher Patientenpositionierung gerecht wird. Abschließend wird ein Framework

vorgestellt, welches eine automatische Quantifizierung und Visualisierung des Behandlungserfolges nach tumorbedingten Wirbelsäuleninterventionen ermöglicht. Jede der oben genannten Methoden bzw. Lösungsstrategien wurde mithilfe einer Vielzahl von klinischen Patientendaten evaluiert, um die benötigte Robustheit, Genauigkeit und Geschwindigkeit der Verfahren zu demonstrieren.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

RF	radiofrequency
RFA	radiofrequency ablation
MR	magnetic resonance
MRI	magnetic resonance imaging
CT	computed tomography
FP-CT	flat-panel detector computed tomography
PET	positron emission tomography
SPECT	single photon emission computed tomography
US	ultrasound
HU	Houndsfield unit
STIR	short tau inversion recovery
SPIR	spectral presaturation with inversion recovery
PVE	partial volume effects
SNR	signal-to-noise ratio
IRV	inter-reader variability
ROI	region of interest
GTV	gross tumour volume
FOV	field of view
DSC	Dice similarity coefficient
FRE	fiducial registration error
FRE_{MS}	fiducial registration error of the multisegmental approach
FRE_{GR}	fiducial registration error of the globally rigid approach
mFRE	minimal fiducial registration error
FLE	fiducial localisation error
ASD	average surface distance
HD	Hausdorff distance
HD95	95th percentile Hausdorff distance

ACM	active contour models
ASM	active shape models
BC-HLS	bias field-corrected hybrid level-sets
CNN	convolutional neural network
SdNN	Siamese deep neural network
ReLU	rectified linear unit
TL	Tversky loss
SVM	support vector machines
FLD	Fisher's linear discriminant
CRF	conditional random field
MI	mutual information
NMI	normalised mutual information
NGF	normalised gradient fields
MS-VB	multisegmental voxel-based
TPR	true positive rate
TNR	true negative rate
GPU	graphics processing unit
GUI	graphical user interface
SHIP	Study of Health in Pomerania
BRATS	brain tumour segmentation challenge
LITS	liver tumour segmentation challenge

1 INTRODUCTION

1.1 MOTIVATION

Due to the enhancement of medical treatment and diagnostic procedures, life expectancy has increased steadily over the last decades. However, this lifetime gain promotes also age-related diseases like cardiovascular diseases, as well as cancer and cancer-induced malicious metastases. Both of them are the most common causes of death nowadays. The survival time of most malicious carcinomas has increased with improved diagnosis and treatment, though, this also promotes spreading of metastases. Besides liver and lung, bone metastases are the third likely, and thereof between two thirds (Harrington, 1986; Wong et al., 1990) and over 90 % (Frangou and Fourney, 2009) are located in the spine, varying according to the reference. The most common primary malignancies that lead to bone metastases are listed in Table 1.1.

Table 1.1: The incidence of skeletal metastases, based on autopsy studies (Maccauro et al., 2011).

Tumour	Incidence
Mammary carcinoma	73 %
Prostatic carcinoma	68 %
Thyroid carcinoma	42 %
Bronchial carcinoma	36 %
Renal carcinoma	35 %
Rectal carcinoma	11 %
Esophageal carcinoma	6 %
Gastrointestinal carcinoma	5 %

Studies showed that most metastases occur in the thoracic spine, followed by the lumbar segment, where as the cervical region is the least involved (Klimo and Schmidt, 2004). More than half of the patients with spinal metastases have lesions in multiple, partly non-contiguous spine segments (Togawa and Lewandrowski, 2006). Following the mechanistic theory, tumour cells will metastasise anatomical regions near to their primary site, e.g. mammary carcinomas preferably infiltrate vertebrae of the thoracic region and prostate cancer usually metastasise the lumbarsacral spine (Algra et al., 1992; Gilbert et al., 1978). The initial anatomical location of spinal metastases is generally the posterior portion of the vertebral body, gradually followed by the anterior body, lamina and pedi-

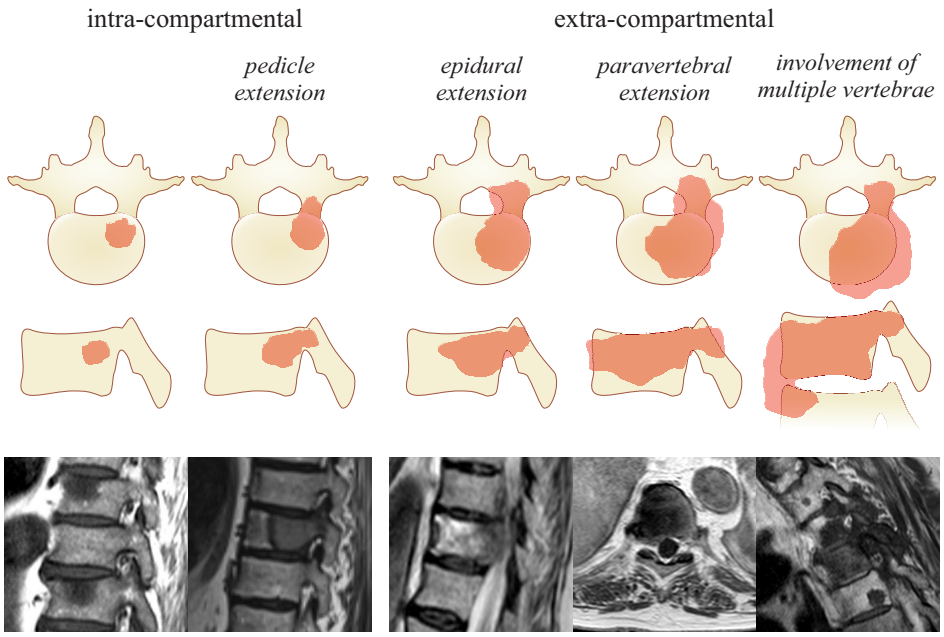

Figure 1.1: Illustrations of different degrees of vertebral involvement in spinal metastases with corresponding exemplary patient cases.

cles with increasing tumour stages (Eleraky et al., 2010; Georgy, 2008). The vertebral involvement in spinal lesion diseases largely influences the treatment strategy and therapeutic decision-making (see Figure 1.1). Osseous metastases can be typically classified as osteolytic, i.e. with increased osteoclastic activity and therefore bone resorption, or osteoblastic (sclerotic), i.e. enhanced osseous tissue production, as well as a mixed combination of the two types (Eleraky et al., 2010). The infiltration of tumour cells causes an imbalance of osteoclastic and osteoblastic activity leading to a release of growth factors, which stimulate bone remodelling and further growth factor production. This results in a dire cycle of bone destruction and local tumour growth (Lipton, 2004; Yin et al., 2005). In some cases, certain types of metastases can be assigned to different primary tumours, e.g. prostate or thyroid tumours predominantly develop osteoblastic metastases, bronchial and renal carcinomas often result in osteolytic types, and mammary carcinomas could lead to mixed sclerotic and lytic lesions (Yin et al., 2005).

Regarding the biomechanics of metastatic vertebrae, destabilisation due to fractures both, under traumatic or normal physiological stress is one of the most common consequences (Georgy, 2008; Whyne et al., 2003), especially if osteolytic metastases weaken the internal bony matrix structures. Spinal stability in a clinical sense includes mechanical stability, as well as the absence of pain, deformity and any neurological signs (Panjabi, 2003). However, spinal metastases could tremendously affect the quality of life and the primarily therapeutic indications are vigorous pain by fractures, bruises, spinal cord and nerve root compressions and therefore, neurological deficits (Klimo and Schmidt, 2004). The latter often occure in advanced stages due to growing extravertebral masses (Guillevin et al., 2007). Once tumours spread and develop metastases, full

recovery is rarely possible and the therapeutic goal is often to stop, delay or shrink its growing masses. Although, a complete cure is often no longer possible, palliative treatment of spinal metastases is indicated with regard to pain palliation and the release of stenosis-related neurological deficits and an overall improved life quality. However, the relative survival rate at five years of metastatic tumours drastically decreases compared to non-metastatic tumours, e.g. for prostate cancer about 70 %, colon cancer about 25 % or renal tumours about 18 % (American Cancer Society, 2019).

To detect spinal metastases, a wide variety of imaging methods can be applied, such as scintigraphs, X-ray radiography, computed tomography (CT), magnetic resonance imaging (MRI) as well as some functional imaging methods like positron emission tomography (PET) and single photon emission computed tomography (SPECT). Conventional radiography may be useful to first detect abnormalities due to lytic or sclerotic bone remodelling, but lacks more detailed information about shape and bone marrow integrity. Furthermore, it commonly detects metastases in later stages with advanced osseous structural loss (Shah and Salzman, 2011). CT imaging is superior in terms of a detailed morphology, its high spatial resolution and an increased soft-tissue contrast. However, similar to X-ray radiography the appearance of metastases depends on their mineralisation and therefore it requires noticeable bone remodelling to be recognised. Thus, early detection of infiltrated bone marrow is hampered and CT scans as a diagnostic imaging technique may be unsuitable. In CT imaging, lytic lesions often appear as soft tissue regions with irregular margins and hypointense image signals compared to osseous structures. In contrast, sclerotic metastases predominantly show hyperdense bone matrix structures with bright image signals, comparable to cortical bone (see Figure 1.2). In comparison to the aforementioned imaging techniques, MRI overcomes restrictions of radiation exposure and combines detailed morphology and high soft tissue contrast as well as enhanced bone marrow visualisation, which makes it an adequate and useful technique for early stage metastasis detection. In addition, MRI is suitable for the assessment of spinal cord compressions and thus for the clarification of symptomatic neurological deficits (Shah and Salzman, 2011). The displayed signal intensities vary with regard to the acquisition sequence and metastatic type due to their degree of mineralisation (see Figure 1.2). Sclerotic metastases predominantly appear hypointense in both T_1- and T_2-weighted sequences compared to healthy bone marrow image signals. Lytic lesions show mainly intermediate to hypointense signals in T_1- and hyper- or isointense signals in T_2-weighted sequences. Furthermore, a contrast agent-induced signal enhancement is usually present with metastases of the lytic type (Shah and Salzman, 2011). The drawback of MRI is the more challenging distinction between remaining active tumour tissue and scarred, necrotised or fractured bone tissue in post-therapeutic scans (Even-Sapir, 2005; Wong et al., 1990). These limitations can be partially compensated by contrast-enhanced sequences, e.g. with gadolinium-based contrast-enhancing agents, since it adds a dynamic component to the imaging process (Kim et al., 1982; Kim et al.,

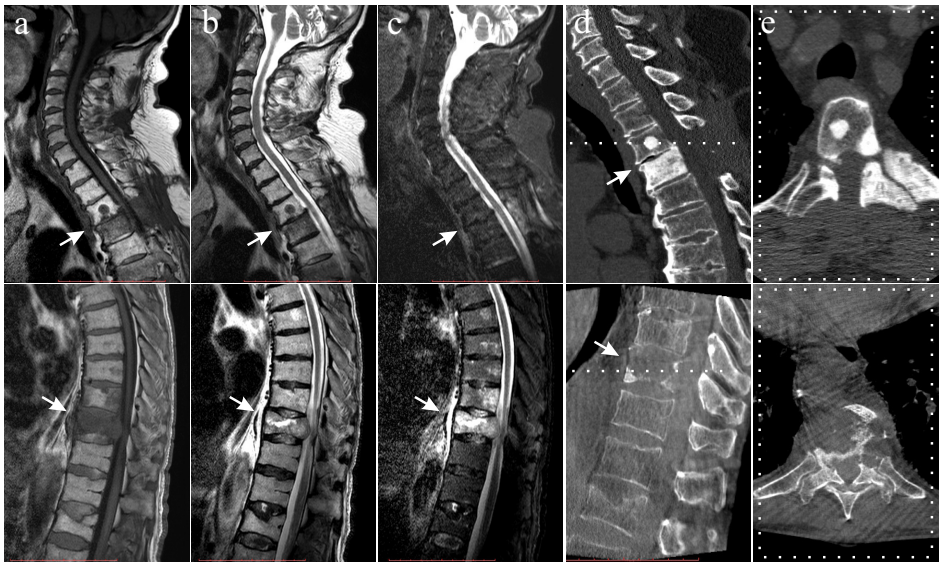

Figure 1.2: Appearance differences of sclerotic (upper row) and lytic (bottom row) spinal metastases (arrows) in MRI and interventional CT imaging. Displayed are two exemplary patient cases with sagittal T_1-weighted (a), T_2-weighted (b), and STIR (c) MRI sequences as well as their corresponding interventional FP-CT scans in a mid-sagittal (d) and an axial cross-section (e). Especially in X-ray-based imaging, the different physiological processes regarding the bone alteration of both lesion types can be vividly visualised.

2003). The major advantage of SPECT or PET imaging as functional and nuclear imaging techniques lies in their potential to detect pathological biochemical and physiological abnormalities due to carcinomas with high specificity (Van Dort et al., 2008). PET imaging with specific radioligands like [^{18}F] flouride could furthermore support the distinction between lytic and sclerotic metastasis types (Barzilai et al., 2018; Even-Sapir, 2005). The decisive disadvantage of these imaging procedures for diagnostic or interventional purposes is their comparatively limited spatial resolution, which could be partly overcome in hybrid imaging techniques like PET/CT, SPECT/CT or even PET/MRI, whose limited availability in clinical routine practice is slowly growing (Shah and Salzman, 2011).

The advances of treatment strategies for metastatic spinal lesions led to the development of the "NOMS" framework, comprising **N**eurological, **O**ncological, **M**echanical, and **S**ystemic assessments to support complex decision-making for therapy techniques across disciplines (Zuckerman et al., 2018). With the aid of NOMS, various key aspects like presence of epidural spine compressions, expected tumour control, vertebral stability, and the risk-benefit ratio of different treatment strategies are assembled to optimise patient care and overall survival (Barzilai et al., 2018).

Historically, the method of choice to treat osseous metastases was conventional external beam radiation, i.e. the target area was radiated by one or two beams. The major drawback of this treatment strategy regarding spinal metastases is the relatively widespread target area including risk structures like the spinal cord, which limits the applied radiation doses

and lead to the necessity of increased fractioning (Barzilai et al., 2018). Hence, the treatment response is commonly delayed and transient. In contrast, spine stereotactic radiosurgery can deliver high-dose ablative radiation in typically one to five fractions to the target (Huo et al., 2017). It utilises image-guided intensity-modulated radiation delivery and steep dose gradients due to highly focused beams, which results in effective doses within the target structures, while protecting adjacent organs at risk (Huo et al., 2017). Therefore, spine stereotactic radiosurgery commonly ensures a fast and durable symptomatic response, i.e. high local tumour control and pain relief. However, radiation therapies are constrained in terms of radioresistant tumours and do not address spinal instabilities, requiring adjuvant procedures like vertebroplasty or balloon kyphoplasty (Barzilai et al., 2018; Georgy, 2008). Chemotherapy or hormonal therapy often tend to fail the desired relief of symptoms of osseous metastases and take time to be effective (Rosenthal and Callstrom, 2012). Owing to the short life span and morbidity of most patients, surgical interventions may often be beyond dispute. These range from the resection of individual vertebrae, discs and surrounding ligaments to ensure en-bloc excisions to intra-tumoural surgeries in order to reduce compressive stress on neural structures. Since it is highly challenging to achieve satisfying resection margins, the risks associated with such procedures often contradict a surgical intervention (Barzilai et al., 2018).

Besides, percutaneous minimally invasive therapies gain reception as promising methods to treat spinal metastases or tumours. These include, among others, various thermal therapies, like microwave ablation, laser ablation, cryoablation and radiofrequency ablation (RFA), which cause necrotisation due to target tissue heating or freezing. Microwave ablation utilises electromagnetic waves with frequencies between 900 - 2450 MHz to heat up the target tissue via molecule agitation, which leads to coagulative necrosis. The main drawback is the low availability of commercial systems, which is probably also the reason why the related literature is rather scarce, especially for spinal metastases (Zhang, 2016). Laser ablation is based on heating up the tumour tissue by infrared light energy through optical fibers and, therefore, induces tissue necrotization of smaller regions. However, there are only very few studies considering laser ablation of bone or spinal metastases (Evans et al., 2020; Rothrock et al., 2020). In contrast to the above, cryoablation rapidly cools the tissue to temperatures of -100°C through partially insulated probes, which insert gaseous or liquid nitrogene or more recently argon into the target tissue (Chu and Dupuy, 2014; Skanes et al., 2004). While cryoablation has the advantage of good visibility in MRI or ultrasound (US) imaging with an easily recognisable ablation zone due to the ice ball formed at the needle tip, its applicability is limited to rather small lesions with a radius of roughly 2 - 2.5 mm, since the cooling effect rapidly declines with increasing distance from the cryogenic center (Khairy et al., 2003; Skanes et al., 2004).

Radiofrequency ablation represents an effective method to treat relatively small tumours and metastases, if surgical resection is inapplicable

and pain reduction is attempted within several hours or few days following the intervention. It has been used to reduce lower back pain caused by facet osteoarithritis (Cho et al., 1997) or osteoid osteoma (Rosenthal et al., 1998) and was introduced approximatley a decade ago to treat osseous spinal metastases (Dupuy et al., 2010). Overall, the number of RFA sessions per year has steadily increased over the last decade (Starr et al., 2019) and is expected to continue to grow immensely (Transparency Market Research, 2019). However, RFA treatment of metastases is - with a few exceptions - not curative, but suitable for pain palliation, to regain lost neurological function, and to contain local tumour growth (Posteraro et al., 2004). RFA utilises frequencies within the range of 300 - 3000 kHz, with no stimulation or interference with neuromuscular or electrolytic processes (Ni et al., 2005). The basic setup consits of a radiofrequency (RF) generator as the source of the RF voltage, a needle-like applicator with electrodes at its tip and a grounding which closes the current circuit with the patient's body as an active element. When the generator is switched on, a high induction flux forms around the active electrodes at the needle tip due to the impedance characteristic of the target volume (Hong and Georgiades, 2010; Kline, 2000).

The physics underlying the RFA is based on the reaction of the target tissue's dipole molecules, i.e. primarily water molecules, which align in the direction of the current and begin to oscillate at the same frequency. This molecular oscillation leads to a friction-induced rise in temperature, which in the end results in coagulation and therefore, target tissue necrotisation (see Figure 1.3). It is worth mentioning, that the electrode at the needle tip itself is not hot or the thermal source of heating, but triggers ionic movement within the adjacent tissue that causes the heating (Hong and Georgiades, 2010). However, this means that the electrical and thermal conductivity of the tissue is of critical importance for successful ablation procedures. In addition, it is essential that the tissue temperature is rising not too fast and not beyond 105°C, since consequent carbonisation or vaporisation could restrict any further energy transmission and lead to incomplete ablation zones (Hong and Georgiades, 2010). At temperatures above 60°C cells start to necrotize (Carrafiello et al., 2007; Frangou and Fourney, 2009) due to irreversible protein denaturation and the destruction of key enzymes (Goldberg and Gazelle, 2001). Furthermore, studies showed that temperatures around 45°C could result in apoptosis, which is of crucial importance regarding the protection of surrounding risk structures like spinal cord or peripheral nerves (Vujaskovic et al., 1994; Yamane et al., 1992). Thus, a typical ablation zone consists of a core region of full necrosis around the needle tip electrodes with an adjacent area of moderate necrotisation and a further zone characterised by apoptosis. These biochemical effects of RFAs occur within seconds to a few minutes, leaving both micro- and macroscopically visible effects, in particular a microcavitation caused by the applicator access path, the ablated tumoural and peri-tumoural tissue, a necrosis-related dark rim and an outer area of inflammation or oedema (Ni et al., 2005). Regarding the RFA devices, it can be noted that almost all medical generators use frequencies between

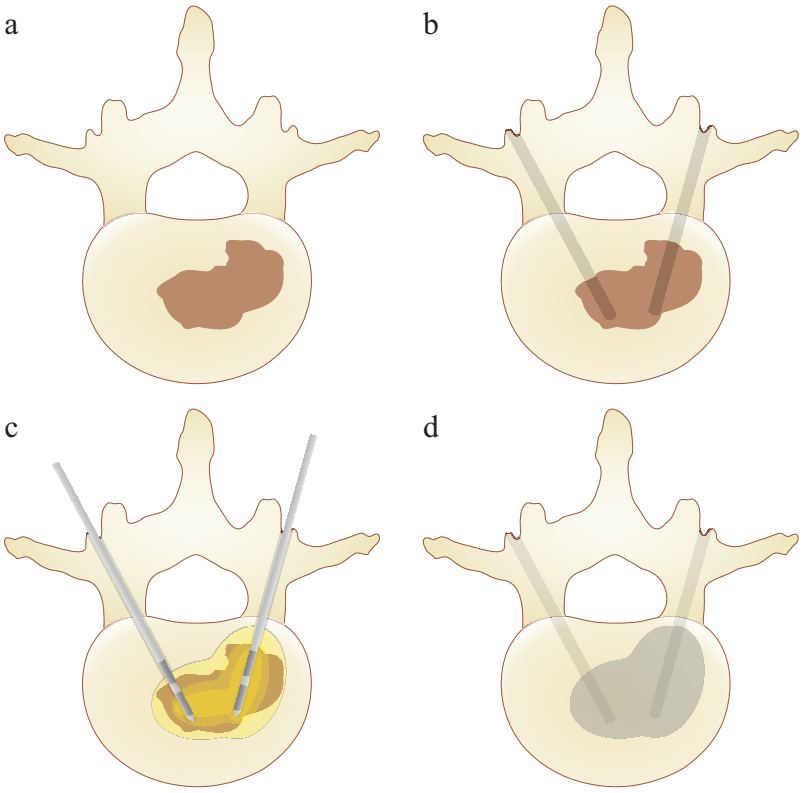

Figure 1.3: Illustration of the procedure of RFA of spinal metastases. Subsequent to the calibration of the image-guided navigation system at the beginning of the intervention (a) the radiologist creates access pathways for the RFA applicators using trocars (b). After the insertion of either single or multiple applicators, an expanding ablation zone develops w.r.t. the elapsed time and the induced energy (c). The resulting shape of the necrosis zone can be influenced by timed switching between various configurations of active pairs of electrodes (d).

450 - 600 KHz, while probe types differ more fundamentally in design, circuitry and feedback mechanism, e.g. monopolar, bipolar, internally cooled, single- or multi-array needle tips (Zhang, 2016).

The following sections deal with the current clinical workflow in detail and show existing limitations and constraints, as well as the potential for improvement through computer-assisted methods.

1.2 CURRENT WORKFLOW

The following description of the workflow refers primarily to clinical procedures in the Department of Neuroradiology at the University Hospital of Magdeburg, but could partly be transferred to related facilities, which treat vertebral metastases by RFA. The whole clinical workflow of RFAs of spinal metastases could be subdivided into three major phases: the pre-interventional therapy planning phase, the image-guided interventional phase and the post-interventional therapy control phase (see Figure 1.4).

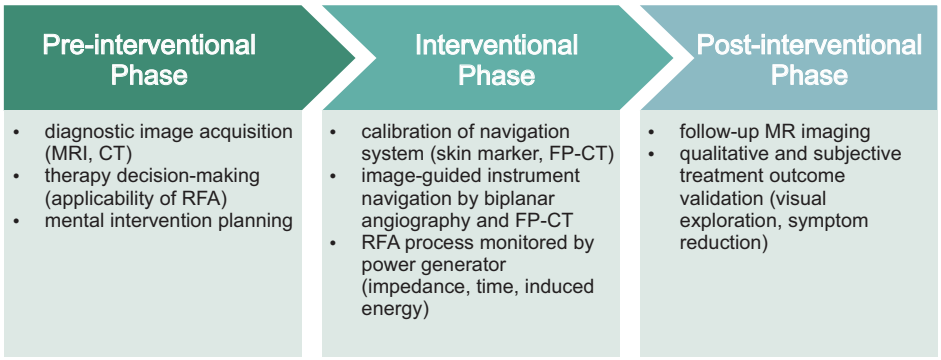

Figure 1.4: Current workflow of spinal metastasis treatment using RFA in the Department of Neuroradiology of the University Hospital of Magdeburg.

PRE-INTERVENTIONAL PHASE Patients with unresolved backpain or with suspected metastases are examined using various imaging protocols. Starting with spinal MRI, containing sagittal and axial native T_1- and T_2-weighted sequences as well as sagittal short tau inversion recovery (STIR)/spectral presaturation with inversion recovery (SPIR) sequences to enhance oedemata typical due to cancerous and metastatic processes. Native T_1- and contrast-enhanced T_1-weighted sequences are commonly the most useful in terms of spinal lesion diagnosis, since intra-vertebral image signals hypointense to surrounding muscles, discs, and normal bone marrow strongly indicate abnormality and marrow replacement (Shah and Salzman, 2011). The different MRI protocols mainly serve the purpose of providing intra-sequence image contrasts, which are highly tissue-specific due to the biochemical composition and the characteristic differences of various MRI sequences in displaying fat and water. Adjuvant CT imaging is used to highlight fracture patterns and bone density alterations, i.e. osteolytic or sclerotic processes due to metastases (Halvorson et al., 2006; Shah and Salzman, 2011). Especially, the integrity of the ventral vertebral rim is assessed in CT scans, since it affects therapeutic decision-making w.r.t. post-interventional stabilisation of the ablated vertebrae. Accordingly, only if the ventral vertebral rim is unimpaired, kyphoplasty comes into question. The circumstance that diagnostic imaging is most commonly performed in the supine position will be of great relevance for the following interventional image-guidance. In general, the current clinical diagnostic and therapeutic decision-making does not include advanced image processing or interactive procedures and is mainly based on image exploration and the experience of the radiologists involved. This applies to the decision whether RFA is feasible and, if so, to the mentally planning of interventional access routes and RFA needle positioning.

INTERVENTIONAL PHASE The first step during the intervention is to place skin markers on the patient's back, calibrate and initialise the CAScination navigation system with a flat-panel detector computed to-

Figure 1.5: Current setting in the operating room in the Department of Neuroradiology of the University Hospital of Magdeburg. On the right hand side, the CAScination navigation system is located (with a 3D model of the patient's thorax on screen; built from the FP-CT volume). The left monitor displays angiographic images, taken regularly to track the RFA needle and guide the metastasis puncture. The power generator for the radiofrequency current is covered by the left monitor.

mography (FP-CT) scan using a rotational C-arm (Siemens Dyna-CT, Doerfler et al., 2015). Subsequently, the CAScination system creates a volume from the Dyna-CT scan for optical navigation and tracking (see Figure 1.5). For the purpose of RFA needle placement, the radiologist hammers and/or drills a trocar into the vertebral bone structure, most commonly through the vertebra's pedicle, i.e. transpedicular, as this is most likely to preserve the structural integrity of the vertebrae (Chen et al., 2016). This is typically feasible only for lumbar and lower thoracic vertebrae, since their pedicles ensure enough space for a stability-preserving insertion through the narrow corridor of cancellous bone tissue. Dependent on factors like age, sex, and height, the pedicle diameter usually varies between 3 to >10 mm from cervical to lumbar vertebrae (Charles et al., 2015; Christodoulou et al., 2005; Liu et al., 2010; Scoles et al., 1988). If spatial restrictions contradict a transpedicular pathway creation, typically for upper thoracic and cervical vertebrae, parapedicular pathways, i.e. along the outside of the pedicles, are used to access the target region (Kothe et al., 2001). For both approaches, biplanar angiographic shots and FP-CT scans are performed to track the current position. However, both image modalities do not directly display the metastases, but rather indicate their location and extent due to advanced alterations of the bone matrix structure. Moreover, metal artifacts caused by inserted instruments further aggravate an accurate tumour localisation (see Figure 1.6). Following the resulting pathways, RFA needles are inserted. Electrodes on the needle tip trigger molecular friction due to 300 - 500 kHz current phase changes and therefore, rising tissue temperature (see Figure 1.7). The metastatic tissue is treated with temperatures above 70°C leading to coagulation necrosis and cell death (Palussiere et al., 2012).

Owing to the poor visibility of metastatic tissue during the intervention, the radiologists have to largely infer the location of the metastases from

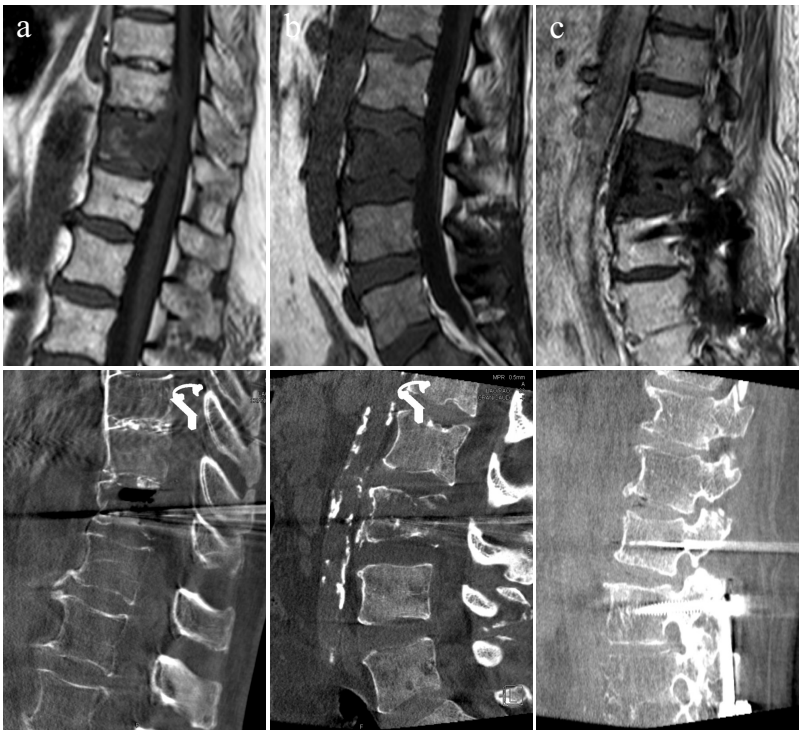

Figure 1.6: Pre-interventional MRI (upper row) and interventional FP-CT scans (lower row) of three exemplary patient cases (a-c) are shown. The overall image quality and the soft tissue contrast of the interventional scans suffer noticeably due to low-dose protocols, as well as beam hardening, streak, and dark band artifacts from the inserted applicators.

the pre-interventionally acquired MRI data and mentally match those images with the intra-interventional imaging. Thus, an exact localisation is only possible to a limited extend and this may result in suboptimal RFA applicator positions and time-consuming corrections.

In contrast to diagnostic imaging, each image during the intervention is moreover acquired in prone patient position, causing intervertebral joint movements and an altered spine flexion compared to the diagnostic images, particularly in thoracic and cervical spine segments. This aspect further increases the cognitive demands of the radiologists for a precise metastasis puncture. In addition, the ablation progress itself cannot be visualised on the interventional images and must be inferred from RF power generator parameters. With progressing ablation of the metastases the impedance will increase due to the absence of conducting tissue, and along with the overall ablation time and energy the radiologist could estimate the coagulation and ablation progress, respectively. Currently, there is no more accurate approach, such as a suitable MRI thermometry or transferred patient-specific necrosis zone predictions as a result of pre-interventional simulations. Subsequent to the ablation process, stabilisation methods are applied if necessary, e.g. kyphoplasty or vertebroplasty (Posteraro et al., 2004).

Figure 1.7: Patient's back with both, skin markers for the navigation and the inserted RFA applicators. In the background stands the power generator for the radiofrequency current displaying the induced energy, the ablation time and the current tissue impedance, which are the parameters to estimate the ablation progress.

POST-INTERVENTIONAL PHASE A few days after the intervention, post-operative MRI scans are acquired to evaluate the treatment outcome, which is currently done by separately exploring and mentally matching metastasis and necrosis zone from pre- and post-interventional images (see Figure 1.8). This, however, is challenging due to the difficulties of correlating spatial positions in both image volumes and estimating the correct spatial extension in three-dimensional space. However, various studies have shown that a reliable assessment of the ablation zone is generally possible by means of follow-up imaging, since they found a strong correspondence between macroscopically and MRI-based measured ablation zones (Palussiere et al., 2012). Therefore, an image-based and quantitative assessment of the necrosis area would be convincing regarding the treatment outcome. Despite this, there are currently no image processing methods like registration or segmentation approaches involved in the post-interventional workflow, leading to time-consuming, inconvenient and rather subjective assessments without any quantitative conclusions of the treatment outcome.

1.3 INTENDED WORKFLOW

The current clinical workflow offers various possibilities to enhance or speed up processing steps by means of computer-assistance, regardless whether pre-, intra- or post-interventional (see Figure 1.9). Additionally, data generation associated with computer-aided methods promotes comprehensibility, reproducibility and clinical documentation.

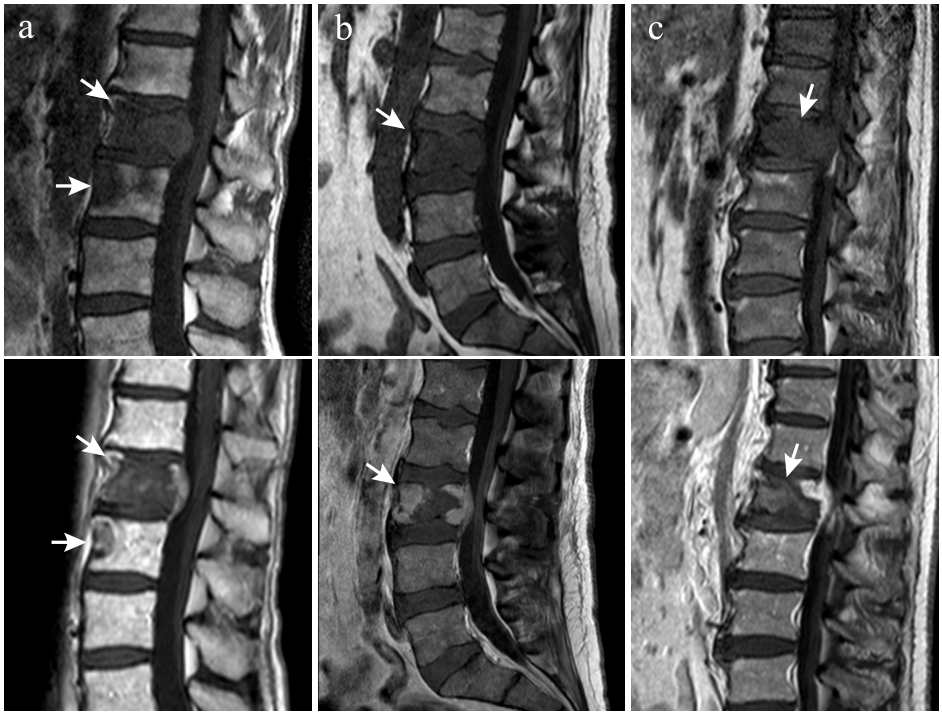

Figure 1.8: Pre- (upper row) and post-treatment (lower row) MRI scans of three exemplary patient cases (a-c) with corresponding metastases and necrosis zones (arrows). Shown are T_1-weighted (pre-RFA) and contrast-enhanced T_1-weighted (follow-up) scans.

PRE-INTERVENTIONAL PHASE Following the image acquisition, diagnosis and decision for RFA treatment, a patient-specific simulation of the RFA can predict the coagulation area considering state-dependent tissue parameters, such as electric and thermoconductive properties (Kröger et al., 2006; Weihusen et al., 2010). In order to assign those tissue parameters to patient-specific anatomical structures, a preceding detection and segmentation step, either manual, semi-automatic or fully automatic needs to be implemented. This includes first and foremost the metastases itself, as well as the surrounding tissues like vertebral bodies, intervertebral discs and organs at risk like the spinal cord. Since manually performed segmentations tend to be highly time-consuming and tedious considering the amount of tomographic image data, automatised approaches are to be preferred. The overall goal of computer-assistance during the therapy planning phase is to define a patient-specific optimal intervention strategy using a numerical simulation which predicts the ablation zone, considering optimised RFA applicator positioning, induced energy per time, and state-dependent tissue parameters of the involved structures (Matschek et al., 2017; Weihusen et al., 2010).

INTERVENTIONAL PHASE Immediately before intervention begin, the previously created and patient-individually optimised treatment plan could be digitally accessed and thus enables a rapid mental preparation to the upcoming intervention. Additionally, pre-interventionally

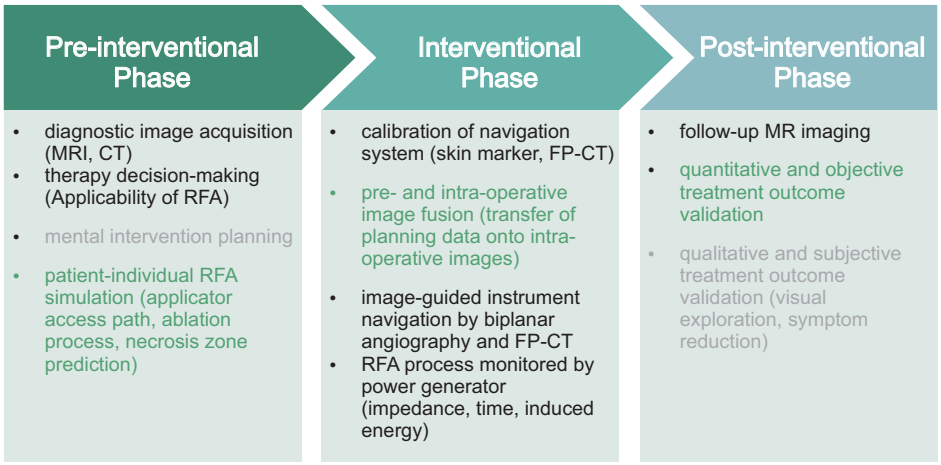

Figure 1.9: The current clinical procedure can be improved by integrating computer-assisted methods (in green) in each of the main workflow phases. These approaches replace in particular time-consuming tasks based on mentally demanding work (in grey). In contrast to the support during the intervention and the subsequent treatment outcome validation, this thesis will not cover the actual RFA simulation itself, but approaches to provide necessary prerequisites for its feasibility.

acquired information could be transferred onto intra-interventional image data via manual or automatised image fusion. Thus, overlays of previously segmented metastases, risk structures, predicted ablation zones and preferable applicator pathways could be projected onto the intra-operative images, which would enhance metastasis puncture precision and speed and thus, positively affect the treatment outcome. Supporting visual context information could reduce the radiologists' cognitive effort arising from the mentally mapping of diagnostic image and planning information onto intra-operative images. However, largely automatic registration methods are to be preferred, since manual fusion would considerably delay the start of the intervention. Considering navigation support by optimised applicator pathways, a real-time tracking of the needle tips could further reduce the required cognitive effort of the radiologists. There are various existing approaches in literature and also commercial solutions regarding this tracking task, which could also enable live updates of the intervention plan (Hirooka et al., 2016; Tomonari et al., 2013). Subsequent to the metastasis puncture, the ablation process is performed under consideration of the parameter settings, i.e. induced energy and ablation time, defined during the preceding simulation.

POST-INTERVENTIONAL PHASE After the RFA, follow-up MRI scans can be used as an input for a computer-supported assessment of the treatment outcome. With the aid of a framework that covers every step of the post-treatment process, i.e. target structure segmentation, image fusion of pre- and post-RFA images, and the computation of quantitative validation measures, the intervention outcome can be reliably and objectively

evaluated. Information regarding tumour coverage and safety margins towards organs at risk, for instance, can enhance prediction performance of tumour recurrence and survival time. Furthermore, by comparison with the pre-interventional RFA simulations, the results of the treatment validation can contribute to the optimisation of prospective ablation zone predictions, in the sense of a feedback loop. Considering such an advanced computer-supported post-treatment workflow, the framework should consist of widely automatised, precise, and fast image processing methods.

1.4 THESIS OBJECTIVES

The previous section has pointed out, that there are various indications to establish a more sophisticated and computer-assisted workflow of RFA interventions of spinal tumours and metastases, as it is done at present. The main purpose of this thesis is to provide solutions in terms of widely automatised image processing approaches to reduce the radiologists' workload and time needed for recurring and time-consuming tasks, to enhance precision and speed of interventional procedures, and to support clinical decision-making. It is of crucial importance that such methods are suitable to be integrated into clinical routine by meeting the requirements defined by radiologists.

Therefore, various aspects throughout the clinical workflow with promising potential of improvement were identified, starting from approaches to support therapy and intervention planning, to approaches that provide relevant assisting information during the intervention, as well as methods to enable a quantitative and reliable treatment outcome validation afterwards. For this purpose, multiple objectives regarding each task were defined in cooperation with the clinical partner in order to develop suitable and adequate solutions. In addition, various aspects of this thesis can also be transferred to analogous clinical issues or interventions, which makes the findings of this work relevant beyond the particular underlying subject matter.

1.5 THESIS STRUCTURE

Since this thesis covers computer-assisted methods supporting various processing steps throughout the clinical workflow of spinal RFAs, it seemed to be most appropriate to align the following chapters with the chronological order of the clinical procedures. This implies that the author will follow the general clinical sequence of a pre-interventional phase, an interventional phase and a post-treatment phase, as introduced in Section 1.3, and subdivide each of the key aspects of this thesis into the usual sections "Introduction", "State of the Art", "Materials and Methods", "Results", "Discussion", and "Conclusion", without losing sight of the overall purpose of the work. This hopefully contributes to the thematic coherence of the individual parts and the readability of the whole thesis.

2 PRE-INTERVENTIONAL PHASE

Subsequent to the initial diagnosis of spinal metastases or lesions and the resulting decision to treat them by minimally invasive RF ablations the therapy and intervention planning phase represents the first part of the treatment workflow. In terms of computer-assistance, the segmentation of relevant anatomical structures represents a pivotal step towards a prospective patient-individual ablation simulation and support during the intervention. In order to numerically simulate heat propagation and related tissue necrotisation volumetric models of all involved tissues are required, i.e. metastases, vertebral bodies and risk structures like the spinal cord. This thesis focuses on the segmentation of the first two structures, as there were still open research issues or limitations in the related literature, while for instance, spinal cord segmentation was adequately addressed in the past by several studies (De Leener et al., 2015; Prados et al., 2016). It is worth mentioning, that this chapter of the thesis does not cover the implementation and design of a numerical simulation itself, which is in development by the cooperation partner Frauenhofer MEVIS (Kröger et al., 2010; Kröger et al., 2006; Weihusen et al., 2010). However, it addresses automatised approaches to provide required prerequisites, which otherwise would have to be produced manually in a very time-consuming and tedious manner. The content of this chapter is based on Hille et al. (2018b) and Hille et al. (2020).

2.1 VERTEBRAL BODY SEGMENTATION

2.1.1 *Introduction*

Although this thesis focuses in particular on minimally invasive interventions of spinal metastases, it is noteworthy that with advancing computer-assisted medicine the segmentation of spinal structures like the vertebral bodies becomes increasingly relevant in other medical fields, too. Prior to the relevance in spinal oncology, the quantitative recording of the spine and thus the vertebrae has a considerable impact on various orthopedic and neuroradiological diagnoses, ranging from scoliosis, stenosis and osteoporosis to vertebral fractures (Brinjikji et al., 2015; Parizel et al., 2010). Besides the diagnostic and therapeutic purposes, automatised segmentation procedures become increasingly important for processing the vast amounts of image data acquired for epidemiological studies (Rak and Tönnies, 2017).

Basically, segmentation methods can be categorised according to the degree of automation, starting with completely manual methods, followed by semi-automatic methods, which for example only require user initialisation, up to fully automatised solutions (Smistad et al., 2015). Each

category offers advantages, but always with the drawback of restrictions or limitations regarding other aspects. While manually performed segmentations most commonly represent the gold standard with regard to the segmentation accuracy and are used as a ground truth for evaluation purposes of more automatised methods, they come at the cost of being highly time-consuming and tedious. With increasing degree of automation, the more challenging and ambitious it becomes to formulate adaptive and generalisable model terms from a priori knowledge and available image information to provide suitable solutions for specific segmentation tasks. In general, it constitutes a balancing act between defining descriptive features as precisely as possible, while being generalisable and capable of representing shape and appearance variabilities of the target structures due to their natural variety as well as different imaging protocols. The goal of most automatised approaches is to reduce the required time and user effort to perform the segmentation task. This, however, usually comes at the cost of a lowered accuracy, since it is hardly possible to represent the vast appearance variability of anatomical structures in a specific model. With the increasing relevance of learning-based methods, the formulation of model terms has been largely eliminated and replaced by the automatised generation of distinctive image features. This has the great advantage that even features can be extracted which would either not have been noticed as important in a manually crafted model, or the implementation of such image information as a model term was not sufficiently applicable. However, learning-based systems must be fed with sufficiently large amounts of training data to widely represent the inherent variance of the data. Since clinical image data is often highly limited, the amount of available patient cases is usually a restricting factor in learning-based segmentation strategies.

While most of the related spine segmentation approaches that focus on CT or radiography (Darwish et al., 2015; Hammernik et al., 2015; Lessmann et al., 2019) benefit from the high contrast of bone tissue as well as the mostly high spatial resolutions, diagnostic MRI became an indispensable technique in clinical decision-making due to amplified soft tissue contrasts and the avoidance of radiation exposure. Besides, CT and radiographic imaging cannot adequately deal with some pathologies like bone tumours and metastases, particularly in early stages (Shah and Salzman, 2011). Therefore, MRI is often essential for diagnostic purposes, for instance, for the scenario addressed in this thesis. However, some characteristics of routine spine MRI tremendously hamper the automation of segmentation approaches. Firstly, highly anisotropic spatial resolution results in partial volume effects and, thus, in blurred delineations between different tissue types especially among adjacent cross-sections (see Figure 2.1). Furthermore, bias field artifacts cause non-homogenous intensities between central and marginal areas. In addition, the image quality and emphasis of different tissues is affected by various imaging parameters, since standarised measurement units like the Houndsfield unit (HU) in CT do not exist in MRI. This means, however, that even native MRI sequences sometimes considerably vary in their image contrasts and

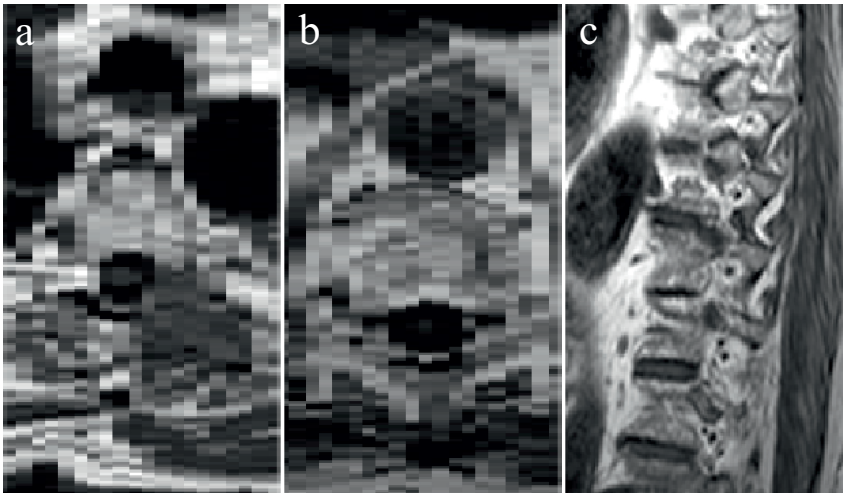

Figure 2.1: Two reconstructed axial (a and b) and one native sagittal slices (c) of a clinical routine T_1-weighted MRI scan demonstrate the difficulty in distinguishing vertebrae from adjacent structures due to the partial volume effects (PVE) caused by high anisotropy. With large sagittal slice spacing, this becomes particularly pronounced at the lateral ends of the vertebrae.

in absolute intensities when it comes to different scanner models or parameter settings. Therefore, a segmentation method that is relevant in clinical settings has to deal with a large variety of MRI sequences and imaging protocols.

2.1.2 *State of the Art*

There are various works in literature regarding vertebral body segmentation in MRI, which differ in their applied methods and the used image data. Besides approaches that were applied to mid-sagittal 2D images (Athertya, Kumar, et al., 2016; Ghosh et al., 2014; Huang et al., 2009), which disregarded valuable spatial information of tomographic imaging, various 3D methods were presented for spinal MRI that will be described in detail below.

Besides largely obsolete approaches based on thresholding, watersheds, and region-growing, segmentation strategies using deformable models, e.g. active contour models (ACM) (Caselles et al., 1997; Kass et al., 1988) or active shape models (ASM) (Cootes et al., 1995) were applied to vertebral body segmentation in spinal MRI. Davatzikos et al. (2002), for instance, trained a deformable shape model to register image data with template images and achieved on average a Dice similarity coefficient (DSC) of $81.5 \pm 3.6\,\%$ on routine images of 14 young healthy volunteers. They used solely T_1-weighted MRI scans with a spatial resolution of $0.93 \times 0.93 \times 3\,\mathrm{mm}^3$.

Štern et al. (2011) also applied a model-based approach, while optimising 29 shape parameters by maximisation of the dissimilarity between inner and outer object intensities driven by image gradients. Their ap-

proach was initialised by marking of each vertebrae in terms of an user input and by approximately specifying the vertebral size as a selection of the respective spine segment (upper/lower thoracic and lumbar). The applied evaluation set contained solely T_2-weighted MRI sequences with in total 75 vertebral bodies of nine healthy subjects, three of them with a resolution of $0.4 \times 0.4 \times 3\,\mathrm{mm}^3$ and six with isotropic spatial resolution ($1 \times 1 \times 1\,\mathrm{mm}^3$). Their approach resulted in an average surface distance (ASD) between the segmented object surface and ground truth points of $1.85 \pm 0.47\,\mathrm{mm}$. Štern et al. (2011) stated a processing time ranging from one to 15 min per vertebra.

Neubert et al. (2011) and Neubert et al. (2012) used ASM to segment both, vertebral bodies and intervertebral discs. They tested their approach on T_2-weighted MRI scans of 14 healthy volunteers with in total 132 vertebrae, acquired with high resolution ($0.34 \times 0.34 \times 1$ to $1.2\,\mathrm{mm}^3$) and achieved a mean DSC of 91 % and a mean Hausdorff distance (HD) of 4.08 mm. However, the average run time per vertebra of 35 min was considerably long. A 10-fold processing time reduction decreased their DSC from 90.8 % (Neubert et al., 2012) to 85 % (Neubert et al., 2011).

Ayed et al. (2012) pursued the idea of formulating the segmentation as a distribution-matching problem. By using an augmented Lagrangian method the distribution of vertebral appearance features was matched to an a priori known reference distribution in order to classify vertebral foreground and non-vertebral background voxel. A mean mid-sagittal 2D-DSC of 85 % was achieved, which leads to the assumption that volumetric quality measures would certainly drop, since most segmentations have difficulties at the lateral ends of the vertebrae due to partial volume effects caused by low laterolateral resolutions or large slice spacings, respectively.

Kadoury et al. (2013) included shape and pose relations between various vertebrae to extend the concept of statistical shape models and to avoid any ambiguities. Using non-linear manifold embeddings improved the shape space representations in contrast to common ASM. They achieved a ASD of on average $2.93 \pm 1.83\,\mathrm{mm}$, while applying solely T_1-weighted sequences of eight subjects with almost isotropic spatial resolution ($1.3 \times 0.9 \times 1\,\mathrm{mm}^3$).

Zukić et al. (2014) combined edge- and intensity-based features, i.e. Canny edges and thresholded gradient magnitudes to a multifeature-based model. Their approach was initialised by a preceding vertebral center detection step using a Viola-Jones detector. The surface mesh of their model was enlarged by balloon forces and constrained by a smoothness term and the approximated vertebral body size. They achieved an average DSC of 79.3 % and a mean ASD of $1.76 \pm 0.38\,\mathrm{mm}$. Their method was evaluated on clinical routine datasets consisting of a variety of MRI sequences including both healthy and pathological vertebrae. Therefore, in contrast to the above mentioned works, their evaluation set was designed to reflect clinical routine imaging.

Chu et al. (2015) fully automatically localised vertebral bodies to define a region of interest (ROI) for a subsequent segmentation step, where

they were using random forest classification for estimating the fore- or background likelihood of each pixel within the produced ROIs. The results were combined with a learned probability map to segment each vertebral body via thresholding. Chu et al. (2015) tested their approach on 23 T_2-weighted images, without stating any pathologies, achieving an overall DSC of 88.7 %, a mean ASD of 1.5 ± 0.2 mm and an average HD of 6.4 ± 1.2 mm. The average computational time per dataset was about 1.3 min.

Korez et al. (2016) introduced a convolutional neural network (CNN)-based approach in spine MRI segmentation. Their method linked active shape models with likelihood maps of the vertebral bodies and achieved an overall DSC of 93.4 %, an average HD of 3.83 mm and a mean ASD of 0.54 mm. Korez et al. trained and tested their methods on the 23 T_2-weighted images made publicly available from Chu et al. (2015).

Goankar et al. (2017) presented a machine learning-based system for vertebral body segmentation on clinical MRI scans of the lumbar spine. In contrast to Chu et al. (2015) and Korez et al. (2016) they examined the applicability of their method to different MRI sequences, though they trained only on T_2-weighted images. The implementation of superpixel-based multiparameter ensemble learning was followed by some morphological post-processing to increase segmentation scores. Goankar et al. had in total 48 sagittal T_2-weighted and 15 T_1-weighted MRI scans and randomly selected six T_2-weighted image volumes for training procedure. The spatial resolution varied in-plane from 0.34×0.34 to 1.1×1.1 mm and the slice thickness was between 0.5 to 5.0 mm. Applying their approach to T_2-weighted images resulted in a mean DSC of 83 %. The segmentation of vertebrae in T_1-weighted images after training on T_2-weighted images expectably resulted in lower DSC scores (on average 75 %).

It is noteworthy that in the time since the elaboration and publication of the work presented in this section, a few more relevant works have been published. Rak et al., 2019, which was the most promising among them, presented a whole-spine segmentation approach for MRI data combining graph cuts including star-convexity constraints and convolutional neural networks, which considerably built on their previous work (Rak and Tönnies, 2017). After a required vertebral patch extraction, a CNN provides likelihood maps in terms of appearance and shape priors, which were the input for a graph cut formulation based on encoding swaps to avoid ambiguous segmentations of neighbouring vertebrae. To evaluate their approach, two databases were used, including T_1- and T_2-weighted image volumes of 64 healthy subjects from an epidemiologic study and 23 T_2-weighted scans from a publicly available source (Chu et al., 2015). Rak et al. (2019) reported a mean DSC of 94.9 %, an average ASD of 0.93 mm and run times below two seconds per vertebra.

The analysis of the existing literature regarding vertebral body segmentation showed that each of the related work was limited by either one or more of the following aspects: insufficient segmentation accuracy, long computational time, limitation to a single MRI sequence or applying image data of higher quality for non-clinical study purposes including

healthy subjects. Accordingly, despite various existing approaches from the related literature, there was still a need for a solution strategy that addresses the specific requirements of the particular clinical purpose in this thesis.

2.1.3 *Objectives*

Considering the settings of clinical routine and the aforementioned goal of supporting therapy planning of spinal metastasis interventions, the following objectives were defined in cooperation with the clinical partner:

- Computational time per patient case < 1 min on current consumer hardware
- Applicable to various MRI sequences for diagnostic purposes (T_1-, T_2-, contrast-enhanced T_1-weighted etc.)
- Segmentation accuracy in the range of the inter-reader variability of field experts

In clinical procedures time plays an important role, expecting suitable computerised approaches to fit into the clinical workflow without significant delay or excessive workload. Supplementary and novel approaches to be integrated should optimally support physicians in clinical decision-making and with repetitive tasks, which tend to be tedious and monotonous, and thus, susceptible to errors due to fatigue. Typically, manual segmentations are repetitive tasks, especially with stacks of slices acquired during 3D tomographic imaging like MRI. Therefore, the support provided by computer-assisted and widely automatised methods in spinal segmentation can drastically reduce the required effort and time. Subsequent to the analysis of the related literature and in consideration of the required time of a manually performed vertebral body segmentation (>10 min per patient case), the objective of a computational time of less than 1 min per patient case was defined. This means a significant reduction of the required time compared to a manual contouring, whereby user interaction, e.g. for initialisation purposes, is still feasible.

Furthermore, approaches with a wide-ranging applicability to clinical routine spinal MRI should deal with various imaging sequences, parameter settings, spatial resolutions, spine sections and healthy vertebrae as well as pathological altered due to fractures, bruises, or metastases. In order to verify the compliance with this requirement, a comprehensive and diverse evaluation set has been compiled with particular focus on clinical data, which is usually highly anisotropic and diverse regarding the imaging protocols. Previous works often shifted away from the challenges of clinical settings by applying their approaches to only one particular MRI sequence or healthy subjects for study purposes.

In order to provide effective support in clinical routine of spinal interventions the segmentation quality should be close to that of an experienced radiologist and in the range of the inter-reader variability (IRV). The

IRV represents an appropriate and robust estimate of an expert-like segmentation accuracy. Methods that produce significantly worse accuracies are only of limited use, since the required post-processing and correction effort would outweigh the time saved by an automatic solution.

Summarising, an approach with reasonable clinical applicability for the addressed scenario should cope with various MRI sequences and imaging protocols, highly anisotropic data, multiple pathological findings in any spine segment, it should require only short computational time, and provide a segmentation accuracy in the range of the IRV.

2.1.4 *Materials and Methods*

Image Data

In order to demonstrate the suitability of the proposed method, an evaluation set was assembled, which included image data of various clinical and research purposes. It consisted of four different databases, including 63 sagittal MRI datasets with overall 419 vertebral bodies of the cervical, thoracic and lumbar spine. The patients or subjects differed in age, sex and presence of spinal pathologies. The evaluation data was acquired in different hospitals and research facilities with various MRI scanners and comprised multiple imaging sequences and protocols (see Table 2.1). Besides this variety, a key characteristic of most datasets was their high anisotropy factor (slice spacing divided by in-plane pixel spacing), ranging from 1.6 to 8.19.

The first database consisted of pre-interventionally acquired MRI data before RFAs of spinal metastases and was representative for the main application case which this thesis addresses. This image data included both vertebrae with metastases from different primary tumours and their adjacent healthy neighbours.

Commonly, the comparison of segmentation approaches and their results between entirely different datasets must be considered as indirect. To overcome this limitation, the evaluation set furthermore consisted of overall 39 image volumes made publicly available together with the related work of Zukić et al. (2014) and Chu et al. (2015). Hence, the produced results could be matched directly with those works. The data from Zukić et al. (2014) included both healthy and pathological spines, e.g. with scoliosis, spondylolisthesis, and vertebral fractures and consisted of various MRI sequences. The third database, published with the work of Chu et al. (2015), comprised 23 T_2-weighted magnetic resonance (MR) images of thoracolumbar spines of voluntary subjects and represented common image data used for research purposes.

The concluding database consisted of epidemiological image data from the Study of Health in Pomerania (SHIP) study (Völzke et al., 2011) and featured spine MR images including T_1- and T_2-weighted sequences. Using this data in particular served the purpose to understand the limits of the presented method regarding low spatial resolution and image quality.

Table 2.1: Characterisation of all datasets used for the evaluation of the proposed method. The individual datasets differed regarding the used MRI sequences (spin echo (SE), turbo spin echo (TSE), fast spin echo (FSE), turbo inversion recovery magnitude (TIRM)), pixel spacing $P_{x,y}$ and slice thickness S_z (both stated in mm), the size of the acquisition matrix M, the anisotropy factor F_A, the number of labelled vertebral bodies $\#_V$, the displayed spine segment SpS (C - cervical, T - thoracic, L - lumbar), the presence of pathology (n.s. - not stated), age, sex, and from whom the reference segmentations originated (neuroradiologists N, trained field experts T). The horizontal lines subdivide the used datasets according to their source (first section - pre-interventionally acquired before RFAs, second section - publicly released by Zukić et al. (2014), third section - publicly released by Chu et al. (2015), fourth section - part of the SHIP (Völzke et al., 2011).

Dataset	MRI_{Seq}	$P_{x,y}$	S_z	M	F_A	$\#_V$	SpS	R	Path.	Age	Sex
preRFA_1	T_1 TSE	0.5	3.3	640 x 640 x 20	6.6	5	C7 - T4	T	-	54	F
preRFA_2	T_1 TSE	0.78	3.3	512 x 512 x 20	4.23	7	T1 - T7	T	+	70	M
preRFA_3	T_1 TSE	0.68	3.3	512 x 512 x 20	4.85	6	T12 - L5	T	+	61	M
preRFA_4	T_1 TSE	0.49	3.3	528 x 528 x 17	6.73	7	T3 - T9	T	+	76	M
preRFA_5	T_1 TSE	0.49	3.3	528 x 528 x 15	6.73	8	T7 - L2	T	+	74	M
preRFA_6	T_1 TSE	0.46	3.3	640 x 640 x 17	7.17	5	T12 - L4	T	+	76	M
Aka2	T_2 FSE	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	21	F
Aka3	T_1 FSE	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	21	F
Aka4	TIRM	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	21	F
Aks5	T_2 FSE	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	22	F
Aks6	T_1 FSE	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	22	F
Aks7	TIRM	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	22	F
Aks8	T_1 FSE	0.70	4	512 x 512 x 15	5.69	8	T10 - L5	T	+	22	F
C002	T_2 TSE	1.12	3.3	448 x 448 x 31	2.96	12	T6 - L5	N	+	74	F
DzZ_T2	T_2 TSE	0.55	4.4	640 x 640 x 12	8.05	8	T10 - L5	T	-	27	M
DzZ_T1	T_1 TSE	0.68	4.4	512 x 512 x 12	6.44	8	T10 - L5	T	-	27	M
F02	T_2 SE	0.5	3.85	768 x 768 x 18	7.7	8	T10 - L5	N	+	51	M
F03	T_2 TSE	1.19	3.3	320 x 320 x 25	2.77	6	T12 - L5	N	+	72	M
F04	T_2 TSE	1.12	3	448 x 448 x 23	2.69	12	T6 - L5	N	+	69	F
S01	T_2 SE	0.47	3.85	640 x 640 x 16	8.19	6	T12 - L5	N	+	65	M
S02	T_2 SE	0.47	3.85	640 x 640 x 16	8.19	7	T11 - L5	N	+	55	F
St1	T_2 SE	0.5	3.85	704 x 704 x 20	7.7	7	T11 - L5	N	+	71	M
Chu (1-23)	T_2 TSE	1.25	2.0	305 x 305 x 39	1.6	7	T11 - L5	T	n.s.	n.s.	F, M
SHIP (1-9)	T_1/T_2 TSE	1.12	4.4	448 x 448 x 15	3.67	5	L1 - L5	N, T	n.s.	29-65	F, M

Methodology

The major steps of the proposed method were as followed (see also Figure 2.2):

1. Initially, a Gaussian filter-based intensity correction was implemented as a pre-processing step to deal with bias field artifacts. The filter kernel size was set to $120 \times 120 \times 30\,\mathrm{mm}^3$ and σ to 20 mm to estimate the bias field of each image volume. In order to remove it, the original image was divided by the bias field estimation. Subsequently, each image volume has been upsampled to ensure spatial isotropy.

2. User interaction initialised the presented approach with three points in a selectable mid-sagittal cross-section to approximate the size, center and sagittal orientation of each vertebral body. For this purpose, both corners of the superior endplate as well as the posterior corner of the inferior endplate were marked. The lateral flection angle could be deduced from interpolating the landmarks' z-coordinates (i.e. in slice direction) of consecutive vertebral bodies.

3. Intensity-based features, e.g. median and variance, were obtained from a cube within the vertebral center and with variable edge length, i.e. two fifths of the specific vertebral body height and length.

4. An abstracted vertebral body shape model was placed upon each vertebral center with the approximate vertebral body length, height and orientation.

5. Within this shape, a pre-segmentation was performed based on adaptive thresholding. The previously gained intensity-based features ensure patient and MRI sequence independence and therefore avoid common difficulties regarding thresholding in MRI. Subsequently, the result was morphologically filtered, first by hole filling and dilating with a 3 mm diameter ball structuring element and followed by removing objects smaller than 1 cm^3. To yield the vertebral body probability map P, the resulting binary image was distance-transformed by a Gaussian convolution (kernel size of $10\,\mathrm{mm}^3$ and σ of 2 mm) and multiplied with the original image volume. This weakened local constraints at the boundaries of the pre-segmented object and enabled level-set convergence away from disadvantageously placed shape models.

6. Boundary feature maps G of each vertebral body were computed by dilating the extracted boundaries of the fitted vertebral body shape model using a 3 mm diameter ball structuring element, subsequently distance-transforming them likewise the probability map and multiplying them with the gradient magnitude images. This feature ensured level-set convergence towards object boundaries

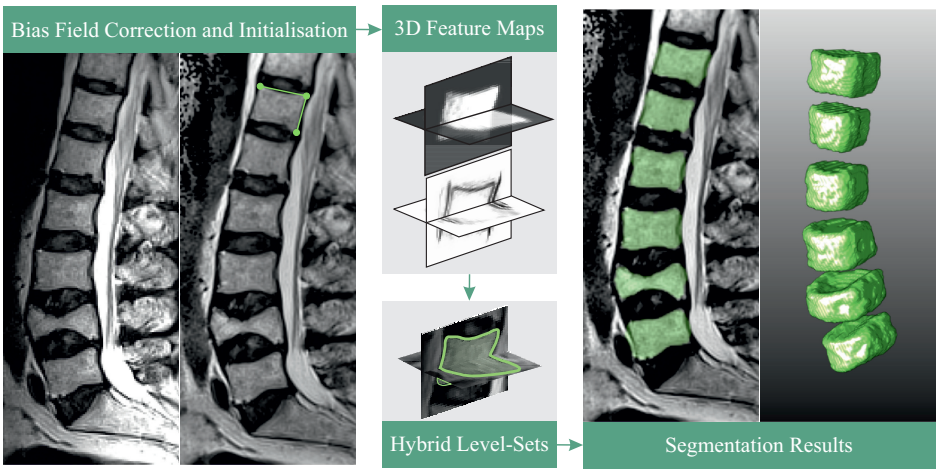

Figure 2.2: Illustration of the pipeline of the proposed vertebral body segmentation approach.

within the range of the model contours. The probability map P and the boundary feature map G defined both terms of the hybrid level-set formulation 2.1.

The proposed 3D bias field-corrected hybrid level-sets (BC-HLS) approach was based on the work presented by Zhang et al. (2008) and combined regional intensity and boundary features with an approximative geometry of the target object in order to steer and constrain the curve towards vertebral body boundaries. Hence, the region information formed a counterweight to attenuate leakage problems frequently emerging in boundary-based methods. The level-set functional to be minimised was defined as:

$$\mathcal{E}(\phi) = -\alpha \int_{\Omega} P \cdot H(\phi) d\Omega + \beta \int_{\Omega} G \cdot |\nabla H(\phi)| d\Omega \tag{2.1}$$

where $H(\phi)$ represented the Heaviside function, Ω was the image domain and the weights α and β were used to balance both terms. The ratio of α to β was empirically determined and set to $4:3$. The probability map P encouraged the level-set contours to enclose regions of a specific per-vertebra intensity range and was defined as:

$$P = g(I_s(\underline{x}), \sigma) \cdot I(\underline{x}) \tag{2.2}$$

where $I(\underline{x})$ was the pre-processed image from step 1 and $g(I_s(\underline{x}), \sigma)$ the result of the distance-transformed and morphologically filtered pre-segmentation $I_s(\underline{x})$ in step 5. The boundary feature map G was the functional of the geodesic active contour term in the hybrid level-set formulation and was defined as:

$$G = g(S_c(\underline{x}), \sigma) \cdot |\nabla I(\underline{x})| \tag{2.3}$$

where $|\nabla I(\underline{x})|$ was the gradient magnitude image and $g(S_c(\underline{x}), \sigma)$ the result of the distance transformation via Gaussian convolution of $S_c(\underline{x})$, which represented a binary image of the extracted and dilated shape model contour.

With ϕ_t defined as a signed distance function:

$$\phi_t = \alpha P + \beta div(G \nabla \phi) \tag{2.4}$$

which could be derived as a simplified iterative approximation of a partial differential equation from the gradient flow applied to 2.1. Like Zhang et al. (2008) initally proposed each iteration step i started with a re-initialisation of ϕ^i and subsequently the embedded function ϕ evolved as an intermediate step with a predefined time step using:

$$\bar{\phi^i} = \phi^i + \triangle_t \alpha P \tag{2.5}$$

After re-initialising $\bar{\phi^i}$, it was updated to ϕ^{i+1} by solving the partial differential equation:

$$\phi_t = \beta div(g \nabla \bar{\phi^i}) \tag{2.6}$$

Evaluation

Ground truth segmentations were available for each dataset created by a radiologist or a trained field expert. For both, the pre-RFA and the SHIP datasets a second reader produced an additional reference segmentation to assess the IRV. In order to quantify the segmentation accuracy, overlap-based and distant-based measures were applied. For that purpose, DSC, ASD, and HD were used to ensure comparability with as many related works as possible. The overlap-based Dice similarity coefficient was defined as:

$$DSC = \frac{2\,|R_1 \cap S_1|}{|R_1| + |S_1|} \tag{2.7}$$

with R_1 and S_1 as foreground voxels of the reference and the produced segmentation. Both distant-based measures referred to the surface delineation of the automatically produced and the reference, i.e. ground truth segmentation. The average surface distant was defined as:

$$ASD(R, S) = |x - y| = \sqrt{\sum |x - y|^2} \tag{2.8}$$

where ASD was the mean Euclidean distance between any point $x \in R$ and their nearest neighbour $y \in S$ with R and S being the surface point sets of the reference and the produced segmentation. The Hausdorff distance HD, which was defined as:

$$HD(R,S) = max\{\check{H}(R,S), \check{H}(S,R)\} \quad with \tag{2.9}$$

$$\check{H}(R,S) = max\{min\{|x,y|\}\} \tag{2.10}$$

being the maximum of both directed Hausdorff distances $\check{H}$ between the two surface point sets R of the reference and S of the produced segmentation in each direction. $\check{H}$ was the maximum distance between any point $x \in R$ and their nearest neighbour $y \in S$ and therefore, represents the worst contour misalignment. Besides the HD, the 95th percentile Hausdorff distance (HD95) can be used especially for learning-based segmentation tasks w.r.t. a more robust percentile than the common 100 % in order to reduce the punishment by single outlier voxels in an otherwise good segmentation.

2.1.5 *Results*

As mentioned before, the IRV was computed for the pre-RFA and the SHIP datasets, ranging from 85.6 % to 91.3 % DSC with on average 88.3 ± 1.6 %, which was similar to the findings of Zukić et al. (2014) (91 %). The mean ASD was 0.84 ± 0.41 mm. Since some of the segmentations were performed by trained field experts, it was additionally examined whether there were significant discrepancies between those segmentations in comparison to two radiologists. The DSCs of both IRV variants barely differed with on average 88.1 % compared to 88.7 %. Similar levels of the IRV were also found for segmentation tasks regarding other organs or structures (Geets et al., 2005; Zou et al., 2004). The proposed BC-HLS approach achieved on average 86.0 ± 3.9 % DSC, 1.59 ± 0.24 mm ASD and 6.86 ± 1.06 mm HD. Average per-vertebra computational time was 5.4 s, whereas the required time was dependent on the spatial resolution and vertebra size. This became apparent considering the computational time differences between the datasets of Zukić et al. (2014) with partially twice the spatial resolution of those datasets of Chu et al. (2015) (see Table 2.2). The complete evaluation results for each dataset are given in the Appendix (see Tables A.1 and A.2).

2.1.6 *Discussion*

In this section, a segmentation approach for vertebral bodies in spinal MRI was presented placing particular importance on clinical requirements and settings. Most of the previously published works regarding this issue were limited solely to study data of healthy subjects and single MRI sequences and thus, largely ignored the difficulties of clinical routine imaging. Furthermore, some of the presented works achieved insufficient

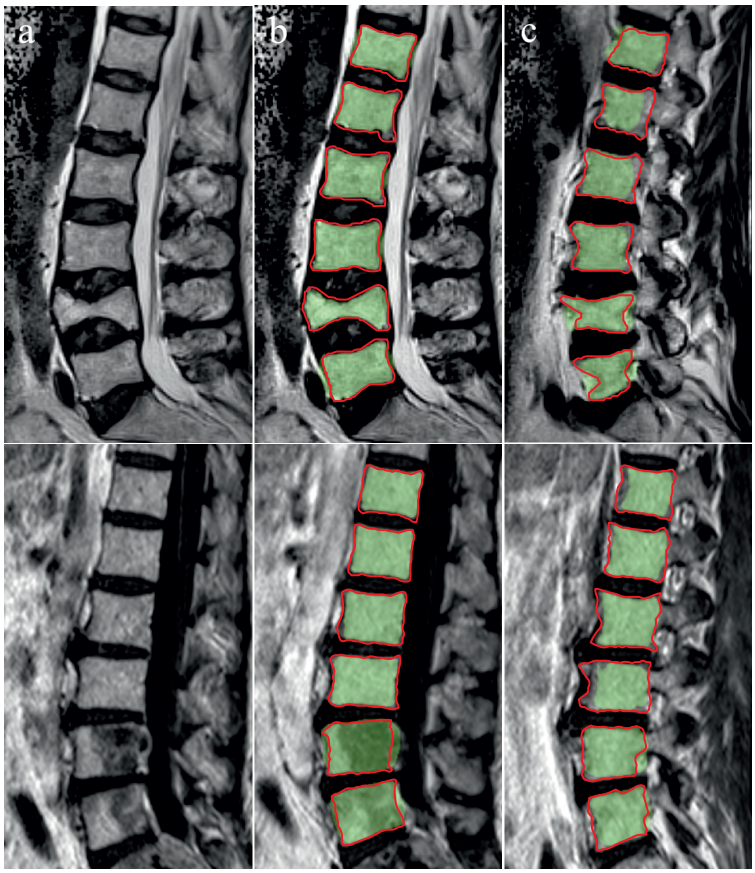

Figure 2.3: Overlay of segmentations produced by the proposed BC-HLS approach (green) and the ground truth (red contours) of dataset *F03* (top row) and *preRFA_3* (bottom row). Mid-sagittal (a and b) and a lateral cross-section are shown (c). The mean DSC of those datasets was 84.3 % and 86.7 %, respectively.

segmentation accuracy or were too time-consuming w.r.t. the above defined objectives. These limitations, however, made them unsuitable for the clinical application addressed in this thesis, and therefore, a novel solution strategy with the BC-HLS approach was presented.

The overall mean DSC achieved with the proposed approach was 86.0 $\pm$ 3.9 %, the mean ASD was 1.59 $\pm$ 0.24 mm and the mean HD was 6.86 mm. With regards to the previously defined objectives, a comparison with the IRV was of critical importance. The latter resulted in on average 88.4 % DSC, indicating that the proposed approach with 86.0 % DSC provided almost expert-like segmentation accuracy, though the standard deviation was considerably higher (1.6 % vs. 3.9 %). Nevertheless, the presented BC-HLS approach achieved on average segmentation accuracies within the range of the IRV (85.6 % to 91.3 %) and thus met the defined clinical objective (see Section 2.1.3). In addition, the required computational time of less than one minute was maintained in all cases, which represents a significant improvement over manual segmentation procedures.

Most discrepancies between the reference and BC-HLS segmentation arose from lateral slices and the impact of PVE, which hamper the detec-

tion of object boundaries (see Figure 2.3). Non-learning-based segmentation strategies are typical trade-offs between data- and model-driven terms, which is why the latter were strengthened at locations of weak image boundaries to avoid leakage problems common for level-set methods (Zhang et al., 2008). However, this increased the dependence on model assumptions, for example the spatial extent of the vertebral body model, promoting under- or oversegmentation at such locations. The correlation of image data and model knowledge is a fundamental issue for any segmentation method based on model knowledge (Neubert et al., 2012; Štern et al., 2011; Zukić et al., 2014). To attenuate level-set convergence towards disadvantageously placed model boundaries, distance-based features were integrated in both the probability and the boundary feature map, in order to enable convergence towards image gradients within the close surrounding. Additionally, pathologies like vertebral fractures or metastases hamper segmentation approaches due to deformations and atypical intensities within and around vertebral bodies, especially if the cortical structures are affected. The latter manifests itself as weak or discontinuous delineations towards surrounding tissues, promoting level-set leakage problems, which could be largely compensated via strengthened local model terms. Nevertheless, the limitations of the presented approach were observed in such rare cases if the vertebra was severely damaged or if the image quality was excessively poor due to artifacts.

A slight difference could be observed in the segmentation quality between T_1- and T_2-weighted sequences, where the latter suffered especially from over-segmentation problems due to similar intensities within the vertebral body and the cerebrospinal fluid of the adjacent spinal cord. This effect was further intensified by PVE due to low spatial resolutions as commonly used in clinical routine spine MRI.

With regard to clinical applicability, the accuracy and precision of segmentation approaches are of primary importance. The overall accuracy of the proposed method was comparable to the state of the art (see Table 2.2), achieving mainly superior results (Goankar et al., 2017; Kadoury et al., 2013; Schwarzenberg et al., 2014; Štern et al., 2011; Zukić et al., 2014), even though few works showed higher accuracies (Korez et al., 2016; Neubert et al., 2012; Rak et al., 2019). However, most of the related work focused only on one particular MRI sequence, despite clinical settings suggesting an applicability for various imaging sequences according to the diagnostic purposes. Furthermore, the high accuracy of the work of Neubert et al. (2012) came at a cost of a considerably longer processing time per vertebra of 35 min (vs. 5.4 s of the proposed BC-HLS method) on recent hardware. A 10-fold processing time reduction decreased their average DSC from 90.8 % (Neubert et al., 2012) to 85 % (Neubert et al., 2011), which was on a par with the results of the presented method, but still unsuitably time-consuming.

Comparable to the evaluation set applied here, only Zukić et al. (2014), Goankar et al. (2017) and the most recent work of Rak et al. (2019) tested their methods on at least two MRI sequences. The first two works each

Table 2.2: Related works and the presented BC-HLS approach in comparison w.r.t. the used MRI sequences, the number of datasets $\#_{DS}$, the number of segmented vertebral bodies $\#_V$, the spine segment SpS (C - cervical, T - thoracic, L - lumbar) as well as the achieved results (DSC, ASD, HD) and the required computation time t_C per vertebra. The superscript to the related works and the proposed approach refer to the used database (^{Z} - datasets publicly provided by Zukić et al. (2014), ^{C} - datasets publicly provided by Chu et al. (2015)).

2D / 3D	Works	MRI_{Seq}	$\#_{DS}$	$\#_V$	SpS	DSC [%]	ED [mm]	HD [mm]	t_C
2D	Huang et al., 2009	T_2	-	52	C, T, L	96	-	-	-
	Ghosh et al., 2014	T_2	13	-	L	84.4 ± 3.8	-	-	-
	Athertya, Kumar, et al., 2016	T_1	16	-	T, L	86.7 ± 4.1	-	5.40 ± 1.12	5.6 s
3D	Štern et al., 2011	T_2	9	75	T, L	-	1.85 ± 0.47	-	1-15 min
	Kadoury et al., 2013	T_1	8	136	T, L	-	2.95 ± 1.85	-	-
	Neubert et al., 2012	T_2	14	132	T, L	90.8 ± 1.8	0.67 ± 0.17	4.08 ± 0.94	35 min
	Zukić et al., 2014^Z	T_1, T_2, TIRM	17	153	T, L	79.3 ± 5.0	1.76 ± 0.38	11.89 ± 2.56	8.3 s
	Schwarzenberg et al., 2014	T_2	2	10	L	81.3 ± 5.1	-	-	19 s
	Chu et al., 2015^C	T_2	23	161	T, L	88.7 ± 2.9	1.5 ± 0.2	6.4 ± 1.2	-
	Korez et al., 2016^C	T_2	23	161	T, L	93.4 ± 1.7	0.54 ± 0.14	3.83 ± 1.04	-
	Goankar et al., 2017	T_1, T_2	57	-	T, L	79 ± 5.0	-	-	-
	Rak et al., 2019	T_1, T_2	128	1412	T, L	93.8 ± 2.6	1.06 ± 0.23	4.06 ± 1.14	1.35 s
	Rak et al., 2019^C	T_2	23	161	T, L	96.0 ± 1.0	0.79 ± 0.25	3.85 ± 2.20	0.9 s
3D	BC-HLS	T_1, T_2, TIRM	63	419	T, L	86.0 ± 3.9	1.59 ± 0.24	6.86 ± 1.06	5.4 s
	BC-HLS^Z	T_1, T_2, TIRM	17	153	T, L	84.1 ± 2.5	1.68 ± 0.24	7.89 ± 1.12	14.0 s
	BC-HLS^C	T_2	23	161	T, L	88.2 ± 1.9	1.66 ± 0.28	6.01 ± 1.01	1.3 s

lacked segmentation accuracy to be considered as sufficiently precise and moreover, Goankar et al. (2017) did not state any computational times.

Nevertheless, a comparison of different segmentation approaches must be considered as indirect, since the evaluated datasets commonly differ from each other and therefore should be interpreted cautiously. Considering this, it is of great benefit if publicly available data could be used for benchmarking purposes. For instance, a straightforward comparison with the work of Zukić et al. (2014) was possible, since they publicly provided a large part of their evaluation datasets with corresponding patient-individually stated results. The following findings and results will refer solely to these shared datasets. Both in mean DSC (79.4 % vs. 84.1 %) and distance measures like mean ASD (1.81 mm vs. 1.68 mm) and HD (12.36 mm vs. 7.89 mm) the presented BC-HLS approach was more precise. Nonetheless, their approach did not inevitably require manual user input, but was based on the detection accuracy of a Viola-Jones detector. Their mean execution time per vertebra was shorter than the presented method (~8.3 s vs. ~14.0 s), however, the beforehand defined objective regarding the computational time could still be met by the proposed method.

Publicly provided datasets of Chu et al. (2015) enabled the direct comparison with their results and those of Korez et al. (2016) and Rak et al. (2019), who applied the same data to their approaches. Again, following results and findings will refer solely to these shared datasets. The results achieved with the proposed BC-HLS approach were on a par with the work of Chu et al. (2015), differing only in the decimal place of the mean DSC or the mean distance measures. Korez et al. (2016) and their deep learning-based method achieved superior results, but aroused doubts about an applicability in clinical settings concerning the computational time. While not stating any information about the processing time, their approach was re-implemented in the course of working on this section's task. To predict a medium sized patch on a consumer GPU it took about 3.75 min, and thus considerably longer than the time required by the proposed BC-HLS approach (1.26 s per vertebra). Additionally, the evaluation data used by Chu et al. (2015) and Korez et al. (2016) included solely T_2-weighted MRI sequences of healthy subjects, whereby no conclusions could be drawn about the transfer to and performance on other MRI sequences or clinical routine patient data. Goankar et al. (2017) rather reported a loss of segmentation quality on unseen MRI sequences. Most recently, Rak et al. (2019) published a work, which surpassed the segmentation accuracy of the proposed BC-HLS approach. They evaluated their method on, amongst others, the publicly available data from Chu et al. (2015) and achieved a DSC of on average 96.0 % and an ASD of on average 0.79 mm, while requiring only 0.9 s per vertebra. Nonetheless, the clinical applicability of CNN-based vertebrae segmentation strategies still needs to be verified with regard to the variety of MRI sequences and pathologically altered vertebrae present in clinical routine imaging.

2.1.7 *Conclusion*

A BC-HLS approach to segment vertebral bodies in diagnostic spinal MRI was implemented and comprehensively evaluated. The method combined regional intensity and boundary features to steer and constrain level-set curves towards vertebral body boundaries. The bias field correction and semi-automatic initialisation with vertebral body size approximation increased the robustness of the segmentation w.r.t. the spine section, imaging sequence and appearance altering pathologies. The segmentation accuracy was sufficient considering the beforehand defined objectives by the clinical partner, while the presented approach was also applicable to a large variety of MRI sequences and parameter settings. By providing results within seconds and requiring only minimal user input for initialisation purposes, the presented method could well be integrated into the pre-interventional workflow. Furthermore, it could be applied to image data with presence of pathologies like fractures, scoliosis or spinal metastases. These are essential requirements for the clinical applicability, which were rarely taken into account by related works so far.

2.2 SPINAL MESTASTASIS SEGMENTATION

2.2.1 *Introduction*

The relevance of treating spinal lesions will further gain importance in the future due to the increased survival time of most malicious carcinomas and the associated probability to develop metastases (Transparency Market Research, 2019). The symptoms make therapeutic solutions with the most immediate effect necessary, since spinal metastases can tremendously affect the quality of life by evoking vigorous pain by fractures, bruises, spinal cord and nerve root compressions or neurologic deficits (Klimo and Schmidt, 2004). Diagnosis and therapy planning can be done with multiple radiological imaging techniques, e.g. planar X-ray radiography, CT, SPECT or MRI. The latter overcomes restrictions of radiation exposure of the aforementioned imaging techniques and offers enhanced soft tissue contrast and direct bone marrow visualisation, which facilitates early lesion detection as well as a robust and reliable assessment of the spatial size and location. Futhermore, the diagnosis and grading of any associated spinal cord compressions is only depictable in MRI (Shah and Salzman, 2011).

Depending on their origin, there are two common types of bone metastases: lytic and sclerotic lesions. The specific type considerably affects the appearance of the metastases in different MRI sequences, ranging from hypo- to hyperintense image signals compared to surrounding spinal structures and non-pathological vertebral bone marrow (see Figure 2.4).

With regard to minimally invasive image-guided interventions like RFA, the segmentation of vertebral metastases represents a pivotal step during the therapy and intervention planning process. Above all, a patient-individual intervention simulation including both the optimisation of the

Figure 2.4: Examples of the appearance variability of spinal lesions (arrows) of different types in STIR (a), T_1- (b) and T_2-weighted MRI sequences (c). Displayed is an epidural metastasis with an osteolytic vertebral body lesion (top row). The sagittal T_1-weighted MR image shows hypointensity with a paraspinal mass, while both the STIR and T_2-weigthed images display the lesion hyperintense. The mid row shows hyperintensity in all acquired MR images, which is typical for benign haemangioma. Sclerotic metastases are displayed in the bottom row, showing characteristic hypointense signals compared to bone marrow in both T_1-weighted and T_2-weighted MRI.

applicator access paths and needle tip positioning, as well as the numerical simulation of the ablative process require an accurate segmentation of the metastases (see Figure 2.5). Furthermore, pre-interventionally defined and segmented target zones transformed onto the intra-operative images can enhance navigation and orientation. This can particularly improve ac-

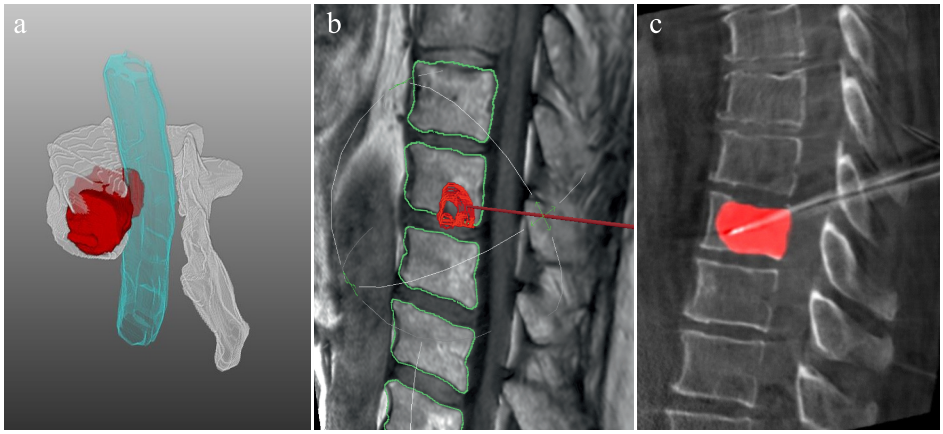

Figure 2.5: Segmentation of spinal metastases can support multiple aspects throughout the therapeutic workflow: interactive visualisations of spatial relations to risk structures (a) and applicator pathway optimisation (b) during the intervention planning, as well as navigation support and visual target zone enhancement during the intervention (c).

curacy and duration of the metastasis puncture, and therefore have a beneficial effect on the treatment outcome. During the post-treatment phase, metastasis segmentations are required in order to perform a computer-assisted treatment outcome validation. Besides these major use cases, other aspects can additionally benefit from such segmented lesions, for instance patient-specific visualisations for pre-interventional patient information (Hille et al., 2017).

In summary, an automatic computer-assisted segmentation of spinal metastases can considerably improve the clinical workflow in several aspects, while drastically reducing the required time and effort, which is generally associated with segmentation tasks of volumetric image data.

2.2.2 *State of the Art*

Segmentation tasks in medical image processing have a long history and numerous solution strategies for a wide range of applications have been developed to date, focussing on all kinds of imaging techniques as well as different structures and organs. The relevance of cancerous diseases suggests that the segmentation of lesions and metastases presents an important subtask within this topic. Besides well-established segmentation methods like threshold-based, region-based, classification- or model-based approaches, deep learning-based techniques have been introduced more recently to lesion detection and segmentation tasks. To this day, however, there are only few publications regarding computer-assisted methods dealing particularly with spinal lesions and most of them are focused on the detection of a specific metastatic type, especially in CT imaging. Therefore, this section will discuss both relevant literature regarding deep learning-based methods of spinal metastases, as well

as some publications focussing on the segmentation of lesions of other organs, i.e. liver and brain tumours in both CT and MRI.

Starting with spinal metastases detection, Yao et al. (2006) presented a computer-aided detection system to find lytic metastases in the spine. Their first step was to automatically segment the spine by thresholding, morphological filtering and a region growing algorithm. After extracting the spinal cord a watershed algorithm was implemented to detect potential lytic vertebral lesions. Subsequently, they assembled a set of 26 features including density, shape, location etc. for each detection and classified them by using support vector machines (SVM). Unfortunately, the authors did not state any quantitative results, neither detection rates and false positives, nor processing times. Furthermore, there was no detailed information whether they tested their method on independent data or how the datasets were split into training and evaluation sets.

Contrary to Yao et al. (2006), Wiese et al. (2012) presented an approach for sclerotic spinal metastases detection in CT images. Their method initially segmented the spine and subsequently detected candidate lesions with a watershed algorithm. Furthermore, 30 features for every detection within the overall 12 clinical cases were computed to train an SVM. They reported a sensitivity rate of 71.2 % with 8.8 false positives per patient case on an independent test set.

Roth et al. (2015) implemented a deep CNN as a second layer in a two-layered cascade framework to spot candidate lesions in CT imaging. Their approach started with a simple thresholding and region growing to coarsely segment the spine. Axial 2D cross-sections were then divided into sub-segments by a following watershed algorithm to spot regions of interest with higher attenuation. After merging neighbouring cross-sections of high attenuation to form 3D seed points for a subsequent segmentation via level-sets, they further classified those candidate lesions with a feature-trained SVM as 'true' or 'false'. This coarse-to-fine method forwarded candidate lesions to the second tier of their detection framework, where a CNN classified them. Each ROI was randomly translated and rotated, resulting in 100 image patches per ROI with 32×32 pixel patch size. While running on high-end hardware, the classification of each ROI took about 30 s. The evaluation set consisted of 532 metastases of 49 patients and they achieved a maximum sensitivity rate of approximately 92 %, though to the account of average 50 false positives per patient. Reducing the sensitivity rate to 80 % and further to 60 % lowers the number of false positives per volume to an average of 9.5 and 1.2, respectively. Similar to Wiese et al. (2012) they focused solely on the detection of the sclerotic lesion type.

While the above-mentioned publications addressed spinal metastases detection in CT imaging, there were even less studies regarding lesions in spine MRI so far. Jerebko et al. (2007) initially started with a spinal cord segmentation and centerline extraction. A median intensity projection along the spinal cord was used to locate discs for separation of the spine into single vertebrae. In a subsequent step the user had to manually place reference points in an axial cross-section to fit an abstract

vertebral body model represented by an ellipse. Within those regions an automatic sclerotic metastases detection was applied, consisting of a primary lesion detection and a subsequent false positive reduction step. To coarsely find candidate lesions a simple adaptive thresholding was implemented, followed by a classification algorithm based on Fisher's linear discriminant (FLD) analysis, which aimed to positively classify at least one candidate of a true lesion, even though multiple false positives were taken into account. Jerebko et al., 2007 tested their approach on 21 patients with 9 diagnosed osteolytic vertebral metastases, resulting in a sensitivity rate of 84.6 % with average 5 false positives per patient. Computational time was stated with approximately 2 min.

More recently, Wang et al. (2017) introduced deep CNNs to vertebral metastases detection in spine MRI. To target the vast variability of size, location and shape they implemented a Siamese deep neural network (SdNN) approach with multiresolution analysis and a weighted averaging of neighbouring cross-sections to benefit from the similarities and to aggregate the detection results. The multiresolution approach produced patches with various spatial resolutions, yielding in total three different representations of the local neighbourhood, while keeping the size of the image matrices constant for a simplified usability within the network. The SdNN comprised three identical multilayer sub-networks to process each image patch resolution. The classification was done on thresholded likelihood maps of each MRI slice predicted by the SdNN. In order to reduce false negatives, they aggregated their 2D likelihood maps using a weighted combination. Wang et al. (2017) assembled their training data from 26 clinical cases, yielding pixel-wise 50 % metastatic samples and 50 % non-metastatic samples. The latter comprised 70 % of randomly selected samples from normal vertebral regions and 30 % from background regions, w.r.t. the more challenging distinction between metastases and healthy vertebral tissue compared to metastases and background. Ten-fold cross-validation resulted in on average 0.2 false positives at a true positive rate (TPR) of 90 %. Most false positives were found in the cerebellum and could be eliminated, e.g. using spine or spinal cord detection.

Even though Wang et al. (2017) and their proposed method can accurately detect spinal metastases resulting in a thresholded likelihood map, those were far from being coextensive with the lesion in the sense of a segmentation task. Their automatic spinal metastases detection can rather be used as an initial step towards subsequent segmentation approaches, either as seed points or as initial contours. To the best of the author's knowledge, the only work addressing the segmentation of spinal metastases has been proposed by Chmelik et al. (2018). They focused on CT images and presented a voxel-wise classification based on a deep CNN with subsequent post-processing to simplify object shapes and produce smooth contours. Starting with an automatic vertebra detection to provide 3D bounding boxes, $16 \times 16 \times 3$ voxel sized patches around each of the bounding box voxels were fed into the CNN to classify them into three categories, i.e. healthy, lytic or sclerotic. Random Forests were used

to reduce false positives as a post-processing step. Their evaluation set consisted of 31 patient cases, including whole spine CT scans, as well as thoracolumbar sections with a spatial resolution of $0.65 \times 0.65 \times 1.00\,\mathrm{mm}^3$. Chmelik et al. (2018) achieved a voxel-wise sensitivity rate of 74 % for sclerotic and 71 % for lytic lesions and a specificity rate of 88 % (sclerotic) and 82 % (lytic). The computational time for analysing a whole-spine CT scan using an NVIDIA Titan Xp GPU was approximately 2 min.

In contrast to the state of the art in spinal metastases, there are various publications addressing the segmentation of hepatic and cerebral lesions. So far, multiple conventional methods were applied to segment brain or liver tumours, starting from thresholding and region-based methods, over clustering approaches to deformable model-based methods (Bauer et al., 2013; Liu et al., 2014). More recently, various learning-based approaches were introduced, either patch-based or with CNNs of an auto-encoder-based architecture. The following analysis will focus on the most promising approaches.

Havaei et al. (2017) applied 2D patch-based networks consisting of two paths, one processed small patches for local features and the other incorporated larger patches to generate rather global and contextual features. Furthermore, they proposed a two-phase training scheme to deal with the highly unbalanced nature of brain tumour segmentation, where the healthy tissue comprises commonly up to 98 % of the total voxel number. The first phase applied data with equiprobable healthy and pathological patches and the subsequent second training phase was performed with a more representative patch distribution. Havaei et al. (2017) trained their approach on the 2013 brain tumour segmentation challenge (BRATS) datasets, consisting of 30 patient cases. They achieved a mean DSC of 88 % on the ten patient cases comprising test set with an average prediction time of 25 s per case on high-end hardware.

In order to incorporate information of the third dimension without drastically increasing the computational costs and the needed graphics processing unit (GPU) memory, works like Lyksborg et al. (2015) tried a compromise solution. They applied an ensemble of three multiplanar reconstructions, i.e. an axial, a coronal and a sagittal view of an MRI volume to a convolutional neural network and applied a subsequent grow cut algorithm to refine the merged segmentations. They achieved a mean DSC of 80 % on the 2014 BRATS challenge data, which consisted of 131 glioma patient cases (91/40 training-test-split).

In contrast to patch-based approaches, auto-encoder-like networks incorporate whole images or volumes, although 3D data is commonly processed as 2D slices due to the required GPU memory and computational costs. Isensee et al. (2017) presented a U-net-like CNN architecture (Ronneberger et al., 2015), while replacing the de-convolutions with upscaling layers and using an exponential linear unit instead of the rectified linear unit (ReLU). They applied their CNN to the 2015 and 2017 BRATS challenge data and achieved on average whole tumour DSC scores of 89.6 % and 85.5 %, respectively.

Shaikh et al. (2017) combined densely connected CNNs with a post-processing dense conditional random field (CRF). The former consisted of the common down-sampling path but intertwined with blocks of densely connected layers. The dense CRF was applied to reduce false positives and to smoothen the segmentation. In total, their approach took less than 30 s for the proposed pipeline of segmentation and post-processing, while achieving a mean DSC of 83 % on the 2017 BRATS challenge test data.

Besides the above-mentioned publications regarding cerebral lesions, there are various works focussing on liver tumour segmentation. Christ et al. (2017) trained two cascade networks to segment both the liver and its lesions in MRI and CT images. After pre-processing the image data, a fully connected CNN segmented the liver (achieving 94.3 % mean DSC on CT data) and forwarded the predicted ROI to the subsequent U-net in order to segment lesions within it. A 3D CRF likewise to the work of Shaikh et al. (2017) was applied to further refine the resulting segmentations, resulting in a lesion DSC of on average 85 % on CT data. Applying MRI data to their cascade CNN instead resulted in a mean liver DSC of 87 % and on average 69.7 % DSC w.r.t. the lesions. Christ et al. (2017) stated a computational time of less than 100 s for both segmentation steps combined.

Another promising approach towards liver lesion segmentation in CT imaging was presented by Li et al. (2017). They combined two densely connected U-Nets: the first extracts 2D intra-slice features, while the following incorporates the original 3D data together with the output of the former network in order to aggregate volumetric context. Therefore, they could alleviate common limitations of 2D approaches, which largely ignore the volumetric context and reduce the computational costs of 3D convolutions by transforming the feature and score maps obtained from the 2D Dense-U-net to a volumetric shape. Data from the 2017 liver tumour segmentation challenge (LITS) was used for evaluation purposes, resulting in averaged DSC scores of 96 % for the liver and 72.2 % regarding the liver lesions. Computational time per patient case ranged between 30 s to 200 s depending on the number of available slices.

In conclusion, while there are several works addressing lesion segmentation in brain and liver, spinal metastases segmentation continues to be a great challenge and is a hardly addressed issue so far, despite the increasing relevance. Currently, the state of the art regarding computer-assisted methods for spinal lesions focuses almost exclusively on the detection both in CT and MRI images. To the best of the author's knowledge, besides the work of Chmelik et al. (2018), no segmentation approaches, especially in MRI, have been published so far.

2.2.3 *Objectives*

Similar to the previous section about vertebral body segmentation, the following objectives were defined in cooperation with the clinical partner:

- Computational time per lesion on average < 1 min on current consumer hardware
- Segmentation accuracy in the range of the inter-reader variability of field experts

Likewise to most medical segmentation tasks, a computer-assisted and automatised approach dealing with spinal lesions should relief radiologists from the time-consuming and tedious task of manually performed contouring of target structures, while providing a segmentation accuracy comparable to the gold standard. Since this task is not part of the current clinical process (see Section 1.2), it is of great importance that an integration into the intended workflow as described in Section 1.3 would result in as little delay and additional workload as possible. Moreover, any added effort must be justified by the quality of the results, which is essential for the implementation of new routines and procedures. Therefore, the goal of this task was to provide expert-like segmentation accuracy with minimum time and interaction requirement. In cooperation with the clinical partner a maximum computational time of one minute was defined in order to significantly reduce the time needed for an alternatively manual segmentation, which usually requires between 5-10 min per metastasis. Since it is difficult to define a particular accuracy value as a threshold for sufficient and satisfactory segmentation results, it seemed to be most appropriate to define an accuracy in the range of the IRV as the goal.

Despite the increasing relevance of this clinical issue, there is hardly any state of the art regarding the automatised segmentation of spinal metastases, which is most likely due to the difficulty and complexity of this ambitious task. Spinal metastases are highly variable in terms of their shape, texture and the emitted image signals, which is particularly dependent on their lesion type and the used MRI sequences. This led to the conclusion that model-based approaches might be unsuitable, since they require substantial a priori knowledge, which is difficult to model in this case. Learning-based approaches, in particular deep convolutional neural networks, are well suited for such tasks due to their ability to automatically find and extract descriptive and distinctive features. Therefore, to meet the clinical objectives the implementation of a deep learning-based approach seemed to be the most promising and suitable.

2.2.4 *Materials and Methods*

Image Data

The evaluation set, which was applied to the proposed approach contained patient cases that underwent radiofrequency ablations of both

single and multiple spinal metastases, mostly in advanced tumour stages. In total, 40 metastases were assembled for this work, originating from renal cell, prostate, cervical, colon, pancreatic, breast, bladder, stomach, lung, caecal, urothelial and spinocellular carcinomas. For diagnostic and therapy planning purposes spinal MRI was performed including sagittal native T_1- and T_2-weighted MRI sequences. Acquisition settings, e.g. magnetic field strength or repetition time, varied within the datasets. The scan resolution ranged from 0.45 to 1.25 mm in-plane and from 3.3 to 4.8 mm in-depth. The acquired MRI data was pre-processed by registering cohesive MRI sequences patient-wise to the respective T_1-weighted image and by cubic resampling of each image volume to a total number of 64 sagittal slices. Hence, the image data was of almost isotropic spatial resolution, while maintaining a fixed slice number to simplify any further processing. A field expert trained by neuroradiologists manually contoured each metastasis slice-wise, producing a binary 3D segmentation mask with a defined center point m_c.

Data Augmentation

Although deep learning-based methods have become evermore popular and were applied to various medical image segmentation tasks, they commonly suffer from the challenge of accumulating sufficiently large datasets to ensure promising training with CNNs, which usually consist of millions of parameters to be optimised. The generalisation ability of such networks, i.e. the performance on unseen test data and the avoidance of overfitting to the training data, can be significantly improved by data augmentation techniques. Besides strategies regarding the network's architecture itself, as well as functional regularisations, data augmentation methods artificially generate new training samples by manipulating and transforming the original images and their correspondingly labelled ground truth (Shorten and Khoshgoftaar, 2019). Accordingly, it is possible to provide larger training datasets for clinical image analysis tasks with comparatively few original patient cases and - equally important - without the vast manual effort of experts for labelling the data. In addition, the amount of available image data acquired during expensive and time-consuming MRI examinations is usually even more limited. Nevertheless, despite all the merits of data augmentation in medical image processing, the differences between artificially altered training images and unseen test data, especially due to the complex natural variety of anatomical structures, constitute the limits of such strategies (Shorten and Khoshgoftaar, 2019).

Considering the segmentation of spinal metastases, the number of 40 available patient cases was considerably small for a CNN-based approach. Therefore, each of the original T_1- and T_2-weighted MRI volumes was augmented using the following techniques:

- *Mirroring:* Flips were applied to each patient volume in all directions. Although for example vertical, i.e. craniocaudal, flips did not occur in the original data and seemed to be inappropriate, it had

proven to be advantageous for the final results, since it prevents fast overfitting.

- *Scaling:* The image volumes were scaled with randomly chosen factors between 0.6 and 1.4.
- *Rotation:* The image volumes were rotated in the range of $\pm$ 30 $^\circ$ around the transversal axis and between $\pm$ 20 $^\circ$ around the sagittal axis.
- *Elastic deformations:* Elastic deformations were applied using random displacement fields with subsequent Gaussian smoothing the grid with a σ ranging between [0, 0.3] (cf. Ronneberger et al., 2015).
- *Gaussian blur:* A Gaussian filter with σ in the range of [0, 0.5] was applied to blur the images.
- *Gamma transformation:* Gamma transformations were applied with γ in the range of [0.5, 2] to modify image intensities.
- *Random cropping:* Each patient volume was translated in a range of $\pm$ 20 voxels in sagittal and vertical direction w.r.t. the center of the metastasis m_c in a random cropping manner and resulted in patches of size 128 $\times$ 128 $\times$ 64.

In conclusion, each augmented volume was whitened by mean subtraction and a subsequent division by the standard deviation. With respect to a stratified cross-validation, the patient data was grouped into 8 folds, splitting each fold into varying 35 training and 5 test patient cases. By using these augmentation techniques 5,250 volumetric and 336,000 cross-section training samples, respectively were generated per fold.

CNN Architecture

With regard to the fact that the commonly used U-net architecture from Ronneberger et al. (2015) was still state of the art in terms of various medical segmentation tasks (Isensee et al., 2018), a minimally modified U-net was implemented, which incorporated 2D image data. In addition, an extension to volumetric data, called vU-net (Milletari et al., 2016) was used to incorporate 3D image volumes. Both variants represented encoder-decoder architectures well suited for medical images, whose decoder paths combined semantic information from the deeper layers of the networks with higher resolution feature maps from the encoder paths via skip connections (see Figure 2.6). The implemented U-nets processed 2D patches of size 128 $\times$ 128, the vU-nets volumes of size 128 $\times$ 128 $\times$ 64. In both network variants each convolutional layer was followed by a batch normalisation and had a kernel size of 3 $\times$ 3 ($\times$ 3), except for the last one which applied a 1 $\times$ 1 ($\times$ 1) kernel to reduce the dimensionality to the desired output size. Furthermore, up-convolutions were replaced by simplified upsampling layers, which have been found to be equally effective while being less computationally expensive (Isensee et al., 2017).

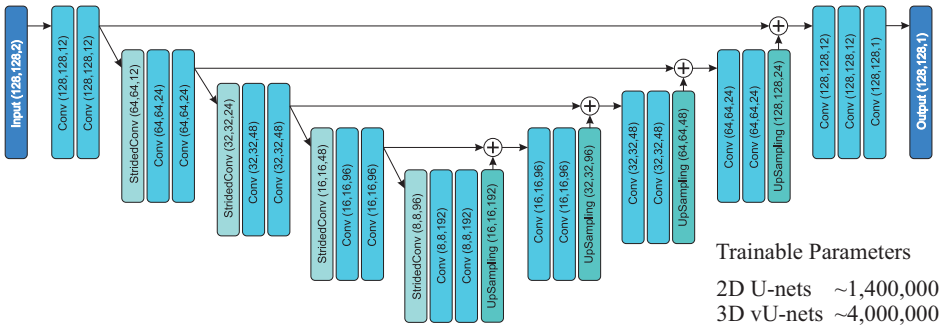

Figure 2.6: The U-net architecture used for 2D image input, with convolutional layers including batch normalisation, strided convolutions for downsampling and upsampling layers. The vU-net architecture for 3D input was analogous, but with an additional dimension for all layers and convolutional kernels. A significant difference between the two variants was the number of trainable parameters, which was about 2.85 times higher for the vU-net variants.

A ReLU was used as activation function for all convolutional layers, except for the last one again, where a sigmoid function was applied to provide values between 0 and 1. Multimodal image input was incorporated in the most straightforward way, i.e. each MRI sequence was represented by an input channel.

The training set of each fold consisted of 5,250 randomly shuffled volumes as 3D input of the vU-nets or 336,000 randomly shuffled slices as 2D input of the U-nets. A single epoch was used, while the number of iterations was equal to the number of available samples. A modified form of the Tversky index (Tversky, 1977) was used as a loss function. The Tversky loss (TL), as proposed by Salehi et al. (2017), was defined as

$$\mathrm{TL}(\alpha,\beta) = \frac{2\sum_{i=1}^{N}(r_{0i}p_{0i})}{\sum_{i=1}^{N}(r_{0i}p_{0i}) + \alpha\sum_{i=1}^{N}(r_{1i}p_{0i}) + \beta\sum_{i=1}^{N}(r_{0i}p_{1i})} \tag{2.11}$$

where p_{0i} was the probability for a voxel i to be a lesion and p_{1i} to be non-lesion. For a lesion voxel r_{0i} is 1 and for a non-lesion voxel r_{0i} is 0, vice versa for r_{1i}. The weights α and β affect the penalties for false positives and false negatives. Adam was used as an optimizer (Kingma and Ba, 2014) with a starting learning rate of 0.01 and mini-batch sizes of two volumetric samples for the vU-nets and 32 for the slice-wise input data of the U-nets. Finally, a threshold of 0.5 was applied within the output layer to producing binary images. The chosen hyperparameters were investigated in preliminary studies with smaller training sets (see Table A.3).

Experimental Design

Multiple network configurations were assessed in terms of their ability to segment spinal metastases. For this purpose, the augmented data was

applied to both network architectures as described above. Despite of the additional dimension of the vU-nets, the basic architecture for both variants was not changed in order to largely exclude further influencing factors and to ensure comparability of the results. The experiments were subdivided according to the used dimensionality, i.e. U-nets and vU-nets, and the input modalities, i.e. single or combined MRI sequences. The training design consisted of stratified 8-fold cross-validation over disjunct subsets of five patients per validation set. The stated results represent the average of all 8 cross-validation folds. Due to the small amount of available data and since no training and design decisions were based on intermediate validation results (no look-ahead bias), it was decided to refrain from a separate test set, as it would have resulted in too few samples for a promising training.

Evaluation

Expertly annotated lesion segmentations were manually produced using co-registered MRI sequences of each patient within a segmentation framework designed for this purpose. Following the ground truth preparation, the resulting labelled image data was fed into the networks described above. To quantify the produced segmentation quality, DSC was used to measure the percentage of volume overlap, as well as voxel-wise TPR and true negative rate (TNR), since some of the related work used both as quality measurements. The DSC was already defined in Section 2.1.4 and the sensitivity (TPR) and specificity (TNR) are defined as follows:

$$\mathrm{TPR} = \frac{|R_1 \cap P_1|}{|R_1|}, \quad \mathrm{TNR} = \frac{|R_0 \cap P_0|}{|R_0|} \tag{2.12}$$

with R_1 and P_1 as foreground voxels of reference and prediction, and analogously R_0 and P_0 as corresponding background voxels. However, sensitivity and specificity are not commonly used to evaluate medical image segmentations since they are highly sensitive to the segment size (Taha and Hanbury, 2015), but they were taken into account since a part of the related work used them as quality measures. The given quantitative results were generated exclusively on patient volumes, even if the U-nets predicted slice-wise segmentation masks, which were subsequently merged patient-wise.

Furthermore, the required manual input to mark the approximate center of each ROI around the metastases for the prediction step meant that the proposed approach was not fully automated end-to-end. However, this prerequisite was justified by the clinical setting which particularly focused on the support of the therapeutic workflow and thus the approximate location of a specific lesion is known due to the diagnostic procedures. However, a preceding automatic detection step as presented by Wang et al. (2017) could substitute this manual user interaction.

2.2.5 *Results*

In order to reasonably classify the achieved results, it was mandatory to investigate the inter-reader variability to get an impression of the congruence, if two expertly annotated segmentations of a specific metastasis were matched. Therefore, a second ground truth was produced from a randomly chosen subset of 15 metastases by a neuroradiologist and afterwards matched with the corresponding primary segmentation masks. For this purpose, DSC, ASD, and HD95 were determined. The computed DSC values per metastasis ranged from 70.4 % to 89.7 %, with an average of 79.4 $\pm$ 6.2 %. The surface distances between two expert segmentations were on average 1.81 $\pm$ 0.59 mm ASD and 7.83 $\pm$ 3.59 mm HD95. The averaged results regarding the segmentation accuracy of the proposed automatic approach can be found in Table 4.1. Averaged over all cross-validation folds, DSC values up to 77.6 $\pm$ 10.8 %, mean sensitivity rates of 78.9 $\pm$ 15.8 % and mean ASD of 2.21 $\pm$ 1.57 mm could be achieved dependent on the applied MRI sequences and the used network variant. Predicting the segmentation of a patient case required approximately 0.29 s (4.51 ms per slice) for the U-nets and on average 0.23 s for the vU-net architecture.

2.2.6 *Discussion*

The segmentation of spinal metastases was of high importance for the addressed scenario of this thesis, since multiple aspects throughout the entire workflow can benefit from the results if these are sufficiently accurate. Therapy planning and simulation, enhanced visual support during the intervention, as well as post-RFA treatment outcome validation require segmentation masks of the metastatic tissue. Since there hardly existed any state of the art regarding this particular issue, a novel approach had to be developed and tested w.r.t. its clinical applicability. For this purpose, an automatic and deep learning-based segmentation method for spinal MRI was proposed.

In order to assess the achieved results of the experiments regarding their general segmentation quality, it was convenient to compare them with the inter-reader variability of manually produced expert segmentations, which could be seen as an indicator of the complexity of a segmentation task. With on average 77.6 % DSC, the results achieved with the U-net incorporating combined T_1- and T_2-weighted input were close to the mean IRV with 79.4 % and well within the its standard deviation of $\pm$ 6.2 %. A similar conclusion could be drawn considering the surface distance measures like ASD, where the IRV showed on average slightly minor discrepancies compared to the automatically produced segmentations (1.81 $\pm$ 0.59 mm vs. 2.21 $\pm$ 1.57 mm). In general, it could be concluded that the achieved results were largely on a par with expertly annotated segmentations, but lacked their robustness and reliability with regard to a few more challenging patient cases, which was reflected in the significantly higher standard deviation (10.8 % vs. 6.2 %).

Table 2.3: Experimental results of each input configuration depending on the used modalities (T_1-, T_2-weighted MRI sequences), as well as a slice-wise (U-net) or volumetric (vU-net) processing. Dice scores (DSC), sensitivity rates (TPR), specificity rates (TNR), average surface distances (ASD) and 95th percentile Hausdorff distances (HD95) were averaged over all cross-validation folds.

		U-net			vU-net		
		T_1	T_2	$[T_1 + T_2]$	T_1	T_2	$[T_1 + T_2]$
DSC [%]	mean	77.4	65.4	77.6	73.7	64.7	74.8
	median	81.8	73.8	80.2	77.5	71.8	78.3
	std	12.4	21.7	10.8	15.6	20.8	13.6
TPR [%]	mean	76.2	71.9	78.9	71.9	65.2	73.7
	median	82.6	77.4	82.7	76.5	72.2	79.6
	std	17.4	21.6	15.8	20.2	24.51	18.6
TNR [%]	mean	98.5	97.5	98.4	98.4	97.85	98.33
	median	99.3	98.4	99.2	99.5	98.7	99.2
	std	2.0	3.3	1.9	2.1	2.7	2.1
ASD [mm]	mean	2.40	2.97	2.21	2.62	3.50	2.61
	median	2.00	1.91	1.79	2.06	2.37	1.84
	std	1.60	2.24	1.57	1.93	2.54	1.88
HD95 [mm]	mean	11.55	12.90	11.12	12.83	15.22	12.28
	median	11.26	7.30	11.14	9.88	10.45	9.50
	std	6.85	11.04	6.57	8.98	12.37	8.11

Besides a sufficient segmentation accuracy, the required time to compute these results was of critical importance and thus defined as an objective of this task. Regardless of the input configuration and network variant the trained models required well under half a second to predict the segmentation masks. In contrast to the time-consuming manual contouring of metastases, the proposed CNN-based approach instantly provided results and therefore will not constitute any delay if integrated into the clinical workflow. Even with a manual post-correction based on the predicted segmentation masks, significantly less time would be required compared to a fully manual alternative.

To go into detail, some representative cases are shown in Figure 2.7, depicting satisfactory results as well as challenging cases with exemplary inaccuracies. Among the latter, falsely positive classified voxels, mostly in

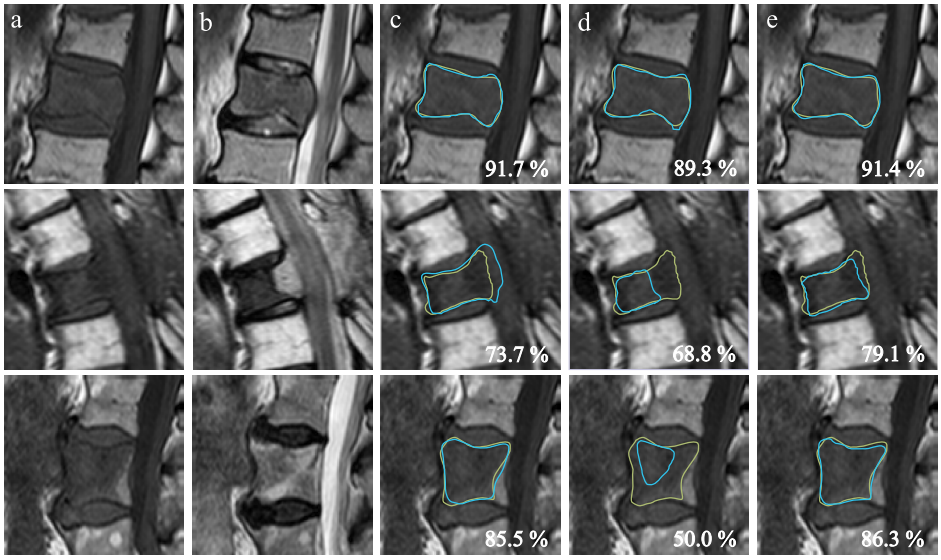

Figure 2.7: Comparison of expertly annotated data (green contours) and automatically produced segmentations (blue contours) using U-nets for three exemplary cases (rows). DSC scores in the lower right corners indicate the segmentation accuracy for each patient case w.r.t. the used MRI sequences. From left to right: (a) original T_1-weighted MRI sequence, (b) original T_2-weighted MRI sequence, (c) result with only T_1-weighted image data, (d) result with only T_2-weighted image data, (e) result with combined T_1- and T_2-weighted image data.

the adjacent tissue of the vertebral bodies, and inaccurate segmentations at the transverse and spinous processes were the most common. Segmentation tasks in this particular area are highly ambitious, since a variety of anatomical structures with similar intensities and textures are in close proximity. Paravertebral and epidural extensions of the metastases were particularly difficult to distinguish from surrounding tissues, which even hampered manual contouring by experts as indicated by the relatively low agreement of two experts. Furthermore, exceptionally shaped metastases, especially if they were not roughly star-convex or metastases with highly heterogenous image signals represented a challenging task for automatic approaches like CNNs. This, however, was very likely due to the fact that the required level of variance was not represented within the provided training set and could only be moderately addressed by data augmentation, which was also indicated by the high standard deviation. Hence, the limited amount of unique training data remained a restricting factor, as previously outlined in Section 2.2.4.

Regarding the used MRI sequences, it was found that T_1-weighted images were the most valuable, since the best results could be achieved in experiments either with T_1-weighted data alone or if it was part of a multimodal image input (see Figure 2.8). This could be attributed to the predominantly uniform appearance of the metastases in T_1-weighted images, whether lytic or sclerotic, which appear hypointense compared to surrounding bony structures. This conclusion has also been confirmed by findings of relevant literature, according to which T_1-weighted MRI

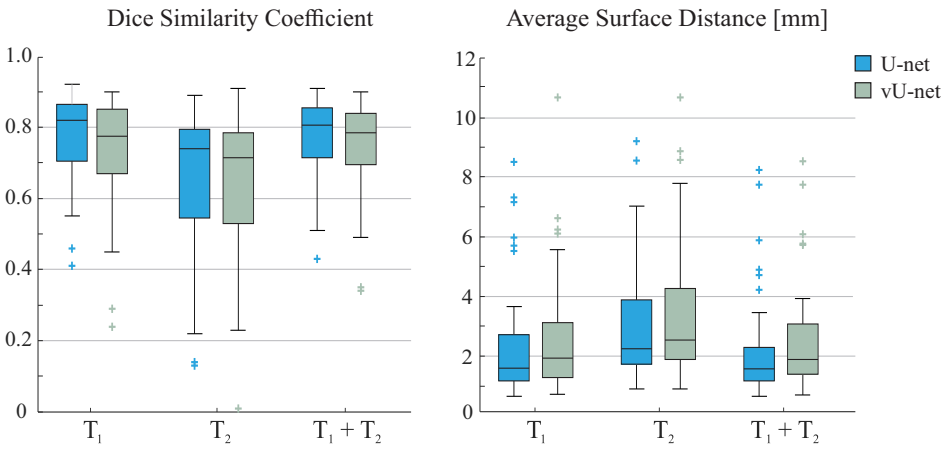

Figure 2.8: Measured DSC and ASD scores depending on the imaging modality and input dimension. Box edges mark the 25th and 75th percentiles, the central box line marks the median value and the whisker marks the most extreme values not considered as outliers.

sequences were found to be most suitable in order to detect any abnormalities of the vertebral bone marrow (Shah and Salzman, 2011). Combining T_1-weighted images with T_2-weighted MRI data showed no significant differences in the mean accuracy compared to an input of exclusive T_1-weighted images (77.6 % vs. 77.4 %), although the standard deviation could be reduced (12.4 % to 10.8 %). The experiments using solely T_2-weighted images yielded the worst results, presumably due to the fact that lytic and sclerotic lesions most notably differ in this particular imaging sequence and therefore, present conflicting information to the CNNs if in the same training set. Hence, T_2-weighted images rather supported and improved the robustness in combination with T_1-weighted input than yielding satisfactory results themselves.

The conclusions above regarding imaging sequences hold true for both network variants, i.e. for slice-wise and volumetric input data. In general, the achieved DSCs with the U-nets were on average 2.4 % higher than with vU-nets. As previously indicated in Section 2.2.4, the ratio between available image data and the number of trainable parameters of CNNs considerably influences their performance. Hence, the discrepancy in the results could be attributed to the increased complexity and number of trainable parameters (1,400,000 vs. 4,000,000) to be optimised when extending the network from 2D U-nets to 3D vU-nets.

It is difficult to compare the achieved results with related works, since there are, to the best of the author's knowledge, no studies regarding automatic spinal metastases segmentation in MRI. Thus, the discussion referred to CNN-based segmentation approaches of liver and brain lesions as well as a work by Chmelik et al. (2018) for spinal CT data. Depending on the used datasets, CNN-based brain tumour segmentations in MR images achieved a DSC up to 88 % (Havaei et al., 2017) in the BRATS challenge 2013, 89.6 % (Isensee et al., 2017) in the BRATS challenge 2015 and 85.5 % (Isensee et al., 2017) in the BRATS challenge 2017. The segmentation of liver lesions in MR images achieved Dice coefficients of 69.7 %

(Christ et al., 2017) and in CT images up to 72.2 % (Li et al., 2017). The stated results enabled an assessment of the segmentation accuracy of the presented method compared to the state of the art in other lesion segmentation tasks. However, the differences between the individual application scenarios are perhaps more complex than their similarities.

Chmelik et al. (2018) were one of the first to adapt a CNN to vertebral metastases segmentation in CT images. They achieved a mean voxel-wise sensitivity rate of 74 % for sclerotic and 71 % for lytic lesions as well as a mean specificity rate of 88 % (sclerotic) and 82 % (lytic). In comparison, the results achieved by the proposed approach including T_1-weighted images are somewhat better (mean sensitivity of 77.6 %), though the experiments using solely T_2-weighted MRI data clearly lacked accuracy. Additionally, it is important to account for the differences in spatial resolution (slice thickness of 0.67 mm in the work of Chmelik et al. (2018) vs. average 3.50 mm of the data applied to the presented approach) and the effects of high spatial anisotropy, i.e. partial volume effects, which considerably hamper the segmentation.

2.2.7 *Conclusion*

In this section a novel approach towards spinal metastasis segmentation in diagnostic MR images based on deep CNNs was presented. For this purpose, a dataset including metastases of both lytic and sclerotic type was comprised and the impact of various input modalities and network configurations on the segmentation accuracy was examined. The achieved experimental results have been quantitatively compared to the inter-reader variability and results in literature, although the latter mostly focused on other organs or imaging modalities. Due to the absence of directly comparable works and the beforehand defined clinical objectives, the IRV was used to assess the quality of the achieved results. With 77.6 % DSC on average the result of the best network configuration of the presented automatic segmentation approach was on a par with the IRV (79.4 %), indicating reasonably accurate segmentation quality and a promising approach towards this ambitious and challenging task. Accordingly, the presented approach met the clinical requirements defined in Section 2.2.3 regarding the required maximum computational time and segmentation accuracy. Therefore, it could be integrated into the intended workflow without delaying it, while automatically producing results of almost expert-like accuracy.

3 INTERVENTIONAL PHASE

After the diagnosis of spinal metastases and the decision for RFA as the most suitable treatment method, the interventional phase of the clinical workflow starts. As already described in Section 1.3 within the introductory chapter, the intervention is performed using image guidance of FP-CT scans and partly biplanar angiography to ensure subcutaneous navigation and tumour puncture. During the intervention, at least one, commonly two applicator needle tips are placed within the lesion (Matschek et al., 2017) and coagulate the surrounding tissue due to high-frequency current-related heating. Based on the continuous measurement of the specific tissue impedance, the ablation time and induced energy, the progress of the ablation and the resulting necrosis zone can be approximately derived. If the impedance value exceeds a certain threshold, since the current flow by means of ionic movement and molecular friction no longer leads to significant tissue heating, this indicates a finished ablation within the target zone and thus the achievement of the interventional goal. A more detailed assessment of the results of the ablation process can only be achieved by the examination of follow-up MRI scans subsequent to the intervention. While the post-RFA procedures will be discussed in Chapter 4, this chapter will address computer-assistance during the intervention itself with the content based on Hille et al. (2018a).

3.1 MULTISEGMENTAL SPINE IMAGE REGISTRATION

3.1.1 *Introduction*

Following the survey work of Viergever et al. (2016) and Maintz and Viergever (1998), respectively, medical image registration considerably gained relevance throughout the last two decades, in particular in the fields of radiotherapy and interventional radiology. Therapy planning, interventional navigation, response monitoring and treatment outcome validation, respectively benefit from image registration techniques, with more and more works that have made the transition from research to clinical settings (Viergever et al., 2016). Likewise, the motivation for this aspect of this thesis was the support of interventional navigation and lesion puncture during RFAs of spinal metastases using a fast and precise multimodal image fusion.

FP-CT and CT angiography are the most common imaging methods regarding image guidance during osseous RFAs (Kavakebi et al., 2017; Palussiere et al., 2012; Posteraro et al., 2004). However, low-dose protocols like intra-interventional FP-CT scans provide a reduced image quality and weak soft tissue contrasts compared to native CT or MRI. This, however, manifests in a decreased signal-to-noise ratio (SNR), beam hardening and

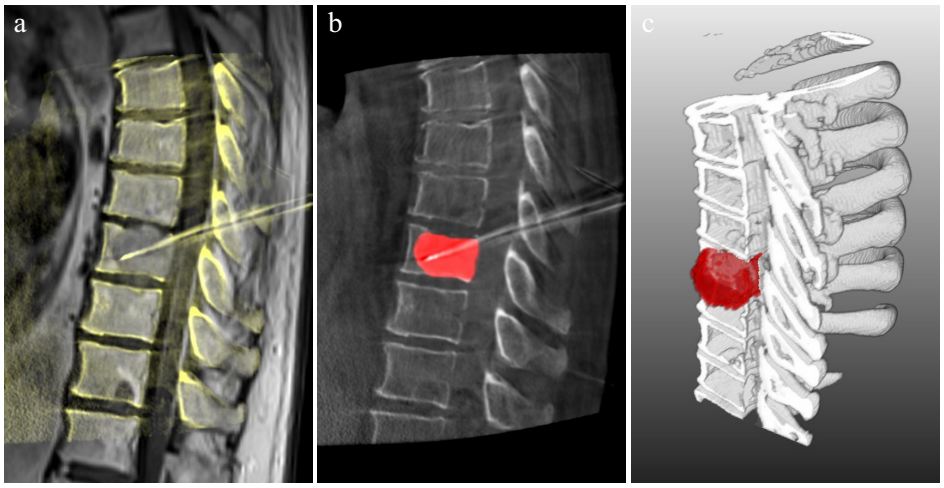

Figure 3.1: Image fusion of pre- and intra-operative images represents a significant improvement of the interventional routine. As a result of the image registration (left; background: MRI, overlay: FP-CT) a transformation matrix could be used to transfer and display pre-interventionally produced information like contoured metastases within the intra-interventional images (middle, right).

scatter artifacts, which hamper precise and reliable metastasis localisation during the interventions. Due to the poor visibility, the radiologists have to infer the metastasis location from pre-interventionally acquired MRI data and mentally match those images with the intra-interventionally acquired FP-CT scans. Thus, an exact localisation is only feasible to a certain degree by orientation on the visible cortical bone structures of the vertebrae. Moreover, each interventional image during the RFA is acquired in prone patient position, causing intervertebral joint movements and altered spine flexion compared to the diagnostic MR images. This aspect further increases the cognitive load of the radiologists for metastasis puncture, particularly if several metastases are treated in a single procedure or if metastases developed large paravertebral extensions that may involve multiple vertebrae (Greenwood et al., 2015; Wallace et al., 2016). These limitations could be overcome by fusing diagnostic MR images with the intra-interventional scans in order to benefit from the typical high soft tissue contrast of MRI during the interventions. Furthermore, additional image information generated during the therapy planning phase, e.g. segmented metastases, predicted ablation zones or optimised RFA applicator trajectories (Baegert et al., 2007; Merten et al., 2019; Seitel et al., 2011), can be visualised in the interventional images and make metastasis puncturing faster and more accurate while less cognitively demanding for the radiologist (see Figure 3.1).

3.1.2 *State of the Art*

Multimodal image registration is well studied in the field of medical imaging. CT and MRI are essential diagnostic procedures and a fusion of them yields hybrid images combining the advantages of each imaging

technique, which enhances the informative content and can support the interpretation of conspicuous structures. Registration methods are commonly categorised regarding the dimensionality, the transformation type and whether they are driven by intrinsic or extrinsic information. Especially registration techniques that base on intrinsic image information, i.e. anatomical landmark-based or intensity-based (voxel-based) approaches became prevalent in recent years (Viergever et al., 2016). They offer the great advantage of fusing multimodal images, where extrinsic information like matching scanner coordinate systems or external tracking systems are not existent. Accordingly, voxel-based image registration of multimodal pre- and intra-operative images are suitable to support interventions with changing patient positioning and percutaneous instrument navigation.

Although various registration approaches have been investigated, only few publications are available regarding musculoskeletal surgery and spinal tumour interventions. The most common limitations in fusing interventional CT and spinal MR images are differences in patient positioning causing intervertebral joint movements and the overall lower image quality of interventional imaging. Since medical image registration is a wide-ranging field with numerous methods and applications that differ fundamentally or only in details, the following analysis of the state of the art will particularly focus on registration approaches for MR/CT spine imaging.

To start with, Kaminsky et al. (2004) introduced a multisegmental image fusion approach, applied to cervicothoracic spine MRI and spiral CT images of one patient. Nine corresponding reference points were placed on each vertebra in both image volumes. Subsequently, each vertebra had to be segmented within the CT image volume for the following multisegmental registration, using thresholding and manually post-correction. The centers of both point clouds were translationally registered and conclusively, the CT volume was rotated in accordance with Procrustes superimposition (Gower, 1975). Fusion time referred to the landmark positioning and took approximately 8 min, although the authors did not state any time needed for the required preceding segmentation procedure. The registration accuracy, defined as the mean fiducial registration error (FRE), was 1.53 mm. While overcoming the limitations of intervertebral joint movements in spine image fusion via their multisegmental approach, it required substantial manual user input to define the corresponding landmark pairs.

Sohn et al. (2009) presented a multisegmental image fusion approach for improved targeting of spinal tumours in image-guided stereotactic radiosurgery. Pre-operative MRI and CT scans of 20 patients with spinal tumours and metastases were acquired and registered via manually defined reference points and delineation of individual vertebral bodies to ensure multisegmental registration. Therefore, similar to Kaminsky et al. (2004), discrepancies caused by patient positioning were widely avoided by multiple rigid registrations of individual vertebrae or spine segments. Sohn et al. (2009) evaluated their registration approach via manual inspection and found it successfully, though not stating any

quantitative registration accuracy measures or the required fusion time. Nevertheless, they assessed discrepancies of the gross tumour volume (GTV), previously segmented in both image modalities and the fused images. Mean relative discrepancies between the GTV from the pre-operative datasets and the fused images were 30.5 % (CT-based GTV) and 14.5 % (MRI-based GTV).

Karlo et al. (2010) presented a feasibility study on MR/CT image fusion of the lumbar spine after spondylosis. Two readers, a radiologist and a spine surgeon registered ten clinical datasets, five with correct and five with incorrect implant positioning, using the OsiriX software (Pixmeo SARL, Switzerland) and applying a landmark-based registration. Fusion time and accuracy were assessed, the latter defined as the largest deviation between MR and CT images at the ventral corner of the cranial endplate of the most cranial vertebra in a midsagittal plane. Karlo et al. (2010) performed two manual registration experiments with different numbers of used landmarks. In session 1 each vertebra was marked with a single reference point in a midsagittal cross-section of the T_2-weighted MRI sequence and the CT image. In contrast, two reference points per vertebra were placed in session 2. The average fusion time was 102 s (first session) and 107 s (second session), the mean discrepancies were 1.02 mm (first session) and 1.91 mm (second session), respectively.

Miles et al. (2013) fused CT/MR spine images in order to allow physicians to visually explore corresponding soft tissue and bony structures in a single dataset. The study included 20 patients who underwent MRI and CT scans of the lumbar spine. Both volumes were registered by a versor-based transformation and mutual information (MI) using the Insight Toolkit, Kitware Inc, USA. For this purpose, the CT images were thresholded to enhance soft tissue details and the user initialised the approach with two corresponding points in each volume. Miles et al. (2013) evaluated their method by the use of 17 anatomical landmarks per image pair and achieved a mean target registration error of 1.9 mm. In contrast to the above-mentioned studies, Miles et al. (2013) presented a spine registration method based on voxel-wise image information and not solely on corresponding landmark pairs and thus, reduced the user initialisation effort. However, they did not state any required computational time and applied a globally rigid transformation, which is less critical for registrations of the rather inflexible lumbar spine, but can considerably affect cervicothoracic spine segments.

After the development and subsequent publication (Hille et al., 2018a) of the content covered in this section, Rashad et al. (2019) presented a hybrid registration method of cervicothoracic spinal MRI and CT images, where locally rigid transformations of single vertebrae were embedded into a global deformation field to fuse both image volumes. For this purpose, the authors used a novel commercial software by Brainlab AG, Germany, which started with an automatic vertebrae segmentation in the CT scans. Subsequently, ROIs were defined based on the preceding segmentations and an initial prealignment was performed by the user, which served as a starting point for the automatic rigid registration of the

first vertebra. The ROIs of adjacent vertebrae were iteratively registered using MI as an image similarity metric, while previous transformations were taken into account to constrain translational and rotational steps. Afterwards, the resulting multiple rigid registrations were embedded into a gloabl deformation field, which has been smoothed between the segmented and rigidly registered vertebrae. Rashad et al. (2019) achieved an FRE of 1.54 mm averaged over 10 patients with multiple MRI sequences each. However, their evaluation procedure casts doubt on the validity of their results, since they transferred landmarks manually defined within the CT scans to the MRI volumes by automatic vertebra-specific rigid registrations using MI, instead of separately defining them in both image volumes. Therefore, Rashad et al. (2019) used the same approach to define corresponding landmark pairs for evaluation purposes, as their proposed method, which generated the results that should thus be evaluated. Hence, the stated results should be considered with reservations. Nonetheless, Rashad et al. (2019) showed the importance of a multisegmental registration strategy, particularly for the cervicothoracic spine, since they additionally fused CT and MRI data of each patient using globally rigid registrations, which led to significantly higher FREs (7.12 mm vs. 1.54 mm).

Most recently, Zhang et al. (2020) presented a case series study regarding the intra-operative localisation of spinal cord lesions. Pre-treatment spinal MRI scans of 22 patient cases were fused with O-arm CT images during interventions using the commercially available software Synergy Cranial from Medtronic, Ireland. However, their publication does not state any information w.r.t. the used registration method and whether it is based on a manual or automatic approach. They evaluated the accuracy of each patient case by a single distance measurement of the maximum discrepancy between the tumour surfaces in both images after fusion. Using this measure, Zhang et al. (2020) stated mean errors of 1.6 ± 0.9 mm and an average computation time of approximately 15 min.

Even though, there were few studies approaching multimodal spine image registration, the clinical applicability of the proposed solutions was limited either by considerably high user interaction and time needed for manual landmark-based registrations, since globally rigid approaches could not sufficiently cope with intervertebral joint movements, or because the registration accuracy could not be assessed due to a lack of information. This led to the conclusion that there remained a need for a precise, fast and automatic fusion of MR and interventional FP-CT.

3.1.3 *Objectives*

Considering an image fusion of pre- and intra-operatively acquired images during spine lesion RFAs, the following objectives shall be met:

- Computational time per patient case on average < 5 min on current consumer hardware
- Registration error (FRE) on average < 3 mm

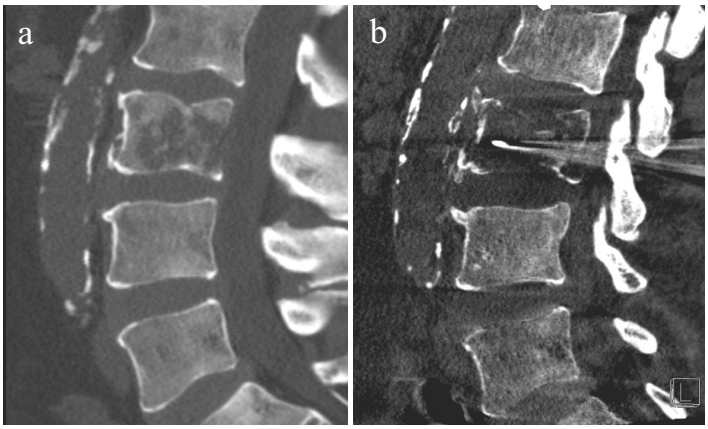

Figure 3.2: Comparison of the image quality between conventional CT (a) and intra-interventionally acquired FP-CT (Dyna-CT) images (b) of the lumbar spine. Overall, weaker image contrast due to low-dose protocols and artifacts like beam hardening at cortical bone structures or metal artifacts from inserted applicators aggravate precise metastasis localisation and puncture.

For efficient and convenient applicability within the interventional workflow, the total procedure should not exceed 5 min and should require only minimal user interaction, in order to be applicable between the calibration of the navigation system and the beginning of the metastasis puncture during the intervention. Most of the relevant state of the art works used landmark-based registration approaches, which is - w.r.t. the above defined objectives - too time-consuming due to the necessity of precisely marking multiple corresponding reference points in both image volumes.

A sufficient level of registration accuracy is generally not known and heavily dependent on the specific task w.r.t. the field of application, the spatial image resolution or the displayed anatomical structures. Additionally, the evaluation of registration results is cumbersome and prone to errors. For instance, in case of calculating accuracies using sets of corresponding landmark pairs, which is a common practice, the user introduces a permanent error within the ground truth due to inaccurate fiducial localisation (fiducial localisation error (FLE)). Considering the application addressed in this section, the mean FRE should be less than 3 mm for being sufficiently precise, which was based on the mean diameter of the vertebral pedicles, ranging from thoracic 3 mm to lumbar 10 mm (Charles et al., 2015; Liu et al., 2010; Scoles et al., 1988). The reason for this was that for most patient cases the RF applicators are inserted via the pedicles into the vertebral body, which is both the narrowest, as well as the pathway with the closest proximity to the spinal cord as a risk structure.

Furthermore, an important aspect of this registration task was the application to intra-operative low-dose FP-CT images, which commonly lack image quality and are more prone to artifacts (see Figure 3.2). Therefore, the intended method should combine a multisegmental and voxel-based

registration approach, to cope with patient positioning-induced spine deformations and to overcome limitations of reduced structural information and the increased presence of image artifacts due to low-dose interventional imaging.

3.1.4 *Materials and Methods*

Image Data

19 patients who underwent RFAs of both, single or multiple vertebral metastases at the Department of Neuroradiology of the University Hospital of Magdeburg, were chosen retrospectively. For evaluation purposes, sagittal native T_1-weighted or contrast-enhanced T_1-weighted MRI sequences and intra-interventionally acquired FP-CT scans (Dyna-CT) of each patient were used. Additionally, five patients were randomly chosen to test the influence of T_2-weighted sequences on the presented registration approach. The in-plane image resolution of the MRI data ranged from 0.47 mm to 1.25 mm (on average 0.63 mm) and the slice spacing was 3.30 mm for all scans. The FP-CT scan resolution ranged from 0.22 mm to 1.10 mm (on average 0.65 mm) in-plane and had a slice spacing ranging from 0.46 mm to 3.00 mm (on average 1.28 mm).

Image Fusion

The presented registration approach was selected due to both the physical characteristics of the spine and the available multimodal images. In the main case of application, in which most patients were in advanced tumour stages and had several vertebral metastases, the interventional target region was not limited to a single vertebra, but partially covered entire spinal segments. A multisegmental, i.e. piecewise rigid registration procedure appeared to be the most suitable approach in order to accurately model the deformation of spine structures, caused by different patient positioning. Therefore, a globally non-rigid image fusion problem was split into multiple local rigid registrations of individual vertebrae or spine segments. A globally deformable registration approach with incorporated regions of local rigidity, as presented by König et al. (2016), was disregarded due to the required time-consuming delineations of rigid structures in both image volumes. Figure 3.3 illustrates the proposed method and its specifications. First, the user had to roughly mark each vertebra or spine segment which should be registered in both modalities as an initialisation. Following this, regions cropped to single vertebrae or segments were transformed so that their centers coincided in the coordinate origin, taking into account the patient orientation and voxel spacing specified in the DICOM tags. This led to a coarse initial image registration. The anteroposterior length l of those regions was set to 10 cm, the laterolateral width corresponded to the MRI volume. The craniocaudal height h was defined depending on the distance between each adjacent marker. Each of the ROIs thus defined was aligned parallel to the verte-

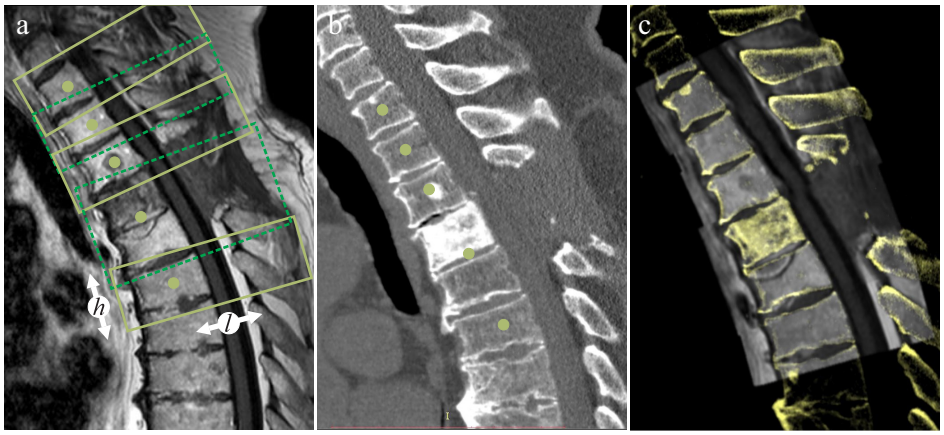

Figure 3.3: For initialisation purposes, the user had to roughly mark the centers (green points) of each region which had to be registered in both imaging modalities (a and b). Subsequently, 3D-ROIs (green contours) were automatically defined and coarsely translated onto the interventional FP-CT image. The proposed multisegmental registration approach then precisely and piecewisely registers each ROI and embeds it in a fused image (c).

bral endplates by rotating it by the orthogonal angle of the connecting line of two adjacent marker points.

Subsequently, a three-level multiresolution and voxel-based rigid registration approach precisely registers each ROI with the interventional image combining normalised gradient fields (NGF) (Haber and Modersitzki, 2006) as the image similarity measure and a Quasi-Newton optimizer, which proved to be the most promising combination in a preliminary study regarding accuracy and computational time. Starting with a rather coarse image resolution, the transformation was subsequently refined on images of increasing resolution until full resolution was reached (downsampling factor was 0.63). NGF are based on a pointwise (continuous) or voxel-based (discrete) distance measure D of the angle between two image gradients, which is defined as

$$\mathcal{D}(T,R) = \frac{1}{2}\int_{\Omega} d(T,R)dx; \quad d(T,R) = \|n(R,x) \times n(T,x)\|^2 \tag{3.1}$$

with two corresponding spatial coordinates x in reference image R and template image T, as well as their related gradient vectors $n(R,x)$ and $n(T,x)$ in the image domain Ω (Haber and Modersitzki, 2006). Both vectors form an angle $\theta(x)$ and since the gradient fields are normalised, the cross product of both vectors is related to the sine of $\theta(x)$. To find a reasonable image registration, the objective function

$$\mathcal{D}(R,T(t)) := \mathcal{D} \to \min \tag{3.2}$$

has to be minimised, for instance by minimising the square of the sine of $\theta(x)$. With respect to the problem of differentiability if image areas are constant, a regularisation term is introduced with

$$n_\epsilon(I, x) := \frac{\nabla I(x)}{\sqrt{\nabla I(x)^T \nabla I(x) + \epsilon^2}} \tag{3.3}$$

The parameter ϵ can be used to regulate the effect of regions with weak gradients $\nabla I(x)$ and can be automatically determined with

$$\epsilon = \frac{\eta}{V} \int_\Omega |\nabla I(x)| dx \tag{3.4}$$

with η as a noise level estimation of the image and V as the volume of the domain Ω (Haber and Modersitzki, 2006). Since NGF are computed by solely taking derivatives into account, it is well suited for multimodal image registration, particularly for intra-operative imaging due to commonly weak image contrasts and high noise levels. The transformed image $T(t) := T \circ t$ is mapped onto the reference image R by a rigid transformation t. Therefore, each transformation TR resulted from the initial translation I_{tr}, multiplied by the image-based registration t and thus $TR = t \cdot I_{tr}$.
Fusing all transformed and labelled ROIs in a joint image was the last step in the proposed procedure. Image intensities of voxels x with more than one label, e.g. within the area of two overlapping ROIs, were weighted depending on their distance to the specific ROI centers:

$$i_x = \sum_{r=1}^{2} \frac{i_{xr}}{1 + e^{-20(||d_{xr}|| - 0.5)}} \tag{3.5}$$

where i_x was the resulting voxel intensity, determined by a distance-based sigmoid function and the intensities i_{xr} of each voxel x of two overlapping regions r. d_{xr} represented the relative normalised distances between such voxels and the nearest two centers of the transformed ROIs with $\sum d_{xk} = 1$. An otherwise linear weighting resulted in a largely blurred and fuzzy display of the anatomical structures.

Evaluation

For each patient, ground truth landmarks were given in both, the diagnostic MRI and the intra-interventionally acquired FP-CT image volumes. Three landmark pairs were determined per vertebra within the sagittal plane of symmetry and on the vertebral rim (superior-anterior, superior-posterior and inferior-posterior corner). The number of labelled vertebrae per patient depended either on the field of view (FOV) of the FP-CT

images or on previously defined target regions. The ground truth data of all patients was prepared by a field expert trained by neuroradiologists. To evaluate the registration quality, the mean FRE via Euclidean distances between N pairs of corresponding landmark pairs (transformed MRI landmark $T(p)$ and the FP-CT landmark q) was determined as

$$\mathrm{FRE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T(p_i) - q_i)^2} \tag{3.6}$$

This was done for the multisegmental approach (FRE_{MS}), as well as for a globally rigid registration (FRE_{GR}) with the same optimizer and metric settings, to verify the need for a multirigid approach to meet clinical accuracy requirements for this particular task. In order to compare the automatic and voxel-based strategy with a manually performed registration, the landmark sets of both image volumes per patient case were directly registered via Horn's quaternion-based method (Horn, 1987). Hence, the resulting minimal fiducial registration error (mFRE) represented the accuracy of an optimal global registration using landmark sets and therefore, indicated the challenge of localising corresponding landmarks in two image volumes. The results reported in the following were averaged over all landmark pairs per patient and subsequently over all patients cases.

Furthermore, to check the inter-reader variability of the ground truth, two additional field experts provided landmark sets in both the MRI and the FP-CT images for a subset of five randomly chosen patients out of the original 19 cases. Additionally, the first reader created a second ground truth of the same five patients 24 h after the first ground truth preparation to check the intra-reader variability. Both FLEs were used to demonstrate that the measured registration accuracy considerably depends on the individual users and their fiducial localisation accuracy, as well as the challenge of a consistent identification of anatomical landmarks. The impact of these introduced inaccuracies within a ground truth are a rarely addressed issue when it comes to evaluating registration approaches (Viergever et al., 2016).

3.1.5 *Results*

As mentioned above, IRV was checked w.r.t. the image modality and resulted in mean FLEs of 1.36 mm for MRI and 1.28 mm for FP-CT scans. The intra-reader variability was on average 1.05 mm for MRI and 1.03 mm for FP-CT images. Manual, i.e. landmark-based registrations of each patient case resulted in a mean mFRE of 1.70 mm. The required time to localise corresponding landmark pairs in both image volumes ranged between 2 and 3 min per vertebra.

Voxel-based and multisegmental registrations of individual vertebrae or spine sections resulted in average FRE_{MS} of 2.35 mm (T_1-weighted MRI) and 2.55 mm (T_2-weighted MRI). Globally rigid registrations of both

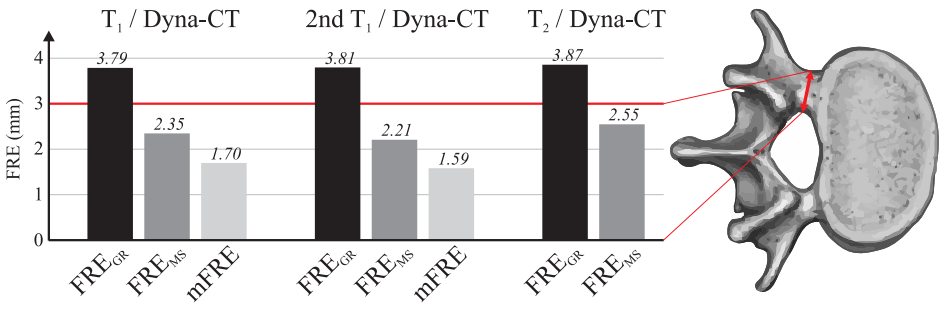

Figure 3.4: Resulting average registration accuracies of both, the globally rigid image registration (FRE_{GR}) and the presented multisegmental approach (FRE_{MS}). '2nd T_1' refers to the additionally produced ground truth of the first reader. The red line at 3 mm represents the clinical accuracy objective derived from the pedicle diameter.

Table 3.1: Intra- and inter-reader variability was determined for five randomly chosen patient cases.

	MRI	FP-CT
FLE_{intra} [mm]	1.05	1.03
FLE_{inter} [mm]	1.36	1.28

image datasets resulted in mean FRE_{GR} of 3.79 mm (using T_1-weighted MRI data) and 3.87 mm (using T_2-weighted MRI data). The computational time per vertebra was on average 24 s and the overall required time per patient, including initialisation, did not exceed 5 min for any patient case on consumer hardware.

3.1.6 *Discussion*

Image fusion has been successfully applied to radiation therapy for the purpose of delineation and enhancement of target fields (Dalah et al., 2008) or to support image-guided interventions (Mauri et al., 2015; Nemec et al., 2007; Sohn et al., 2009). However, in musculoskeletal radiology, especially in neuroradiology of the spine, image fusion applications have rarely been reported. Image fusion of diagnostic MRI and interventional FP-CT scans can efficiently support image-guided interventions of spinal metastases due to the valuable gain in image information. MRI is considered the method of choice regarding tumour and metastasis delineation as well as for assessing compressions of spinal nerve roots and the spinal cord, due to its emphasised soft tissue contrast. However, RFAs are performed under FP-CT image guidance, which besides a pronounced image contrast of cortical bone structures withholds relevant image information regarding soft tissues like metastases. Therefore, fusion of both modalities can provide various additional information to support applicator guidance and needle tip placement beyond the established methods of

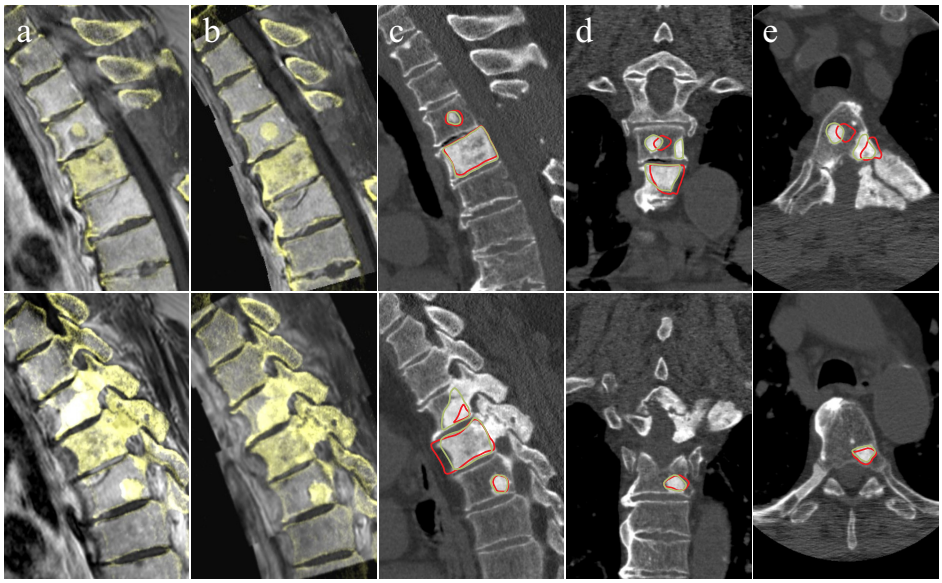

Figure 3.5: Comparison of fused images using a globally rigid (a) and a multisegmental (b) registration approach. Shown are a mid-sagittal (top row) and lateral (bottom row) cross-section of an exemplary patient case. Additionally, the masks of previously segmented metastases were transformed according to the resulting transformation matrices of the globally rigid (red contours) and the proposed multirigid approach (green contours). Sagittal (c), coronal (d) and axial (e) cross-sections of the FP-CT scan are shown with overlayed metastasis contours. The FRE_{GR} of this patient case was 4.58 mm and the FRE_{MS} was 1.84 mm.

current navigation systems (Kavakebi et al., 2017; Wallach et al., 2014). Considering this, the proposed multisegmental voxel-based (MS-VB) registration approach represents a key aspect of this thesis and the intended intervention support.

Table 3.2 displays an overview of relevant information about the presented and related state of the art works, although some of the latter did not state any quantitative results, which restricts comparability. In contrast to the related work of spinal image fusion, the presented approach focused on interventional rather than native CT imaging, which meant qualitatively inferior images due to low-dose protocols. So far, either landmark-based approaches (Kaminsky et al., 2004; Karlo et al., 2010; Sohn et al., 2009) or MI as a registration metric (Čech et al., 2006; Hu and Haynor, 2004; Miles et al., 2013) have been used. The former usually have the disadvantage of a time-consuming initialisation procedure, which grows proportionally with the number of used landmarks. Automatised landmark detection approaches, e.g. by edge detection, come at the cost of solving the complex correspondence problem between two sets of landmarks. In contrast to the mean computational time of 24 s per segment of the proposed MS-VB approach, the study of Zhang et al. (2020) for instance, required on average 15 min, while Hu and Haynor (2004) reported an overall required time of approximately 60 min per patient case, which is hardly compatible with any clinical procedures. Additionally, the registration accuracy of manual procedures is highly dependent on the

user's precision and care, whereas especially a high workload and lack of time in clinical routine could negatively affect the results. Likewise, a preceding vertebrae segmentation to define the regions to be registered, as presented by Čech et al. (2006) or Hu and Haynor (2004), is yet more time-consuming, if manually performed or if automatically performed requires similar user initialisation as the proposed MS-VB approach (Chu et al., 2015; Rak et al., 2019; Zukić et al., 2014). The presented MS-VB method, however, required only an approximate marking of the vertebrae to be registered, which reduced the required time for an initialisation as well as the cognitive load of radiologists and could therefore, be more easily integrated into the clinical workflow. However, this manual initialisation can be replaced in the future by automatic vertebra and spinal metastases detection procedures like presented by Chen et al. (2015), Rak and Tönnies (2016), and Wang et al. (2017). Yet it remains to be seen, how robust vertebra detection in the immediate vicinity of metastases performs. After the implementation and publication of the proposed MS-VB approach, Rashad et al. (2019) presented a largely similar registration strategy using a new commercially available software. They followed the pipeline of vertebra-specific multirigid registrations and the subsequent embedding in fused images. As already mentioned in Section 3.1.2, their results should be considered with reservations, since their evaluation strategy was unsuitable to unbiasedly assess the registration accuracy of their proposed method. In addition, their registration accuracy was highly dependent on the initial prealignment, which partially led to FREs over 2 cm, if the outmost segmented vertebrae were used as starting points. However, the findings of Rashad et al. (2019) indicated the importance of a multisegmental approach to cope with spine deformations (mean FRE_{MS} of 1.54 mm vs. mean FRE_{GR} of 7.12 mm) and the impact of the spatial resolution on the registration accuracy (mean FRE_{MS} of 0.91 mm with high resolution images vs. mean FRE_{MS} of 1.77 mm with low resolution images).

Similar to the findings of Rashad et al. (2019) regarding the impact of spine deformations, the average FRE_{MS} with 2.35 mm achieved with the proposed multisegmental approach was significantly lower than the average FRE_{GR} of 3.82 mm using a global transformation (see Figure 3.5). The latter would not have met the clinical objectives. This was also evident considering the individual patient cases (see Table A.4), with 9 out of the total 19 patient cases not meeting the accuracy criteria if registered globally. Applying the MS-VB approach in contrast, resulted only in two cases with an insufficient accuracy (one of them achieved 3.12 mm instead of the required 3 mm). The worst patient case resulted in an FRE_{MS} of 4.43 mm averaged over 9 pairs of corresponding landmarks, although the mFRE, i.e. the mean error of a manual registration of this case, was among the largest with 2.37 mm (1.70 mm overall patient cases). Accordingly, the actual image registration could have been sufficiently precise, but the considerably high mFRE indicated difficulties in accurately identifying corresponding landmarks and thus also had a negative impact on the measured registration accuracy of the proposed approach. In general, the

Table 3.2: Related work and the presented MS-VB approach in comparison w.r.t. the used image modalities (MRI, CT, FP-CT), the number of patient cases N_P, the usage of interventional imaging data (I_{Int}), the registration metric (MI - mutual information, NGF - normalised gradient fields, LMB - landmark-based), the used transformation type, as well as the achieved average fiducial registration error FRE and the required computational time per dataset t (t* - average time required per vertebra). The results of Rashad et al. (2019) were questionable w.r.t. to their evaluation strategy.

Works	Image Data	N_P	I_{Int}	Metric	Transformation	FRE [mm]	t
Hu and Haynor, 2004	MRI / CT	1	-	MI	multirigid	n.s.	>60 min
Kaminsky et al., 2004	MRI / CT	1	-	LMB	multirigid	1.53	>8 min*
Čech et al., 2006	MRI / CT	3	-	MI	multirigid	n.s.	n.s.
Sohn et al., 2009	MRI / CT	20	-	LMB	globally rigid	n.s.	n.s.
Karlo et al., 2010	MRI / CT	10	-	LMB	globally rigid	1.46	~2 min
Miles et al., 2013	MRI / CT	20	-	MI	globally rigid	1.90	n.s.
Rashad et al., 2019	MRI / CT	10	-	MI	multirigid	(1.54)	n.s.
Zhang et al., 2020	MRI / CT	22	-	n.s.	globally rigid	1.6	~15 min
MS-VB	MRI/FP-CT	19	+	NGF	multirigid	2.35	24 s*

magnitude of the mFRE can be explained in particular by the voxel size of the MRI scans, which had a diagonal of approximately 3.5 mm. Besides the spatial resolution component, it is not trivial to identify corresponding anatomical landmarks in images of different modalities, since the representation and display of the same tissue types differ considerably. Hence, the mean mFRE of 1.70 mm demonstrated the challenging task of landmark localisation in two image volumes even for field experts.

Furthermore, intra- and inter-reader variability regarding the FLE were investigated. As expected, average intra-reader variability was significantly lower than the inter-reader variability (1.04 mm vs. 1.32 mm), since various readers subjectively interpret the given landmark positioning guidelines slightly different. Even though, the anatomical landmark positions were clearly defined in theory, inaccuracies occurred, because, e.g. the vertebral rim was several voxels wide and the sagittal plane of symmetry was partially difficult to determine objectively and reproducibly. For instance, the choice of the respective sagittal cross-section could already lead to significant discrepancies between different readers. The FLEs of landmarks placed in the MR images were slightly higher than those in the interventional FP-CT, which was due to the lower spatial resolution of the MRI. With 0.8 ± 0.4 mm, a similar FLE of the intra-reader variability in MRI (spatial resolution of $0.3 \times 0.3 \times 3$ mm) was found by Miles et al. (2013).

Considering the used MRI sequences, slightly higher FREs have been observed when using T_2- instead of T_1-weighted MR images. This could partly be attributed to the somewhat better contrast of the T_1-weighted images at the transition between the dorsal vertebral rim and the spinal canal, while the image intensities of both, vertebral bone marrow and cerebrospinal fluid were roughly similar in the T_2-weighted MR images.

In terms of the used registration metric, NGF proved superior to MI (average FRE_{MS} of 2.35 mm vs. 2.87 mm). This could most likely be attributed to the overall reduced tissue contrast and SNR of low-dose interventional FP-CT imaging. In most FP-CT scans only the cortical bone of the vertebra emitted sufficient imaging signals and high image contrast, while surrounding tissues or the cancellous bone was hardly visible and thus contributed scarcely to the MI criterion.

3.1.7 *Conclusion*

In this section, a precise registration approach was presented, which fused diagnostic MR and interventional FP-CT images for the purpose of intervention support. The proposed registration strategy combined a voxel-based and multisegmental approach and took interventional image data into account. The choice of the image fusion method was determined by both physical characteristics of the spine as well as the requirements of the workflow during RFA interventions of spinal metastases. It could be shown, that a multisegmental approach with NGF as an image similarity measure could cope with spine deformations due to patient positioning and avoid the time-consuming initialisation of landmark-based

registration approaches. The sufficient registration accuracy and short computational times enabled the presented method to provide practical and applicable intervention support without significantly delaying the clinical workflow or inducing too much additional workload.

4

POST-INTERVENTIONAL PHASE

Subsequent to the intervention, follow-up MRI scans are performed to assess the treatment outcome. Primary goal of most interventions is to reduce pressure on the spinal cord and peripheral nerves, as well as on the periosteum of the vertebrae, which is highly sensitive to pressure pain (Guillevin et al., 2007; Klimo and Schmidt, 2004). This results in a relief from severe pain, neurological deficits, and paralysis in particularly advanced stages. Apart from these overall qualitative assessments of the treatment outcome, it is a crucial step forward to an enhanced and computer-assisted therapeutic workflow to produce quantitative and reproducible evidence. For instance, both, the ablated and the residual tumour volume (partial remission) in relation to the pre-operative metastases, as well as the remaining safety margin of the ablation zone to nearby risk structures are key parameters for the evaluation of the therapy outcome. Based on the combination of such quantitative and qualitative assessments, a more accurate and reliable prediction about recurrent and progressive metastases can be made, and thus whether further interventions or adjuvant therapies are indicated. The base for such quantitative validations is a segmentation of the necrosis zones and an image fusion of follow-up MRI scans with the pre-operative images, in order to superimpose necrosis zones and former metastases. Both issues are addressed in the following sections and based on Steffen and Hille et al. (2020) and Hille et al. (2019).

4.1 NECROSIS ZONE SEGMENTATION

4.1.1 *Introduction*

Follow-up MRI scans are used to qualitatively assess the treatment outcome subsequent to the RFA intervention. Conventionally, this assessment is carried out by side-by-side comparison of diagnostic and follow-up imaging in order to compare geometric properties like size and shape of both the former metastases and the resulting necrosis zones. This, however, is challenging due to the difficulties of correlating spatial positions in separate image volumes or estimating the correct spatial extension in three-dimensional space, as well as the considerably high interaction effort, as the user has to slice through each volume separately. Hence, the conventionally required mental image fusion is highly demanding, time-consuming, prone to human error, and hardly lead to a reliable and robust assessment of the treatment outcome. Since tumour recurrence due to a partial remission after RFAs can be diagnosed comparatively early by size and geometry alterations observed in follow-up scans (Egger et al., 2015), a precise volumetrisation of the necrosis zone is of crucial

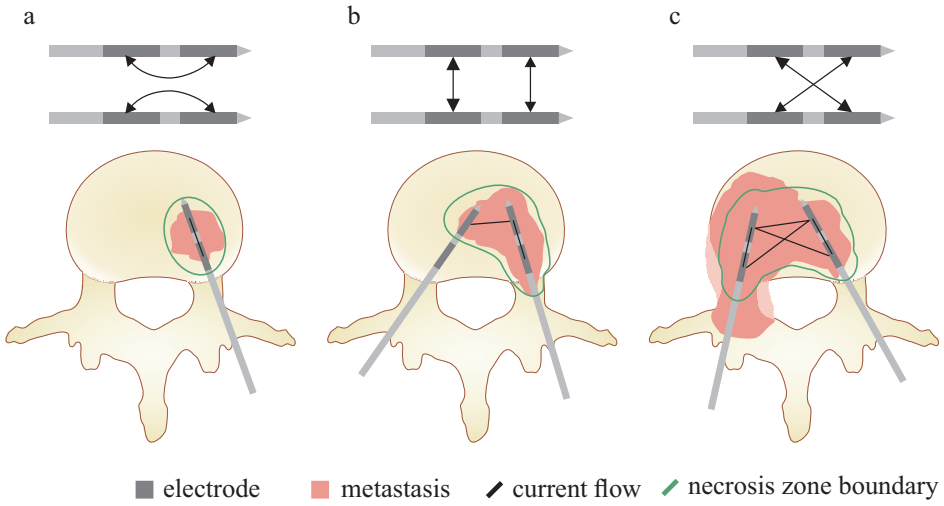

Figure 4.1: Basic configurations of the current flow between active pairs of electrodes at the applicator needle tips w.r.t. mono- or multipolar ablation concepts (top row). While monopolar RFAs result in ellipsoidal shapes (a), multiple applicators could shape more complex geometries (b and c) due to timed switching between the electrode configurations (bottom row).

importance. Additionally, it constitutes the basis for any computation of quantitative validation measures such as tumour coverage, residual tumour volume or maintained safety margins towards adjacent organs at risk. Beyond the benefit of the individual therapy, such acquired data can yield additional insights in the context of retrospective studies. For instance, it can contribute to a better understanding of the causes of tumour recurrence or it can further improve ablation protocols and procedures by comparison with previously predicted ablation zones during the therapy planning phase. Nowadays, the manufacturers of RFA devices commonly state a prospective size of the ablation zones, which could be used as a good approximation, although the extent is generally rather overestimated (Kröger et al., 2010; Matschek et al., 2017). Accordingly, accurate segmentations of the resulting necrosis zones can provide beneficial feedback for the simulation and outcome prediction of ablative procedures.

Nevertheless, some specific aspects of necrosis zone segmentations tend to make this a highly challenging and ambitious task, e.g. hardly noticeable intensity differences between necrosis and remaining metastasis, inflammation processes caused by the coagulation or possible hematomas. Moreover, an overall high shape variability of the necrosis zones due to the former metastases shapes, the configuration of multipolar RFA, as well as the tissue-specific heat propagation of target and adjacent structures hampers automatic segmentation approaches. Regarding the shape of necrosis zones, there are two basic concepts: for monopolar RFAs only one applicator needle is inserted, resulting in rather small and ellipsoidal ablation zones, whereas for multipolar RFAs two or more needles are inserted with their tips placed within the tumour tissue (see Figure 4.1).

Besides increased ablation zones, the major advantage is the additional degree of freedom, since multiple configurations of active pairs of needle electrodes can shape more sophisticated ablation zones and allow a finer control of the induced temperatures (Matschek et al., 2017). This becomes even more relevant when considering that the heat transfer abilities differ notably between the cancellous and cortical bone of the vertebrae (Greenberg et al., 2014). Therefore, the intra-vertebral location of the metastases and thus the positioning of the RFA applicator needle tips significantly influences the ablation process and consequently the resulting shape of the necrosis zones. A segmentation approach will have to cope with these challenges to be suitable and applicable in clinical settings.

4.1.2 *State of the Art*

With the increasing number of ablative procedures on the spine (Starr et al., 2019, Transparency Market Research, 2019) the subject of a volumetric assessment of resulting necrosis zones is gaining relevance. Despite this, regardless of the application area, there currently are only few works addressing ablation zone segmentation subsequent to any ablation procedure. To the best of the author's knowledge, all published research works focused on liver tumour treatment in CT and none of them dealt with spinal necrosis zones or MRI. Therefore, the following state of the art analysis will focus on related approaches for ablative interventions of hepatic lesions. Segmentation approaches regarding metastatic lesions, as mentioned in Section 2.2.2 could be considered relevant in a broader sense, but will not be discussed here once more.

Bricault et al. (2006) developed a computer-aided analytical tool, which focused on assessing local recurrences of liver metastases after RFA treatment in CT. For this purpose, a semi-automatic 3D segmentation approach based on a watershed algorithm was implemented. On average, the segmentation took approx. 4 min, but accuracy results were not stated.

Keil et al. (2010) proposed a semi-automatic and user-initialised segmentation approach of liver lesions and subsequent ablation zones in CT imaging combining a 3D region growing approach with various morphological operations. Evaluating their approach with 49 patient cases resulted in a mean 99 % concordance correlation coefficient with a 95 % confidence interval. However, their approach included a manual correction possibility by iteratively adapting the region growing thresholds and no distinction was made between pre- or post-correction results w.r.t. the accuracy.

Weihusen et al. (2010) presented a workflow-oriented software support for CT image-guided RFA of focal liver malignancies. In their treatment outcome validation tool they also included a semi-automatic necrosis zone segmentation based on a morphological region growing algorithm, though Weihusen et al. (2010) stated no quantitative results whatsoever.

Passera et al. (2013) proposed a semi-automatic approach to segment hepatic lesions in diagnostic and ablation zones in post-interventional CT scans of the liver. The input of their segmentation approach consisted of

a binary mask of the liver generated by a preceding user-guided live-wire algorithm and up to three reference pixels within the lesion or necrosis marked by an operator. A fuzzy C-means clustering with additional morphological filtering subsequently segmented the target structures. Their approach was applied to pre- and post-interventional images of 10 patient cases and took on average approximately 10 min, while achieving a mean TPR of approximately 96 %. Both segmentations were part of a processing pipeline to quantitatively assess tumour coverage after RFAs of liver lesions, whereby the following Section 4.2.2 will focus in particular on strategies for treatment outcome validation.

The currently most advanced method was proposed by Egger et al. (2015), whose semi-automatic graph cut-based approach was used to segment ablation zones in hepatic CT imaging. They included an interactive refinement option by manually dragging the seed point around and generating new segmentations based on the current set of graph nodes. Egger et al. (2015) achieved a mean accuracy of 77 % DSC on their evaluation set of 12 patients and reported computational times of a few seconds.

In conclusion, the previous studies focused solely on post-interventional CT imaging of liver tumour ablations, which is hardly transferable to necrosis zone segmentation after ablative interventions of spinal metastases in MRI. Primarily differences in shape and appearance of the ablation zones due to the different heat propagation processes among soft tissues and osseous structures, as well as the variety of adjacent tissues in the complex geometry of the spinal region distinguish hepatic and spinal necrosis zone segmentation. Furthermore, the different imaging techniques, i.e. CT imaging in most state of the art works and MRI in this specific use case, as well as the so far predominantly manual procedures in literature, required a fundamentally new approach towards this ambitious and challenging segmentation task.

4.1.3 *Objectives*

Analogous to the segmentation task described in Section 2.2, the following objectives were defined in cooperation with the clinical partner:

- Computational time per necrosis zone on average < 1 min on current consumer hardware
- Segmentation accuracy in the range of the inter-reader variability of field experts

Since there are no directly related works regarding an automatic and fast segmentation of necrosis zones after RFAs of spinal metastases, a novel approach had to be developed to meet the defined clinical objectives. First of all, a thorough problem analysis has been carried out in order to find a principally suitable segmentation strategy from the existing state of the art of medical image segmentation. Subsequently, the most promising approach had to be adapted to the specific requirements of this task.

Likewise to most segmentation tasks, the time required plays a decisive role, since a reasonable support must reduce the radiologist's workload and occupy as little time as possible. Preliminary studies and the preparation of the ground truth have shown that manually performed contouring of a single necrosis zone took approximately 5-10 min. In order to considerably speed up this task and to ensure that even in the case of a necessary manual post-correction significantly less time is required, the objective of a maximum computational time of one minute per necrosis zone was defined.

Besides, a reliable and feasible segmentation approach must provide a sufficient accuracy to be considered applicable for clinical purposes. Accordingly, the resulting average segmentation accuracy must be within the range of the inter-reader variability of this particular or related tasks, e.g. spinal metastases segmentation, and thus reflect expert-like performance.

The conceptual similarity of this task and its requirements to the segmentation of spinal metastases as presented in Section 2.2.1 w.r.t. the image modality, the region of interest as well as the shape and appearance variability of the target structures suggested the implementation of a similar segmentation approach. Hence, a state of the art deep learning-based approach seemed to be suitable and most promising.

4.1.4 *Materials and Methods*

Image Data

In order to evaluate the proposed method a dataset was assembled, consisting of 26 follow-up MR images of patients who underwent RFAs of spinal metastases at the Department of Neuroradiology of the University Hospital of Magdeburg. These images were acquired between one and three days after the intervention including, among others, sagittal contrast-enhanced T_1-weighted and native T_2-weighted MRI sequences. Both sequences were chosen since they are most commonly used for visual examination regarding treatment outcome validation by neuroradiologists in this particular case. The image data differed w.r.t. specific acquisition parameters and scanner configuration. The scan resolution within a slice ranged between the individual patient cases and imaging sequences from 0.45 mm to 1.25 mm, the spacing between adjacent slices ranged from 3.3 mm to 4.8 mm. The image volumes of each patient case were pre-processed by a cubic interpolation between the original number of slices (between 15 and 25) to a fixed number of 64 to yield almost isotropic spatial resolution and simplify any following processing steps. An experienced neuroradiologists manually contoured each necrosis slice-wise. Thus, the input data could be used for training purposes as individual slices or as patient-wise merged volumes.

Data Augmentation

Due to the relatively small amount of available data, each of the 26 original MRI volumes were extensively augmented using the following techniques:

- *Gaussian blur:* The images were blurred with a Gaussian filter with σ in the range from 0 to 0.5.
- *Gamma transformation:* Gamma transformations with γ in the range from 0.5 to 2 were applied to modify image intensities.
- *Mirroring:* Each patient volume was flipped in all directions. This included vertical flips, i.e. craniocaudal. Even though it may appear inappropriate, preliminary studies have proven it to be advantageous for the final results due to the avoidance of fast overfitting.
- *Scaling:* The image volumes were scaled with randomly chosen factors between 0.6 and 1.4.
- *Rotation:* Rotations were applied to the image volumes in the range of $\pm$ 30 $^\circ$ around the transversal axis and between $\pm$ 20 $^\circ$ around the sagittal axis.
- *Elastic deformations:* Random displacement fields with subsequent Gaussian smoothing the grid with a σ ranging between 0 and 0.3 were used to elastically deform the image volumes (cf. Ronneberger et al., 2015).
- *Random Cropping:* Finally, each patient volume was translated in a random cropping manner within a range of $\pm$ 20 voxels in sagittal and vertical direction w.r.t. the centers of the necroses m_c and subsequently cropped to patches of a fixed size of 128 $\times$ 128 $\times$ 64 voxels.

After the augmentation each image volume patch was whitened by mean subtraction and a subsequent division by the standard deviation. It was ensured that each patch contained at least fractions of necrotised tissue. For the purpose of a stratified cross-validation, the patient data was grouped into six folds with a 21/5 (training/validation) split for two folds and a 22/4 split for the remaining four. Each patient volume within the training set was augmented 150 times, yielding in total 3,150 volumetric and 201,600 cross-section training samples, respectively for both 21/5-split folds, and 3,300 and 211,200, respectively for the remaining four folds.

CNN Architecture

Likewise to the similar data augmentation strategy implemented for the spinal metastases segmentation (see Section 2.2.4), U-net and vU-net architectures were applied to this task (see Figure 4.2). Again, a patch size of 128 $\times$ 128 pixels for 2D slices as U-net input and 128 $\times$ 128 $\times$ 64 voxels

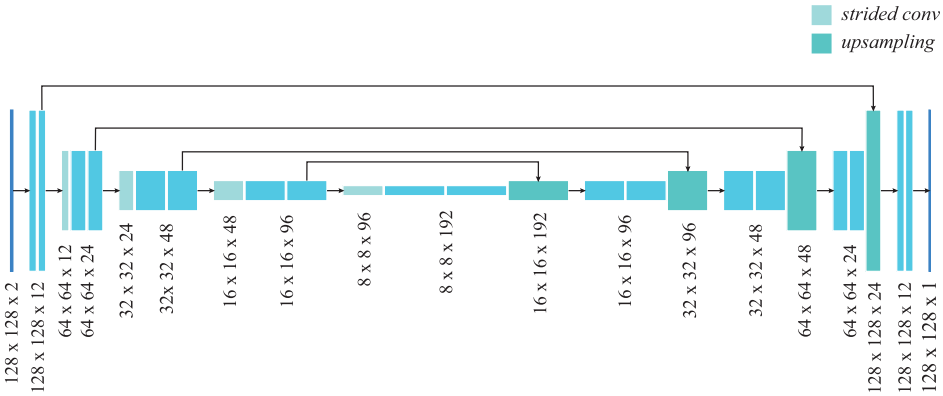

Figure 4.2: The U-net architecture used for multimodal 2D image input. The vU-net architecture for 3D input was analogous, but with an additional dimension for all layers and convolutional kernels. A significant difference between the two variants was the number of trainable parameters, which was about 2.85 times higher for the vU-net variants.

for volumetric vU-net input was defined, while an additional dimension contained either one or multiple MRI sequences as channels. According to the implemented network variant, the convolutional layers had a kernel size of 3×3 ($\times$ 3), except for the last one, which applied a 1×1 ($\times$ 1) kernel to reduce the dimensionality to the desired output size. Subsequent to each convolutional layer a batch normalisation followed. Downsampling was performed by strided convolutions (stride of 2) and simplified upsampling layers replaced the commonly used up-convolutions (Isensee et al., 2017). ReLU was implemented as the activation function of each convolutional layer, except for the last layer, where a sigmoid function was used to transform the resulting values in the range between zero and one. Again, only a single epoch was used with iterations equal to the number of available training samples, Tversky loss was implemented as a loss function and Adam was used as an optimizer with a starting learning rate of 0.01. With regards to the limited GPU memory resources, especially if incorporating image volumes, mini-batch sizes were defined as 2 for volumetric input of the vU-nets and 32 for 2D image slices as U-net input. Finally, a threshold of 0.5 was applied to produce binary output images.

Experimental Design

Similar to the experimental design described in Section 2.2.4, various input and network configurations were tested. Thus, different MRI sequences were applied either as monomodal or multimodal input to both, the U-nets and vU-nets. Dependent on the number of available patient cases, stratified 6-fold cross-validation over disjunct subsets was performed, with either four or five patients within each validation set. The results stated in the following refer to the average over all 6 cross-validation folds. The problem of limited training data was even more critical for this task, as only two thirds of the patient data were available compared to the spinal metastases segmentation of Section 2.2.4. Hence, a separate test set

was not comprised, since it would further reduce the available samples which hardly seemed to indicate promising training results. This, however, did not lead to biased results or undermined an un-biased evaluation of the trained models, since no training and design decisions were based on intermediate validation results (no look-ahead bias).

Evaluation

Reference, i.e. ground truth, segmentations of the necrosis zones were performed by a neuroradiologist using contrast-enhanced T_1-weighted and native T_2-weighted MRI sequences of each patient. For this purpose, a manual segmentation framework was set up to enable slice-wise contouring of the necrosis zones with subsequent 3D merging to volumetric masks. These segmentations were fed to the CNNs along with the corresponding image data in order to train the model, which subsequently predicted binary segmentation masks of unseen necrosis zones. DSC, sensitivity and specificity, as well as ASD and HD95 were used as quality measures to evaluate the produced segmentations. Those measurements always referred to 3D patient volumes, which were either directly produced by the vU-nets or as patient-wise merged 2D predictions by the U-nets.

4.1.5 *Results*

For the purpose of assessing the segmentation accuracy of the proposed method and to classify the results as sufficient or not, the IRV was used as a benchmark. Therefore, a second ground truth segmentation of each patient was produced in order to match it with those of the first expert. The DSCs between these two expert segmentations ranged between 75.2 % and 88.9 % with on average 82.4 $\pm$ 5.9 %. The surface distances were on average 1.43 $\pm$ 0.52 mm ASD and 6.71 $\pm$ 3.10 mm HD95. Table 4.1 shows the automatically produced results by the specific CNNs. Depending on the applied MRI sequences and network variant, the proposed method achieved DSC scores up to 77.2 $\pm$ 15.6 %, sensitivity rates up to 86.4 $\pm$ 15.4 % and ASD of 1.59 $\pm$ 1.09 mm. Prediction time was on average 4.29 ms per slice and therefore approximately 0.28 s per patient case for the 2D U-nets, and on average 0.21 s per patient case for the 3D vU-net configurations.

4.1.6 *Discussion*

The segmentation of RFA-induced necrosis zones either done manually or with the aid of computer-assisted methods, is a necessary prerequisite for any objective therapy outcome validation, e.g. as presented in the following section. To perform quantitative and volumetric evaluations of tumour coverage and to draw conclusions about the protection of risk structures, segmentation masks of the resulting necrosis zones are indispensable. With regards to clinical applicability, the accuracy, degree

Table 4.1: Experimental results for each input configuration depending on the used modalities (contrast-enhanced (ce) T_1-, T_2-weighted MRI sequences), as well as a slice-wise (U-net) or volumetric (vU-net) processing. Dice scores (DSC), sensitivity rates (TPR), specificity rates (TNR), average surface distances (ASD), and 95th percentile Hausdorff distances (HD95) were averaged over all cross-validation folds.

		U-net			vU-net		
		ceT_1	T_2	$[ceT_1 + T_2]$	ceT_1	T_2	$[ceT_1 + T_2]$
DSC [%]	mean	76.7	62.2	77.2	72.7	60.4	75.9
	median	83.5	65.9	82.3	78.1	63.3	80.7
	std	19.0	21.7	15.6	18.7	27.4	17.2
TPR [%]	mean	81.4	69.2	81.6	77.7	63.1	77.8
	median	84.5	76.9	86.4	86.2	71.1	82.6
	std	17.1	20.3	15.4	21.5	26.5	19.6
TNR [%]	mean	99.2	98.8	99.2	99.1	98.7	99.2
	median	99.6	99.1	99.5	99.4	99.0	99.6
	std	0.9	1.0	0.8	0.9	1.3	0.9
ASD [mm]	mean	1.59	2.49	1.69	2.18	2.99	1.90
	median	1.14	2.25	1.18	1.68	2.69	1.52
	std	1.09	1.53	1.08	1.60	1.76	1.42
HD95 [mm]	mean	6.27	10.16	7.28	9.11	10.89	7.32
	median	3.99	7.12	4.38	4.85	8.70	6.42
	std	5.75	10.56	6.92	9.28	11.24	6.42

of automatisation, and the required computational time are of crucial importance. Manual segmentation procedures have proved to be time-consuming and tedious due to their limitation to slice-by-slice processing and required up to 10 min per necrosis zone. The presented automatic approach, however, could reduce the required computational time per patient case to well below half a second and thus drastically speed up the post-interventional treatment validation procedure.

With respect to the segmentation accuracy, the inter-reader variability was the defined benchmark. The best results were achieved combining 2D patches of the contrast-enhanced T_1- and the T_2-weighted MRI sequences applied to a U-net and resulted in a mean DSC of 77.2 %. Although the results were on average qualitatively not quite at an IRV level (mean DSC of 82.4 %), they were still within the defined objective range of the

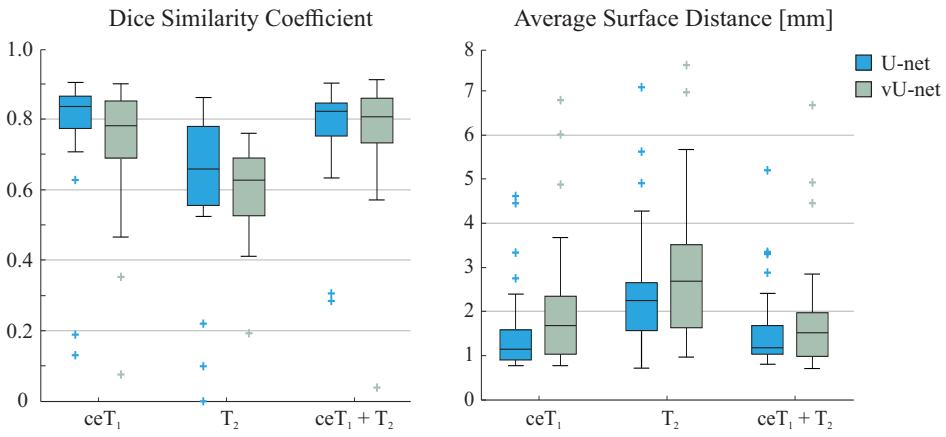

Figure 4.3: DSC and ASD scores depending on the imaging modality and input dimension. Box edges mark the 25th and 75th percentiles, the central box line marks the median value and the whisker marks the most extreme values not considered as outliers.

expert variability (75.2 % and 88.9 % DSC). Nonetheless, five patient cases out of the total 26 did not achieve the required accuracy with this best input and network configuration and would have needed a manual post-correction to meet the clinical objective. These patient cases were either hampered by distinctly poor image contrasts or constituted an anomalous exception due to the presence of inflammation processes and hematomas (see Figure 4.4). As indicated above, the best results were achieved with either the contrast-enhanced T_1-weighted data alone or if it was part of a multimodal image input (see Figure 4.3). This could most likely be attributed to the predominantly high image contrasts between necrosis and surrounding bone marrow signals, which were additionally enhanced by the injection of contrast agents. Although, the segmentation accuracy could only be slighty improved in combination with T_2-weighted MRI data, the standard deviation could be reduced considerably from 19.0 % to 15.6 % for the 2D U-nets (see Table 4.1). Applying solely T_2-weighted images yielded worse results, most likely due to the inferior image contrast between necrosis and surrounding tissues. Hence, they rather supported and improved robustness in combination with the contrast-enhanced T_1-weighted input. Similar findings have already occurred with regard to spinal metastasis segmentations, and likewise, it was found that on average the 2D U-nets yielded more accurate segmentation results than the 3D vU-net variants. Again, it is reasonable to assume that the higher number of trainable parameters to be optimised in 3D CNNs (4,000,000 vs. 1,400,000) and, therefore, the increased network complexity had a rather disadvantageous effect if applied to such a limited amount of training data.

Since this was, to the best of the author's knowledge, the first study to tackle necrosis zone segmentation in spinal MRI, a comparison with existing literature was only possible to a limited extent. Moreover, almost all relevant works regarding necrosis zone segmentation did not state any quantitative results, except for the work of Egger et al. (2015). The best

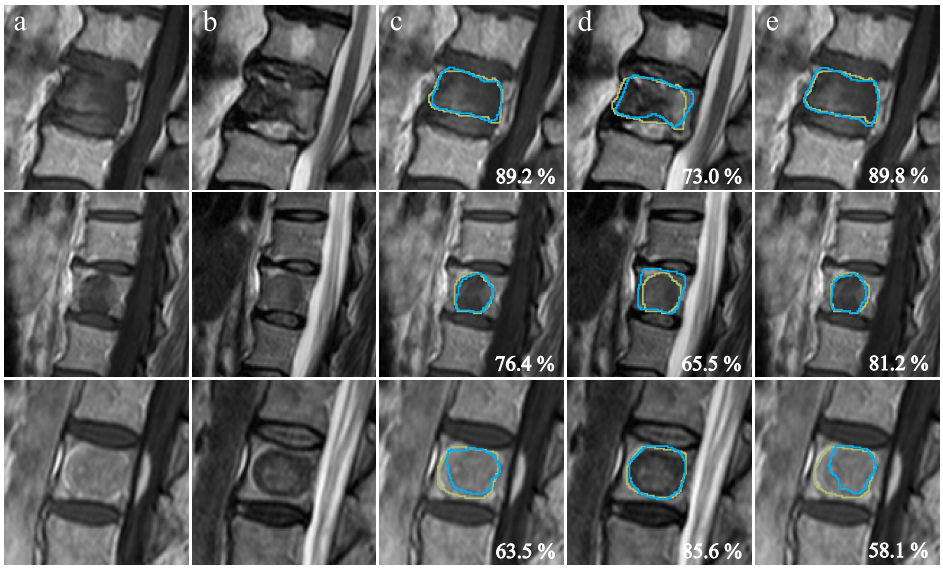

Figure 4.4: Three exemplary patient cases, representing very high (top row), average (middle row) and unsatisfactory (bottom row) segmentation accuracies produced with 2D image input. Corresponding DSC scores are stated in the lower right corners. Green contours display the expertly annotated data as ground truth and blue contours represent the automatically produced segmentations. From left to right: (a) original T_1-weighted MRI sequence, (b) original T_2-weighted MRI sequence, (c) result with only T_1-weighted image data, (d) result with only T_2-weighted image data, (e) result with combined T_1- and T_2-weighted image data.

results achieved with the presented approach were on a par with those of their semi-automatic method (77.2 % vs. 77.0 %). Due to the lack of a directly related state of the art, it might be interesting to take works of automatic lesion segmentation as a similar issue into consideration. For instance, Chmelik et al. (2018) adapted a CNN to vertebral metastasis segmentation in CT images and achieved voxel-wise sensitivities of 74 % for sclerotic and 71 % for lytic lesions as well as specificities of 88 % (sclerotic) and 82 % (lytic). Section 2.2.5 of this thesis showed the results of a spinal metastasis segmentation in MRI that applied the same approach as described above. Considering spinal metastases, the best network configuration achieved a mean DSC of 77.6 %, which is on a par with the average 77.2 % DSC of the necrosis zone segmentation presented in this section. Although the results were quite similar, the segmentation of spinal metastases represented a more ambitious task due to the higher appearance variability depending on the lesion type (lytic, sclerotic or mixed) and MR sequences. In contrast, necrosis zones predominantly emit hypointense image signals in almost all MRI sequences w.r.t. surrounding bone marrow. Accordingly, higher accuracies could be expected for this task, but these were presumably undermined by even less available training data, which further restricted generalisability of the CNNs. In addition, the segmentation of necrosis zones in spine MRI was hampered by distinct difficulties, e.g. the challenging delineation from remaining

metastases and inflammation processes or since possible hematomas could overlap with the necrosis zones (see Figure 4.4).

4.1.7 *Conclusion*

Automatic necrosis zone segmentation in follow-up MRI scans after RFAs of spinal metastases has the potential to quantify and objectify the treatment outcome validation. Additionally, it can provide important information regarding the improvement of ablation procedures and may help to understand and predict possible tumour recurrence more reliably. In this section, a CNN-based segmentation approach was presented and it was examined how various input modalities and dimensions affected the segmentation accuracy. The best network and input configuration achieved on average 77.2 % DSC with a median of 82.3 %, showing overall sufficient segmentation accuracy, although rare deficient cases have been observed. The latter could be attributed in particular to a distinctly poor image quality and the presence of inflammatory processes and hematomas. Overall, the presented method indicated promising results in the vast majority of the patient cases with expert-like segmentation quality, and therefore, constitutes a valuable approach towards this ambitious and challenging issue.

4.2 TREATMENT OUTCOME VALIDATION

4.2.1 *Introduction*

A framework for a comprehensive and automatised evaluation of the treatment outcome of RFAs of spinal metastases represented the concluding step of a computer-aided and advanced therapeutic workflow. In order to assess and quantify the results of the ablation process, it is essential to compare pre-interventional and follow-up MRI scans and evaluate the congruence of both the former metastasis and the necrosis zone via distance and volume overlap measures. Furthermore, to draw conclusions regarding the integrity or potential impairment of surrounding risk structures, such as the spinal cord, the automatic computation and visualisation of safety margins towards the ablation zone is equally important.

Currently, the treatment outcome validation is done by separately examining and mentally matching the pre- and post-interventional images without any computer-assisted image fusion or superimposed segmentation masks. Accordingly, the current procedure is inconvenient and highly demanding for any radiologist due to the mental fusion of 3D image volumes, which takes considerably long time without being very accurate or reproducible. For instance, only small contrasts between necrosis and residual tumour tissue as well as inflammation processes significantly hamper the assessment of tumour coverage and may lead to misinterpretations. These aspects hardly enable the radiologists to draw objective,

quantifiable, robust and reliable conclusions about the RFA outcome. However, at this very point an appropriate computer-assisted validation tool can help to remedy this lack in clinical practice and support the radiologists in their mandatory assessment, without any additional workload. For this purpose, such a framework should cover all necessary steps and aspects of a post-RFA treatment validation procedure and automatise and speed up the entire process itself.

4.2.2 *State of the Art*

There are several works regarding computer-assisted approaches for treatment outcome validation based on image registration of pre- and post-treatment images, however, they mainly focused on hepatic lesions in CT imaging. McCreedy et al. (2006) presented a tool to guide physicians through segmentation, registration and visualisation steps in order to evaluate ablation procedures of liver tumours. Multiple registration methods were included in their tool, ranging from user-guided landmark-based to automatic voxel-based approaches. Besides the computational times of 8 - 10 min for affine and 2 - 3 h for deformable methods, no quantitative results regarding the registration accuracies were stated. In addition, they did not provide any information about the computation of treatment validation measures, but rather focused on supporting visualisations.

Fujioka et al. (2006) proposed an automatic rigid registration and subsequent fused visualisation of pre- and post-interventional liver CT data. In 16 out of their 20 patient cases, the automatic registration based on normalised correlation coefficients was sufficient enough to avoid manual post-correction and achieved registration errors of on average 3.3 ± 1.2 mm. The total time required for their image fusion was <10 min. Again, no treatment validation measures were computed.

Based on their previous work regarding a user-initialised automatic rigid registration of pre- and post-RFA CT images (Rieder et al., 2012), Rieder et al. (2014) assessed an outcome validation software in a retrospective clinical study regarding hepatic lesions. They focused in particular on validation support via suitable visualisations, including traffic light colour-coded tumour coverage and a spherical mapping of the tumour surface, which was previously presented in Rieder et al. (2010). The mean registration accuracy measured as FRE was 6.24 mm and the mean computational time per registration was 45 s (Rieder et al., 2012). The main finding of their study was that in contrast to the conventional image reading, a validation tool, as presented, increased the percentage of correctly identified local tumour recurrence, while the prediction performance was noticably affected by the registration accuracy.

Passera et al. (2013) presented a tool to register pre- and post-treatment CT images of hepatic lesions to quantitatively assess the therapeutic success of RFAs. Considering this, they included segmentation (described in Section 4.1.2), registration as well as validation and visualisation procedures. B-spline deformations with normalised mutual information (NMI) as similarity measure were used for the image registration and required

approximately 40 min on medium consumer hardware. In 2 out of their 10 patient cases their registration approach failed, although no accuracies w.r.t. the successful cases were stated whatsoever. Subsequent to the preceding segmentation and registration of the lesion and ablation zone, an overlap measure classified the tumour coverage in fully or not-totally treated, which additionally was shown in 2D and 3D visualisations of the segmentation masks as well as colour-coded volumes of ablated and residual tumour tissue.

Similar to Passera et al. (2013), the work of Laura et al. (2016) implemented a deformable registration approach for liver CT images of minimally invasive interventions. They combined surface landmarks obtained from a preceding liver segmentation with information from internal vessel segmentations to register pre- and post-interventional images via a finite element method in a more physiologically accurate way as, for instance, with solely liver surface-driven approaches. They used manually identified landmarks on hepatic vessel structures for evaluation purposes and achieved a mean FRE of 1.2 mm. However, no required computational time was stated. Subsequent to a semi-automatic segmentation of the former lesion and resulting necrosis zone, the treatment efficiency was quantitatively measured via tumour coverage and a classification was provided in order to support the physician's decision-making regarding re-treatment of residual tumours.

In contrast to the above-mentioned works, which all focused on liver lesion ablations in CT imaging, von Dresky et al. (2018) stated tumour coverage measures, while optimising generator modulation and electrode positions for RFAs of spinal metastases. However, they neither described any registration procedure nor stated the accuracy, which is crucial for such a validation tool to be reliable, since the subsequent treatment outcome assessments are based on these results, as demonstrated by Rieder et al. (2014).

In conclusion, while there are several works regarding a computer-assisted treatment outcome validation of minimally invasive interventions like RFA, except for the work of von Dresky et al. (2018), all focused on hepatic lesions and CT imaging. Besides, all of the above-mentioned works were limited by either one or multiple of the following constraints: inaccurate registration accuracy (respectively no stated accuracy), long computational time or considerably large user input was required. Hence, there was still a need for an RFA outcome validation framework that combines target structure segmentations, a fast, precise and automatic image registration of spinal MRI as well as the computation and intuitive visualisation of quantitative validation measures for decision support of radiologists.

4.2.3 *Objectives*

In order to support the radiologist with a computer-assisted treatment outcome validation framework, the following objectives were defined in cooperation with the clinical partner:

- Framework must provide all image processing aspects necessary for a throughout treatment validation support
- Computational time of the integrated image registration per patient case on average < 1 min on current consumer hardware
- FRE in subvoxel range of each patient case (half of the diagonal of a voxel)
- Automatic computation and visualisation of relevant validation measures

Since all related works focused on hepatic interventions or had limitations, a novel framework had to be designed and implemented for this specific task. First, such a computer-assisted tool must guide and support the radiologist throughout each step of the validation procedure, i.e. segmentation support for the purpose of metastasis and necrosis contouring, a precise and fast image registration, as well as the computation of relevant quantitative validation measures. Additionally, interactive 2D and 3D visualisations should help the radiologist to quickly and comprehensively capture the tumour coverage or possibly violated safety margins. Considering the segmentation support, the intended framework should offer automatic segmentations as proposed in Section 2.2.1 and 4.1.1, as well as an assisted but largely manually performed method as a post-correction or fall-back solution. Since both automatic segmentation approaches were described and discussed in previous sections of this thesis, they will not be addressed once more in the following. This section will mainly focus on the required registration and subsequent treatment outcome validation.

Analogous to the objectives of the previous tasks, short computational times were also critical for this step of the workflow. Manual image registration, especially in 3D, is commonly time-consuming, cumbersome and unnecessarily ties up valuable resources. This was proven, for instance, during the preparation of the ground truth of a single patient case, which consisted of multiple corresponding landmark pairs in the pre- and post-interventional MRI and required approximately 10 min. This procedure should be shortened significantly through the implementation of an automatic registration approach, which should not exceed one minute while being sufficiently accurate at the same time. A particular threshold regarding the registration accuracy in order to classify a result as sufficient should be chosen with particular care, since the subsequent computation of validation measures and their effect on the decision support of the radiologists depends on the registration accuracy (Rieder et al., 2014). Therefore, the threshold was set as low as possible and yet as high as

reasonably achievable. For this purpose, a maximum FRE of half the diagonal of a voxel was defined, ensuring a registration accuracy of subvoxel size. In conclusion, the required treatment validation framework had to combine segmentation support, precise image registration, as well as compute and visualise relevant RFA validation measures like the tumour coverage or the safety margin towards risk structures.

4.2.4 *Materials and Methods*

The outcome validation tool was developed as an application prototype in the modular medical image processing and visualisation framework MeVisLab (MeVis Medical Solutions AG, Germany). The graphical user interface (GUI) as well as most of the features of the proposed tool were based on such available software modules and integrated third party libraries and techniques. To start with, the pre- and post-RFA MRI scans could be imported as DICOM or Nifti image data, which represent the most common types of medical image data. The segmentation tasks could be performed within the presented tool either interactively with a live-wire-based approach or previously performed segmentations could be imported and edited as 3D binary masks or as MeVisLab internal contour segmentation objects (CSO). Automatic segmentations could be performed with the approaches presented in the Sections 2.2.4 and 4.1.4.

The implemented registration approach was initialised by a superposition of the centers of mass of both segmentation masks. Subsequently, a multiresolution image-based rigid registration approach precisely registers both corresponding image volumes combining NMI and a Quasi-Newton optimizer. NMI was used as it is established as one of the most common metrics for multimodal registrations. NMI binning was set to 32 and depending on the original image resolution 4 to 6 multiresolution levels were used, which was automatically determined.

Following the image fusion step, various treatment outcome validation measures were automatically calculated (see Figure 4.5). Starting with the residual tumour size T_R, which was defined by all voxels belonging to the tumour mask but were not included in the contoured necrosis zone (in relation to the whole metastasis volume). Furthermore, surrounding non-cancerous tissue should be spared as much as possible, which was also considered as a validation index N_{NC}. It is defined as the fraction of necrotised non-cancerous tissue with regard to the total ablation volume. The distance d_C between the centers of mass of metastasis and necrosis zone was computed, providing a measure to assess RFA applicator placement and the resulting ablation zones w.r.t. the target volumes. Finally, a minimal surface distance d_R between the necrosis zone and any adjacent and defined risk structure was computed to ensure the integrity of critical safety margins. For qualitative validation support, a slice-wise visualisation including superimposed segmentations of the metastasis and the corresponding necrosis as well as an interactive volume-rendered scene have been integrated. A traffic light colour scheme showed the individual volume fractions of residual tumour (red), ablated non-cancerous tissue

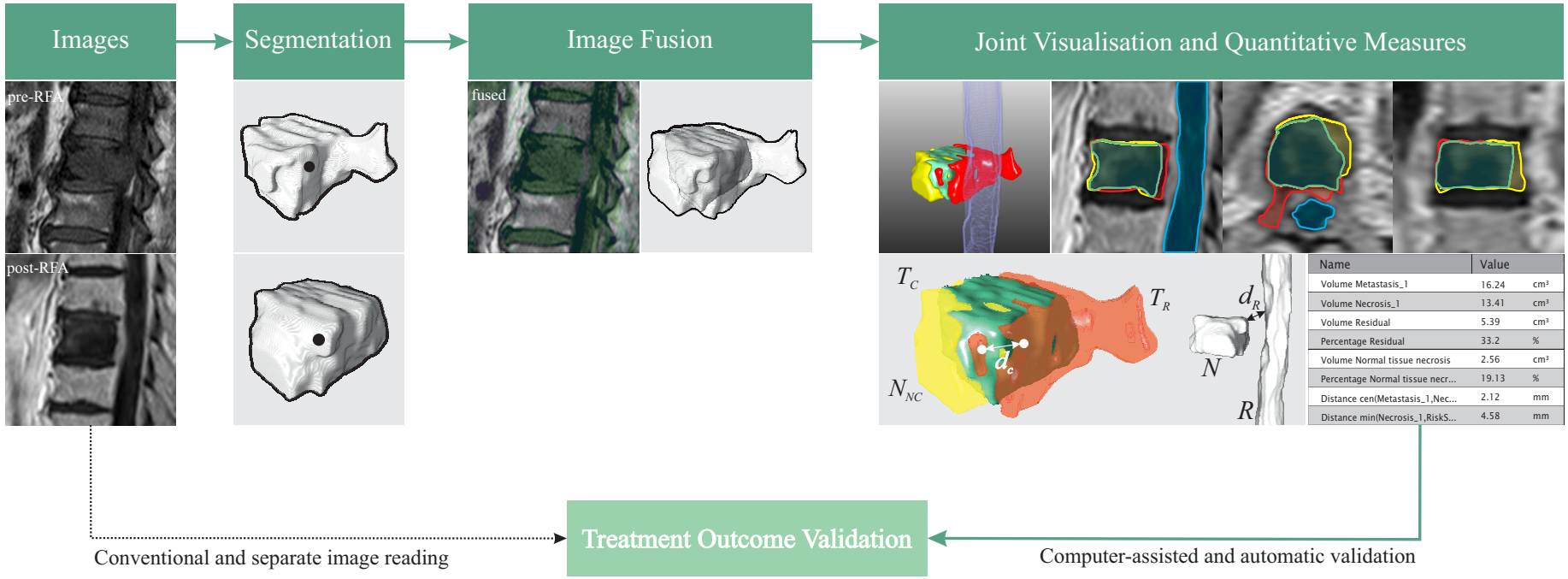

Figure 4.5: Illustration of the proposed computer-assisted treatment outcome validation framework: First, the target structures, i.e. metastasis and necrosis zone, are segmented in the pre- and post-RFA MR images. Subsequent to an initial prealignment an automatic voxel-based registration fuses both image volumes and enables the computation of quantitative outcome validation measures as well as joint visualisations. A traffic colour light scheme simplifies the interpretation of spatial relationships and context information, e.g. the tumour coverage T_C (green), the fraction of necrotised non-cancerous tissue N_{NC} (yellow) or the residual tumour volume T_R (red). Additionally, d_R as the minimal distance between the necrosis zone N and a risk structure R (e.g. spinal canal) and d_C as the distance between the centers of the metastasis and necrosis after the registration process are computed.

(yellow) and ablated tumour (green). A visualised margin surrounding the metastasis, as commonly integrated in RFA validation softwares of hepatic lesions did not seem to be useful for this particular task. It was not implemented, since ablations beyond the size of the metastases in the confined space of the vertebral bodies are scarcely possible, especially due to the natural limitation by surrounding cortical bone tissue (Greenberg et al., 2014).

In terms of an application example, the proposed framework was applied to a retrospective study of 30 patients who underwent RFAs of spinal metastases at the Department of Neuroradiology of the University Hospital of Magdeburg. Each patient case consisted of pre-interventional T_1- and follow-up contrast-enhanced T_1-weighted MRI sequences, which are typically used for diagnostic and validation purposes (Shah and Salzman, 2011). The in-plane image resolution ranged from 0.47 mm to 1.25 mm (average 0.63 mm) and the slice spacing was 3.30 mm for all scans. In addition, a radiologist contoured metastases and necrosis zones of each patient case using the integrated semi-automatic and live-wire-supported segmentation tool.

4.2.5 *Results*

Image Registration

Since the accuracy of the image registration was of crucial importance for the following treatment outcome validation, it was examined for each patient case. The ground truth for this evaluation consisted of corresponding pairs of manually placed landmarks on the vertebrae adjacent to the metastasis and necrosis zone, respectively in both image volumes. Thus, a field expert marked four fiducials within the sagittal plane of symmetry of the vertebra; two on the cranial endplate of a vertebra above and two on the caudal endplate below the metastasised vertebra. For this purpose, the slice spacing was interpolated to provide a quasi-isotropic spatial resolution in order to reduce the FLE to a minimum. Registering both fiducial sets of each patient directly via Horn's quaternion-based method resulted in mFRE of average 0.72 ± 0.15 mm. The mFRE represented both, the result of a manual registration by landmarks and the minimum possible registration error that can be achieved, while measuring accuracy with the given fiducials. The average FRE, i.e. the accuracy of the integrated automatic voxel-based registration approach was computed between each pair of corresponding fiducials per patient via Euclidean distances. The discrepancies were averaged over all fiducial pairs per patient and subsequently over all patient cases and resulted in an FRE of 1.57 ± 0.32 mm. The entire registration procedure never exceeded 15 s.

Outcome Validation

Table 4.2 lists the results obtained with the outcome validation tool for each patient of the retrospective study. The metastasis sizes of the exam-

ined clinical dataset ranged from 2.85 cm^3 to 52.88 cm^3 with a mean of 24.84 $\pm$ 13.28 cm^3. In less than 15 % of the given patient cases the necrosis zone was larger than the corresponding metastasis, resulting in a mean size of 17.06 $\pm$ 8.69 cm^3 (min. 3.83 cm^3, max. 42.11 cm^3). Due to the restrictions imposed by limited intravertebral space and the need for safety margins towards surrounding risk structures, there was no patient case with an entirely covered tumour (the closest to this were 98.1 % coverage of patient 22, see Table 4.2). On average 44.1 $\pm$ 20.82 % of the original tumorous tissue was not covered by the necrosis zone in the follow-up scans and therefore remained untreated aside from apoptotic impacts. In addition, on average 23.26 $\pm$ 24.87 % of the segmented necrosis zone did not overlap with the metastasis mask and thus could be labelled as necrotised non-cancerous tissue. The minimum surface distance d_R between the necrosis zone and the spinal canal as an adjacent risk structure was on average 3.58 $\pm$ 1.86 mm with no intersection between the two structures in any patient case. The distance between the centers of mass of each metastasis and corresponding necrosis was on average 3.05 $\pm$ 2.84 mm with a minimal distance of 0.17 mm (patient 6) and a maximum distance of 13.46 mm (patient 13).

4.2.6 *Discussion*

A pre- and post-interventional image registration with preceding segmentation procedures can be a valuable tool for efficient and objective treatment outcome validation (Laura et al., 2016; Rieder et al., 2014). Figure 4.6 shows the prototypal GUI of the presented framework, combining automatically generated quantitative measures with interactive visualisations. This supports the radiologist with a straightforward and intuitive representation of the spatial relationships of the individual structures, which is considerably more challenging to comprehend if both images were viewed separately in individual viewers or even on different monitors as it is common practice. Furthermore, measures derived from the overlap of pre-interventional metastasis and post-RFA necrosis zone quantify the treatment outcome and enable a comprehensible, reproducible, and more objective validation. It has already been proven that this can improve clinical decision-making, for instance, in terms of identifying local tumour recurrence (Rieder et al., 2014). However, the prediction performance with the aid of computer-assisted validation frameworks largely depends on the accuracy of the integrated registration procedures (Rieder et al., 2014). Therefore, it is required to ensure high precision, which means FREs ideally in a subvoxel range. The registration approach integrated in the proposed treatment outcome validation framework achieved accuracies of less than half the patient-individual voxel diagonals (on average 1.71 mm vs. on average 1.57 mm FRE) for each patient case. Considering additionally the mFRE of on average 0.72 mm, it indicated sufficient image fusion accuracy of the presented tool.

Since the patients in this particular study underwent RFAs of spinal metastases for palliative treatment due to advanced tumour stages and of-

Table 4.2: Quantitative validation results of a retrospective study concerning RFAs of spinal metastases. For each patient case (ID - patient case) the tumour size T, the necrosis zone size N, the percentage of residual tumour T_R w.r.t. the pre-RFA size, the percentage of necrotised non-cancerous tissue N_{NC} w.r.t. the total ablation zone, the minimum distance d_R between necrosis zone and a risk structure, and the distance d_C between the centers of mass of metastasis and necrosis zone were automatically computed.

ID	T [cm^3]	N [cm^3]	T_R [%]	N_{NC} [%]	d_R [mm]	d_C [mm]	ID	T [cm^3]	N [cm^3]	T_R [%]	N_{NC} [%]	d_R [mm]	d_C [mm]
1	22.51	14.91	45.33	17.46	5.07	1.52	16	11.55	3.83	70.50	11.14	6.15	1.99
2	10.19	9.85	68.37	67.26	3.99	4.11	17	16.24	13.41	33.21	19.13	4.88	2.12
3	30.92	16.77	50.95	9.57	1.06	2.05	18	30.78	19.23	41.70	6.69	4.05	0.02
4	32.84	22.40	35.41	5.32	0.25	5.96	19	2.90	14.02	26.41	84.76	5.26	3.95
5	17.61	13.33	27.38	4.05	3.16	3.57	20	4.59	2.91	42.06	8.75	0.64	2.82
6	42.30	18.03	79.39	51.65	3.77	0.17	21	52.88	11.92	78.57	4.91	8.91	3.37
7	2.85	9.26	13.66	73.47	3.78	3.36	22	13.41	18.84	1.87	30.19	2.27	6.64
8	27.34	29.58	18.85	25.00	3.83	1.86	23	20.62	11.41	47.01	4.27	3.25	0.60
9	46.19	16.29	70.10	15.24	5.58	1.16	24	38.61	28.40	29.09	3.59	1.77	2.35
10	25.88	42.11	35.96	60.65	3.32	0.82	25	30.37	23.28	27.63	5.59	0.90	1.85
11	38.90	17.29	56.29	1.67	3.14	0.43	26	28.72	17.69	40.44	3.30	4.65	1.65
12	34.65	19.66	45.85	4.58	3.72	1.70	27	13.55	12.72	13.50	7.82	4.16	9.56
13	18.48	9.01	53.76	5.21	2.17	13.46	28	11.46	5.16	73.31	40.78	1.86	4.68
14	31.73	30.59	56.78	55.17	3.81	1.17	29	30.41	26.95	20.34	10.10	5.22	1.67
15	14.26	9.41	69.04	53.08	6.15	3.75	30	42.55	23.57	48.68	7.36	1.63	3.20

ten highly space-occupying metastases, there were no tumours classifiable as fully-treated. The percentage of residual tumour tissue was between 1.9 % and 79.4 % with a mean of 44.1 %. This was primarily due to the spatial limitations within the vertebrae and the required safety margins between necrosis zones and adjacent surrounding risk structures, such as the spinal cord or nerve roots. Compliance with the safety requirements for this study was proven by the fact that no intersection between necrosis zone and spinal canal could be observed in any patient case.

The centers of mass distance between the metastasis and necrosis represents a valuable measure for validating and optimising applicator placement and configurations of multipolar RFAs. Especially the intra-vertebral position of the applicator needle tip has a crucial impact on the resulting ablation zone, since cortical bone tissue notably differs from cancellous bone and bone marrow w.r.t. heat propagation and thus influences the time-dependent development of the necrosis zone differently (Greenberg et al., 2014). Therefore, the distance between the centers of mass of both segmentations may play an important role in therapy planning and simulation of RFAs in addition to the outcome validation. The retrospective comparison of both the simulation and the actual treatment outcome enables a successive adaptation and improvement of future simulations for more precise and reliable outcome predictions.

Considering the related work, almost all of the previous studies focused on hepatic tumour coverage and CT images, only one study was found dealing with RFAs of spinal metastases (von Dresky et al., 2018). Each of the relevant works showed limitations regarding either the registration accuracy (Rieder et al., 2012), the required computational time of their approaches (Fujioka et al., 2006; McCreedy et al., 2006; Passera et al., 2013) or they did not provide information on at least one of these matters (Laura et al., 2016; McCreedy et al., 2006; Passera et al., 2013), which limited the assessment of their clinical applicability. In contrast to all relating studies, the proposed framework also included surrounding risk structures in the treatment validation, since in RFAs of vertebral metastases the spinal canal in particular must be protected from necrotising effects and therefore, constitutes a critical indicator for the treatment outcome. Regarding RFAs of spinal metastases, von Dresky et al. (2018) evaluated their proposed approach on two patient cases and stated tumour coverages of 91.8 % and 85.3 %, as well as 46.4 % and 24.1 % necrotisation of non-metastasised tissue w.r.t. the total ablation zone. In comparison with the tumour coverage results given above, on average a substantially larger fraction of the metastases was ablated in the study of von Dresky et al. (2018), while the proportion of necrotised non-tumorous tissue was similar. However, due to the comparatively small dataset (n = 2) and the lack of any information about the registration approach and its accuracy, significance of the results and the conclusions was limited.

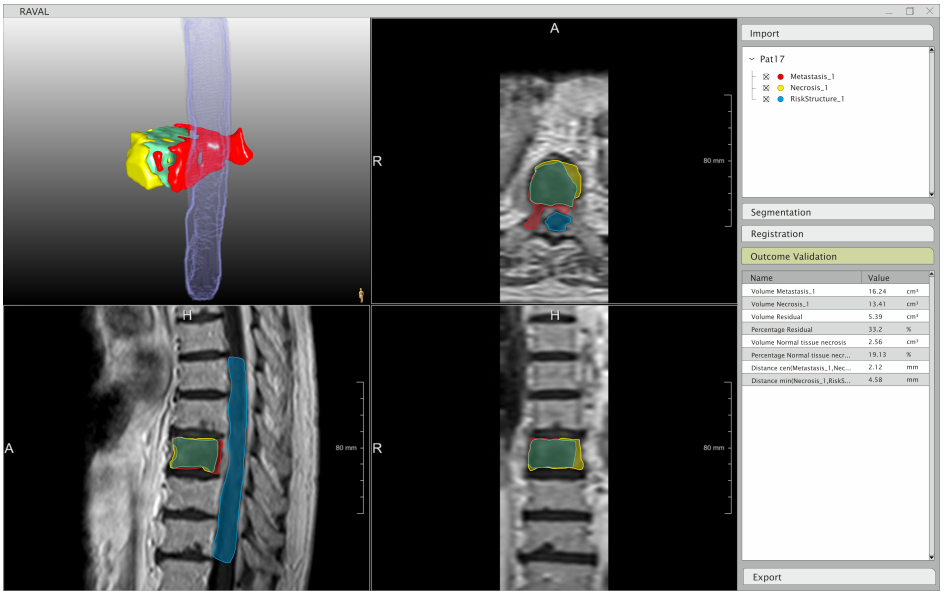

Figure 4.6: Graphical user interface of the proposed treatment outcome validation tool with both 2D and 3D visualisations, as well as quantitative measures for a patient-individual outcome validation.

4.2.7 *Conclusion*

This section presented a novel framework to provide the radiologists with qualitative and quantitative information to support treatment outcome validation of RFAs of spinal metastases. The proposed framework covered each aspect of a computer-assisted post-treatment workflow: target structure segmentations, image fusion of pre- and post-RFA images, and the computation and visualisation of all relevant validation measures. The integrated image registration proved to be fast and accurate, which is of crucial importance for a reliable assessment of the interventional results and clinical decision-making. The framework itself can be integrated at the end of the clinical workflow to finally evaluate and document the results and outcome of the preceded treatment, without additional time and work effort for the radiologists. This is therefore the concluding step of a more sophisticated and purposeful computer-assisted workflow as presented in the introductory chapter in Section 1.3.

5

SUMMARY

With increasing survival time of most patients with carcinomas due to advanced diagnosis and treatment strategies, the incidence of spinal metastases is steadily rising. Currently, about 60-70 % of all patients with systemic cancer will develop spinal metastases. Accordingly, there is a growing need for computerised approaches to support and relieve radiologists with the steadily increasing number of spinal interventions. In the past decades, various medical image processing methods for a vast number of purposes have been proposed, but only few were tailored to the specific requirements and needs of clinical routine. In contrast, this thesis presented novel strategies to support radiofrequency ablations of spinal metastases, which were primarily developed for this clinical scenario and were comprehensively evaluated on data of numerous patient cases. For that purpose, a thorough analysis in cooperation with the clinical partner from the Department of Neuroradiology of the University Hospital of Magdeburg was done to identify multiple aspects of the current treatment workflow, which could be improved by novel computer-assisted approaches. The focus was set on reducing the radiologists' workload and time needed for recurring routine procedures, to enhance precision and speed of interventional procedures like metastasis puncture, and to support decision-making and post-treatment validation tasks. In this respect, the state of the art in medical image processing was analysed and the existing gaps and limitations were overcome by the proposed approaches in order to ensure suitable and applicable assistance starting from the therapy and intervention planning phase, to the intervention itself, to the subsequent assessment of the treatment outcome.

Considering the support of the pre-interventional planning phase, a precise, fast and robust vertebral body segmentation and a novel segmentation method for spinal metastases were presented. The former consisted of a hybrid level-set approach combining regional intensity and boundary features with a preceding bias field correction. A wide range evaluation set comprised of image data from different sources, e.g. patient cases with various spinal pathologies as well as healthy study participants, demonstrated sufficient accuracy with a mean DSC of 86.0 %, short computational times with only 5.4 s per vertebra and robustness towards multiple MRI sequences and protocols, all of which is indispensable for clinical applicability. In addition, this thesis addressed the highly challenging task of spinal metastasis segmentation in MRI and presented - to the best of the author's knowledge - the first approach for this purpose. The main challenge of this task was the high shape and appearance variability of the metastases, which differ in dependence of their origin, lesion type, spine and intra-vertebral location as well as dependent on the applied MRI sequences. Patient data with spinal metastases of vari-

ous origins and different lesion types were applied to U-net-like CNNs in order to investigate the most suitable input configuration w.r.t. MRI sequence, 2D or 3D patches and mono- or multimodal image input. With up to 77.6 % DSC, the proposed approach achieved on average almost expert-like segmentation accuracy, while only rare outliers lowered the mean accuracy and the vast majority of the patient cases showed DSCs of >80 %, resulting in a slightly higher median DSC compared to the inter-reader variability. Additionally, computational times under half a second represent an enormous advantage over manual procedures and do not cause any delay in the clinical workflow. Both of the above-mentioned methods, the vertebral body and spinal metastases segmentation, were able to meet all of the beforehand defined objectives by the clinical partner. Nevertheless, it is noteworthy that a crucial part regarding an enhanced therapy and intervention planning is in development by the cooperation partner MeVis Medical Solutions AG, Germany. This refers to a numerical simulation of the ablation process with a resulting necrosis zone prediction, which previously have been published focusing on hepatic lesions and metastases.

In order to assist the radiologists during the intervention itself, an image fusion method was presented to transfer planning data and supporting visual information, respectively onto the intra-operatively acquired images. A voxel-based and multiresolution registration approach precisely and rapidly matched diagnostic and poor quality interventional image data, while additionally addressing spine deformations due to different patient positioning by a multisegmental strategy. With a mean FRE of 2.35 mm and 24 s per segment the proposed approach demonstrated suitable applicability w.r.t. the defined clinical objectives.

Regarding the post-interventional phase of RFAs of spinal metastases, this thesis proposed a novel necrosis zone segmentation approach as well as an innovative framework to support the radiologist in the treatment outcome validation. The volumetry of the resulting necrosis zones after the ablation process is a crucial requirement for a reliable and objective assessment of the therapy results. For this purpose, a CNN-based necrosis zone segmentation strategy for follow-up MRI scans was presented, which was, to the best of the author's knowledge, the first for this particular issue. With a mean DSC of 77.2 % and an average ASD between the produced and the ground truth segmentation of 1.59 ± 1.09 mm the proposed method demonstrated sufficient segmentation accuracy, while keeping the computational times well under half a second. The thus produced necrosis zone segmentations can be used in combination with masks of the former metastases as an input to the subsequently presented treatment outcome validation framework. The superposition of the centerpoints of both segmentations initialised a precise image registration, which is required to compute quantitative validation measures, such as the tumour coverage. The presented framework automatically computes decisive validation measures and furthermore provides intuitive 2D and 3D visualisations of all relevant structures. Studies have already shown that computer-assisted tools for treatment outcome assessment - like the

presented framework - have the potential to enhance prediction of local tumour recurrence. The proposed framework concludes the intended computer-assisted workflow of RFA of spinal metastases.

With regard to future work, most of the presented methods in this thesis were applied retrospectively, i.e. a comprehensive integration into a future workflow is still pending and much development work is still required. This of course applies in particular to the certification processes regarding medically used software. Besides, with learning-based methods still on the rise, some of the presented approaches of this thesis could further benefit from deep learning techniques. For instance, the integration of computerised detection approaches could make the proposed segmentation methods fully automatic or could overcome manual user initialisation of the pre- and intra-operative image fusion. Moreover, a deep learning-based joint framework combining segmentation and registration could elastically register multimodal images considering pre-defined rigidities of multiple segmented anatomical structures (Estienne et al., 2019). Nonetheless, common limitations regarding such strategies will remain challenging, for instance, the limited amount of clinical data or the time-consuming preparation of labelled reference data. However, active learning strategies indicate promising results in order to identify the most informative data samples and could therefore, reduce the amount of required labelled data to ensure promising training.

In conclusion, all of the proposed approaches were in particular tailored for clinical needs and thus were able to meet the defined objectives of their specific tasks. Using real patient data and ground truth annotations from radiologists and trained field experts for evaluation purposes, each method proved its suitability and applicability to clinical data. Existing gaps and limitations of the state of the art were appropriately identified and elaborated, while the relevance and novelty of the presented approaches have been proven by publications in recognised journals as well as in proceedings of national and international conferences. Most of the presented approaches could be adapted to similar application fields and interventional treatment procedures, e.g. for lung or liver lesions as well as osseous metastases apart of the spine, which makes them relevant beyond the particular underlying subject matter of this thesis.

APPENDIX

A

Table A.1: Complete evaluation results of the proposed BC-HLS approach (Part 1 of 2). This part refers to the pre-RFA and SHIP datasets. Since a second ground truth was available for each case, the IRV could be computed.

	DSC [%]		ASD [mm]		HD [mm]	
dataset	BC-HLS	IRV	BC-HLS	IRV	BC-HLS	IRV
preRFA_1	78.4	87.6	1.3	0.3	7.5	2.6
preRFA_2	81.9	87.6	1.4	0.4	7.7	4.0
preRFA_3	86.2	86.7	1.3	0.6	7.6	5.3
preRFA_4	85.0	89.3	1.1	0.3	7.0	4.4
preRFA_5	83.7	90.0	1.3	0.3	9.0	3.7
preRFA_6	84.4	91.3	1.5	0.3	8.6	3.4
mean $\pm$ std	83.3 $\pm$ 2.8	88.7 $\pm$ 1.7	1.3 $\pm$ 0.1	0.4 $\pm$ 0.1	7.9 $\pm$ 0.7	3.9 $\pm$ 0.9
SHIP_1_T_1	85.7	88.4	1.5	1.1	7.2	5.6
SHIP_1_T_2	84.0	-	1.7	-	7.9	-
SHIP_2_T_1	84.6	86.9	1.5	1.2	6.7	5.6
SHIP_2_T_2	81.9	-	1.8	-	7.7	-
SHIP_3_T_1	84.9	88.2	1.5	1.1	7.1	5.2
SHIP_3_T_2	83.6	-	1.7	-	7.8	-
SHIP_4_T_1	84.4	89.6	1.5	0.9	6.4	4.4
SHIP_4_T_2	83.1	-	1.7	-	6.9	-
SHIP_5_T_1	85.2	85.6	1.7	1.5	8.1	5.5
SHIP_5_T_2	83.7	-	1.9	-	8.6	-
SHIP_6_T_1	87.3	90.9	1.5	1.0	7.3	5.0
SHIP_6_T_2	86.1	-	1.6	-	7.1	-
SHIP_7_T_1	81.0	87.3	2.0	1.2	8.9	6.3
SHIP_7_T_2	80.7	-	2.0	-	8.6	-
SHIP_8_T_1	82.9	88.5	1.7	1.1	7.5	5.7
SHIP_8_T_2	81.5	-	1.9	-	8.0	-
SHIP_9_T_1	85.4	87.3	1.6	1.3	6.9	5.9
SHIP_9_T_2	84.9	-	1.7	-	7.2	-
mean $\pm$ std	83.9 $\pm$ 1.8	88.1 $\pm$ 1.5	1.7 $\pm$ 0.2	1.1 $\pm$ 0.2	7.6 $\pm$ 0.7	5.5 $\pm$ 0.5

Table A.2: Complete evaluation results of the proposed BC-HLS approach (Part 2 of 2). This part refers to the publicly available data and corresponding results of Zukić et al., 2014 and Chu et al., 2015.

	DSC [%]		ASD [mm]		HD [mm]	
dataset	BC-HLS	Zukić et al.	BC-HLS	Zukić et al.	BC-HLS	Zukić et al.
AKa2	88.1	82.3	1.6	1.7	7.7	10.8
Aka3	83.2	81.6	1.9	1.6	8.8	10.8
Aka4	85.7	83.2	1.7	1.6	6.9	9.3
Aks5	86.4	84.2	1.4	1.4	6.5	10.8
Aks6	83.1	80.6	1.6	1.8	8.0	10.6
Aks7	84.8	78.0	1.5	2.0	6.6	11.3
Aks8	85.1	83.7	1.5	1.5	7.5	9.1
C002	77.9	74.5	1.8	1.8	7.8	11.8
DzZ_T2	86.4	84.3	1.6	1.8	8.5	12.0
DzZ_T1	87.8	78.9	1.3	2.0	7.2	13.6
F02	80.2	74.2	2.0	2.0	8.6	15.7
F03	84.0	76.1	1.9	2.2	8.6	14.1
F04	84.7	72.6	1.5	2.2	6.0	11.4
S01	83.6	76.7	2.1	2.1	9.8	19.6
S02	82.5	79.7	1.6	1.4	7.6	9.7
St1	82.8	79.9	2.1	1.9	10.3	17.2
mean $\pm$ std	84.1 $\pm$ 2.5	79.4 $\pm$ 3.8	1.7 $\pm$ 0.2	1.8 $\pm$ 0.3	7.9 $\pm$ 1.1	12.4 $\pm$ 3.0
dataset	BC-HLS	Chu et al.	BC-HLS	Chu et al.	BC-HLS	Chu et al.
#1	89.6	88.6	1.6	1.7	5.7	6.5
#2	89.1	86.9	1.3	1.6	5.9	8.3
#3	89.5	86.7	1.5	1.8	6.1	8.1
#4	85.4	85.1	1.9	1.9	7.0	8.3
#5	89.6	88.4	1.5	1.5	5.5	6.7
#6	86.6	89.2	1.6	1.5	6.3	6.2
#7	88.3	88.2	1.7	1.6	5.8	7.3
#8	87.6	87.1	1.9	1.6	6.2	6.2
#9	87.7	88.3	1.6	1.6	5.4	5.9
#10	87.7	89.5	1.4	1.4	5.3	5.1
#11	87.0	89.5	1.8	1.4	6.1	7.8
#12	88.6	89.9	1.7	1.4	5.8	5.6
#13	86.7	86.6	1.7	1.5	5.2	6.1
#14	87.8	88.5	1.5	1.6	5.6	5.5
#15	87.7	88.0	1.8	1.7	6.9	7.0
#16	86.7	87.3	2.0	1.6	6.2	6.9
#17	90.2	91.8	1.7	1.3	5.3	6.1
#18	87.5	91.0	1.8	1.3	6.5	4.6
#19	87.7	87.3	1.7	1.5	5.7	7.4
#20	90.0	91.2	1.5	1.3	5.7	5.0
#21	86.7	87.9	1.8	1.5	6.9	6.6
#22	89.6	91.9	1.7	1.3	6.0	4.3
#23	89.8	91.7	1.5	1.3	5.1	4.7
mean $\pm$ std	88.2 $\pm$ 1.9	88.7 $\pm$ 2.9	1.7 $\pm$ 0.3	1.5 $\pm$ 0.2	6.0 $\pm$ 1.0	6.4 $\pm$ 1.2

Table A.3: Preliminary study on the effect of different learning rates (l_r) and optimizers (Adam, SGD - stochastic gradient descent) on the quality of spinal metastasis segmentation. DSC scores were computed over all cross-validation folds for the 2D multimodal image input.

		Adam				SGD	
	l_r	0.1	0.01	0.001	0.0001	0.1	0.01
DSC [%]	mean	74.92	76.06	76.02	73.18	73.89	75.75
	std	13.40	13.17	14.45	16.48	14.02	15.04

Table A.4: Registration accuracy of the proposed MS-VB registration approach for pre- and intra-interventional images of spinal metastases. For each patient case the registration accuracy (fiducial registration error of the globally rigid approach (FRE_{GR}), fiducial registration error of the multisegmental approach (FRE_{MS}), the fiducial registration error of a manual registration (mFRE)) and the specific spatial resolution (internal voxel diagonals) for the MRI (S_{MRI}) and FP-CT ($S_{FP\text{-}CT}$) scans are stated. Values marked with an asterisk symbol denote, that the used pre-interventional images were T_2-weighted MRI sequences instead of the otherwise used native T_1- or contrast-enhanced T_1-weighted data.

	patient case																			Ø	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	mean	std
FRE_{GR} [mm]	7.67	1.89	3.57	2.64	5.37	8.07	3.07	2.35	2.49 3.28*	5.23	2.77	2.92	4.58 5.83*	4.13 3.96*	2.60	3.16	2.64 2.30*	3.37 3.96*	3.46	3.79 3.87*	1.72 1.29*
FRE_{MS} [mm]	4.43	2.08	3.12	2.02	3.03	3.00	2.86	1.47	2.16 2.67*	2.11	2.59	1.78	1.84 2.11*	2.28 2.79*	1.63	2.28	2.06 2.25*	1.83 2.93*	1.96	2.34 2.55*	0.70 0.35*
mFRE [mm]	2.37	1.41	1.98	0.88	1.53	1.28	1.74	1.32	0.97	4.66	2.46	1.06	0.98	1.38	2.15	1.60	1.62	1.67	1.24	1.70	0.85
S_{MRI} [mm]	3.39	3.34	3.33	3.34	3.37	3.34	3.37	3.37	4.43	3.81	4.43	3.37	3.34	3.34	3.39	3.39	3.37	3.35	3.34	3.49	0.35
$S_{FP\text{-}CT}$ [mm]	0.66	0.66	0.66	0.66	0.66	0.66	0.69	0.69	0.68	0.68	0.68	1.05	0.61	1.06	1.27	1.27	1.23	1.23	0.66	0.83	0.26

Algra, P. R., Heimans, J. J., Valk, J., Nauta, J., Lachniet, M., and Van Kooten, B. (1992). "Do metastases in vertebrae begin in the body or the pedicles? Imaging study in 45 patients." In: *American Journal of Roentgenology* 158.6, pp. 1275–1279.

Athertya, J., Kumar, G. S., et al. (2016). "Fuzzy clustering based segmentation of vertebrae in T1-weighted spinal MR images." In: *International Journal of Fuzzy Logic and Intelligent Systems* 6, pp. 23–34.

Ayed, I. B., Punithakumar, K., Minhas, R., Joshi, R., and Garvin, G. J. (2012). "Vertebral body segmentation in MRI via convex relaxation and distribution matching." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 520–527.

Baegert, C., Villard, C., Schreck, P., Soler, L., and Gangi, A. (2007). "Trajectory optimization for the planning of percutaneous radiofrequency ablation of hepatic tumors." In: *Computer Aided Surgery* 12.2, pp. 82–90.

Barzilai, O., Fisher, C. G., and Bilsky, M. H. (2018). "State of the art treatment of spinal metastatic disease." In: *Neurosurgery* 82.6, pp. 757–769.

Bauer, S., Wiest, R., Nolte, L.-P., and Reyes, M. (2013). "A survey of MRI-based medical image analysis for brain tumor studies." In: *Physics in Medicine and Biology* 58.13, pp. 97–129.

Bricault, I., Kikinis, R., Morrison, P. R., vansonnenberg, E., Tuncali, K., and Silverman, S. G. (2006). "Liver metastases: 3D shape-based analysis of CT scans for detection of local recurrence after radiofrequency ablation 1." In: *Radiology* 241, pp. 243–250.

Brinjikji, W., Luetmer, P. H., Comstock, B., Bresnahan, B. W., Chen, L., et al. (2015). "Systematic literature review of imaging features of spinal degeneration in asymptomatic populations." In: *American Journal of Neuroradiology* 36.4, pp. 811–816.

Carrafiello, G., Laganà, D., Ianniello, A., Dionigi, G., Novario, R., et al. (2007). "Post-radiofrequency ablation syndrome after percutaneous radiofrequency of abdominal tumours: One centre experience and review of published works." In: *Australasian radiology* 51.6, pp. 550–554.

Caselles, V., Kimmel, R., and Sapiro, G. (1997). "Geodesic active contours." In: *International Journal of Computer Vision* 22.1, pp. 61–79.

Čech, P., Andronache, A., Wang, L., Székely, G., and Cattin, P. (2006). "Piecewise rigid multimodal spine registration." In: *International Conference on Bildverarbeitung für die Medizin (BVM)*, pp. 211–215.

Charles, C. Y., Yuh, R. T., Bajwa, N. S., Toy, J. O., Ahn, U. M., and Ahn, N. U. (2015). "Lower thoracic pedicle morphometry: male, taller, and heavier specimens have bigger pedicles." In: *Spine* 40.6, pp. 323–331.

Chen, H., Shen, C., Qin, J., Ni, D., Shi, L., Cheng, J. C., and Heng, P.-A. (2015). "Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks." In: *International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI)*, pp. 515–522.

Chen, Y.-R., Deb, S., Pham, L., and Singh, H. (2016). "Minimally invasive lumbar pedicle screw fixation using cortical bone trajectory–a prospective cohort study on postoperative pain outcomes." In: *Cureus* 8.7.

Chen, Y.-T. and Wang, M.-S. (2004). "Three-dimensional reconstruction and fusion for multi-modality spinal images." In: *Computerized Medical Imaging and Graphics* 28.1-2, pp. 21–31.

Chmelik, J., Jakubicek, R., Walek, P., et al. (2018). "Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data." In: *Medical image analysis* 49, pp. 76–88.

Cho, J., Park, Y. G., and Chung, S. S. (1997). "Percutaneous radiofrequency lumbar facet rhizotomy in mechanical low back pain syndrome." In: *Stereotactic and Functional Neurosurgery* 68, pp. 212–217.

Christ, P. F., Ettlinger, F., Grün, F., et al. (2017). "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks." In: *arXiv:1702.05970*.

Christodoulou, A. G., Apostolou, T., Ploumis, A., Terzidis, I., Hantzokos, I., and Pournaras, J. (2005). "Pedicle dimensions of the thoracic and lumbar vertebrae in the Greek population." In: *Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists* 18.6, pp. 404–408.

Chu, C., Belavỳ, D. L., Armbrecht, G., Bansmann, M., Felsenberg, D., and Zheng, G. (2015). "Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method." In: *PloS one* 10, e0143327.

Chu, K. F. and Dupuy, D. E. (2014). "Thermal ablation of tumours: biological mechanisms and advances in therapy." In: *Nature Reviews Cancer* 14.3, pp. 199–208.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). "Active shape models-their training and application." In: *Computer Vision and Image Understanding* 61.1, pp. 38–59.

Dalah, E. Z., Nisbet, A., Reise, S., and Bradley, D. (2008). "Evaluating commercial image registration packages for radiotherapy treatment planning." In: *Applied Radiation and Isotopes* 66.12, pp. 1948–1953.

Darwish, A. A., Salem, M. A.-M., Hegazy, D., and Ebeid, H. M. (2015). "Vertebrae segmentation techniques for spinal medical images." In: *International Conference on Information, Communications and Signal processing (ICICIS)*. IEEE, pp. 110–115.

Davatzikos, C., Liu, D., Shen, D., and Herskovits, E. H. (2002). "Spatial normalization of spine MR images for statistical correlation of lesions with clinical symptoms 1." In: *Radiology* 224.3, pp. 919–926.

De Leener, B., Cohen-Adad, J., and Kadoury, S. (2015). "Automatic segmentation of the spinal cord and spinal canal coupled with vertebral labeling." In: *IEEE Transactions on Medical Imaging* 34.8, pp. 1705–1718.

Doerfler, A., Gölitz, P., Engelhorn, T., Kloska, S., and Struffert, T. (2015). "Flat-panel computed tomography (DYNA-CT) in neuroradiology. from high-resolution imaging of implants to one-stop-shopping for acute stroke." In: *Clinical Neuroradiology* 25.2, pp. 291–297.

Dupuy, D. E., Liu, D., Hartfeil, D., Hanna, L., Blume, J. D., et al. (2010). "Percutaneous radiofrequency ablation of painful osseous metastases." In: *Cancer* 116, pp. 989–997.

Egger, J., Busse, H., Brandmaier, P., Seider, D., Gawlitza, M., et al. (2015). "RFA-cut: semi-automatic segmentation of radiofrequency ablation zones with and without needles via optimal st-cuts." In: *International Conference on Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 2423–2429.

Eleraky, M., Papanastassiou, I., and Vrionis, F. D. (2010). "Management of metastatic spine disease." In: *Current Opinion in Supportive and Palliative Care* 4.3, pp. 182–188.

Estienne, T., Vakalopoulou, M., Christodoulidis, S., Battistela, E., Lerousseau, M., et al. (2019). "U-ReSNet: ultimate coupling of registration and segmentation with deep nets." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 310–319.

Evans, L. T., Vega, R. A., and Tatsui, C. E. (2020). "Spinal laser interstitial thermal therapy for metastatic tumors." In: *Central Nervous System Metastases*, pp. 623–634.

Even-Sapir, E. (2005). "Imaging of malignant bone involvement by morphologic, scintigraphic, and hybrid modalities." In: *Journal of Nuclear Medicine* 46.8, pp. 1356–1367.

Frangou, E. and Fourney, D. R. (2009). "Minimally invasive treatment of spinal tumors." In: *International Conference on Seminars in Spine Surgery*. Vol. 21. 2, pp. 112–120.

Fujioka, C., Horiguchi, J., Ishifuro, M., Kakizawa, H., Kiguchi, M., et al. (2006). "A feasibility study: evaluation of radiofrequency ablation therapy to hepatocellular carcinoma using image registration of preoperative and postoperative CT." In: *Academic Radiology* 13.8, pp. 986–994.

Geets, X., Daisne, J.-F., Arcangeli, S., Coche, E., De Poel, M., Duprez, T., Nardella, G., and Grégoire, V. (2005). "Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI." In: *Radiotherapy and Oncology* 77, pp. 25–31.

Georgy, B. (2008). "Metastatic spinal lesions: state-of-the-art treatment options and future trends." In: *American Journal of Neuroradiology* 29.9, pp. 1605–1611.

Ghosh, S., Malgireddy, M. R., Chaudhary, V., and Dhillon, G. (2014). "A supervised approach towards segmentation of clinical MRI for automatic lumbar diagnosis." In: *International Conference on Medical*

Image Computing and Computer-Assisted Intervention (MICCAI). Vol. 17, pp. 185–195.

Gilbert, R. W., Kim, J.-H., and Posner, J. B. (1978). "Epidural spinal cord compression from metastatic tumor: diagnosis and treatment." In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 3.1, pp. 40–51.

Goankar, B., Xia, Y., Villaroman, D., Ko, A., Attiah, M., Beckett, J., and Macyszyn, L. (2017). "Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images." In: *IEEE Journal of Translational Engineering in Health and Medicine* 5, pp. 1–12.

Goldberg, S. N. and Gazelle, G. S. (2001). "Radiofrequency tissue ablation: physical principles and techniques for increasing coagulation necrosis." In: *Hepato-gastroenterology* 48.38, pp. 359–367.

Gower, J. C. (1975). "Generalized procrustes analysis." In: *Psychometrika* 40.1, pp. 33–51.

Greenberg, A., Berenstein Weyel, T., Sosna, J., Applbaum, J., and Peyser, A. (2014). "The distribution of heat in bone during radiofrequency ablation of an ex vivo bovine model of osteoid osteoma." In: *The Bone & Joint Journal* 96.5, pp. 677–683.

Greenwood, T. J., Wallace, A., Friedman, M. V., Hillen, T. J., Robinson, C. G., and Jennings, J. W. (2015). "Combined ablation and radiation therapy of spinal metastases: a novel multimodality treatment approach." In: *Pain Physician* 18.6, pp. 573–581.

Guillevin, R., Vallee, J.-N., Lafitte, F., Menuel, C., Duverneuil, N.-M., and Chiras, J. (2007). "Spine metastasis imaging: review of the literature." In: *Journal of Neuroradiology* 34.5, pp. 311–321.

Haber, E. and Modersitzki, J. (2006). "Intensity gradient based registration and fusion of multi-modal images." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 726–733.

Halvorson, K. G., Sevcik, M. A., Ghilardi, J. R., Rosol, T. J., and Mantyh, P. W. (2006). "Similarities and differences in tumor growth, skeletal remodeling and pain in an osteolytic and osteoblastic model of bone cancer." In: *The Clinical Journal of Pain* 22.7, pp. 587–600.

Hammernik, K., Ebner, T., Stern, D., Urschler, M., and Pock, T. (2015). "Vertebrae segmentation in 3D CT images based on a variational framework." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 227–233.

Harrington, K. (1986). "Metastatic disease of the spine." In: *The Journal of Bone & Joint Surgery* 68, pp. 1110–1115.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., et al. (2017). "Brain tumor segmentation with deep neural networks." In: *Medical Image Analysis* 35, pp. 18–31.

Hille, G., Becker, M., Saalfeld, S., and Tönnies, K. (2019). "Treatment outcome validation tool for radiofrequency ablations of spinal metastases." In: *International Conference on Computer- und Roboterassistierte Chirurgie (CURAC)*, pp. 134–139.

Hille, G., Merten, N., Serowy, S., Glaßer, S., Tönnies, K., and Preim, B. (2017). "Assessing the benefits of interactive patient-specific visualisations for patient information." In: *International Conference on Bildverarbeitung für die Medizin (BVM)*, pp. 224–229.

Hille, G., Saalfeld, S., Serowy, S., and Tönnies, K. (2018a). "Multi-segmental spine image registration supporting image-guided interventions of spinal metastases." In: *Computers in Biology and Medicine* 102, pp. 16–20.

Hille, G., Saalfeld, S., Serowy, S., and Tönnies, K. (2018b). "Vertebral body segmentation in wide range clinical routine spine MRI data." In: *Computer Methods and Programs in Biomedicine* 155, pp. 93–99.

Hille, G., Steffen, J., Dünnwald, M., Becker, M., Saalfeld, S., and Tönnies, K. (2020). "Spinal metastases segmentation in MR imaging using deep convolutional neural networks." In: *arXiv preprint arXiv:2001.05834*.

Hirooka, M., Koizumi, Y., Imai, Y., Ochi, H., Nakamura, Y., et al. (2016). "Assessment of a needle-tracking system for bipolar radiofrequency ablation." In: *Journal of Medical Ultrasonics* 43.2, pp. 185–191.

Hong, K. and Georgiades, C. (2010). "Radiofrequency ablation: mechanism of action and devices." In: *Journal of Vascular and Interventional Radiology* 21.8, pp. 179–186.

Horn, B. K. (1987). "Closed-form solution of absolute orientation using unit quaternions." In: *Journal of the Optical Society of America A* 4.4, pp. 629–642.

Hu, Y. and Haynor, D. R. (2004). "Multirigid registration of MR and CT images of the cervical spine." In: *International Conference on SPIE Medical Imaging*. Vol. 5370. International Society for Optics and Photonics, pp. 1527–1539.

Huang, S.-H., Chu, Y.-H., Lai, S.-H., and Novak, C. L. (2009). "Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI." In: *IEEE Transactions on Medical Imaging* 28.10, pp. 1595–1605.

Huo, M., Sahgal, A., Pryor, D., Redmond, K., Lo, S., and Foote, M. (2017). "Stereotactic spine radiosurgery: Review of safety and efficacy with respect to dose and fractionation." In: *Surgical Neurology International* 8.

Isensee, F., Kickingereder, P., Bonekamp, D., Bendszus, M., Wick, W., Schlemmer, H.-P., and Maier-Hein, K. (2017). "Brain tumor segmentation using large receptive field deep convolutional neural networks." In: *International Conference on Bildverarbeitung für die Medizin (BVM)*, pp. 86–91.

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. H. (2018). "No new-net." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–244.

Jerebko, A., Schmidt, G., Zhou, X., Bi, J., Anand, V., et al. (2007). "Robust parametric modeling approach based on domain knowledge for computer aided detection of vertebrae column metastases in MRI."

In: *International Conference on Information Processing in Medical Imaging (IPMI)*. Springer, pp. 713–724.

Kadoury, S., Labelle, H., and Paragios, N. (2013). "Spine segmentation in medical images using manifold embeddings and higher-order MRFs." In: *IEEE Transactions on Medical Imaging* 32.7, pp. 1227–1238.

Kaminsky, J., Rodt, T., Zajaczek, J., Donnerstag, F., and Zumkeller, M. (2004). "Mehrsegmentale Bildfusion an der Wirbelsäule/Multisegmental Image Fusion of the Spine." In: *Biomedizinische Technik/Biomedical Engineering* 49.3, pp. 49–55.

Karlo, C., Steurer-Dober, I., Leonardi, M., Pfirrmann, C., Zanetti, M., and Hodler, J. (2010). "MR/CT image fusion of the spine after spondylodesis: a feasibility study." In: *European Spine Journal* 19.10, pp. 1771–1775.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). "Snakes: active contour models." In: *International Journal of Computer Vision* 1.4, pp. 321–331.

Kavakebi, P., Freyschlag, C., and Thomé, C. (2017). "How I do it: optimizing radiofrequency ablation in spinal metastases using iCT and navigation." In: *Acta Neurochirurgica* 159.10, pp. 2025–2028.

Keil, S., Bruners, P., Ohnsorge, L., Plumhans, C., Behrendt, F. F., et al. (2010). "Semiautomated versus manual evaluation of liver metastases treated by radiofrequency ablation." In: *Journal of Vascular and Interventional Radiology* 21.2, pp. 245–251.

Khairy, P., Chauvet, P., Lehmann, J., Lambert, J., Macle, L., et al. (2003). "Lower incidence of thrombus formation with cryoenergy versus radiofrequency catheter ablation." In: *Circulation* 107.15, pp. 2045–2050.

Kim, J. H., Hahn, E. W., and Antich, P. P. (1982). "Radiofrequency hyperthermia for clinical cancer therapy." In: *National Cancer Institute Monographs* 61, pp. 339–342.

Kim, Y. H., Choi, B. I., Cho, W. H., Lim, S., Moon, W. K., Han, J. K., Weinmann, H.-J., and Chang, K.-H. (2003). "Dynamic contrast-enhanced MR imaging of VX2 carcinomas after X-irradiation in rabbits: comparison of gadopentetate dimeglumine and a macromolecular contrast agent." In: *Investigative Radiology* 38.9, pp. 539–549.

Kingma, D. P. and Ba, J. (2014). "Adam: A method for stochastic optimization." In: *arXiv:1412.6980*.

Klimo, P. and Schmidt, M. H. (2004). "Surgical management of spinal metastases." In: *The Oncologist* 9, pp. 188–196.

Kline, M. (2000). "Radiofrequency techniques in clinical practice." In: *Interventional Pain Management*.

König, L., Derksen, A., Papenberg, N., and Haas, B. (2016). "Deformable image registration for adaptive radiotherapy with guaranteed local rigidity constraints." In: *Radiation Oncology* 11.1, p. 122.

Korez, R., Likar, B., Pernuš, F., and Vrtovec, T. (2016). "Model-based segmentation of vertebral bodies from MR images with 3D CNNs." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCA)*. Springer, pp. 433–441.

Kothe, R., Strauss, J. M., Deuretzbacher, G., Hemmi, T., Lorenzen, M., and Wiesner, L. (2001). "Computer navigation of parapedicular screw fixation in the thoracic spine: a cadaver study." In: *Spine* 26.21, E496–E501.

Kröger, T., Pätz, T., Altrogge, I., Schenk, A., Lehmann, K., et al. (2010). "Fast estimation of the vascular cooling in RFA based on numerical simulation." In: *The Open Biomedical Engineering Journal* 4.

Kröger, T., Altrogge, I., Preusser, T., Pereira, P., Schmidt, D., Weihusen, A., and Peitgen, H.-O. (2006). "Numerical simulation of radio frequency ablation with state dependent material parameters in three space dimensions." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 380–388.

Laura, C. O., Drechsler, K., Wesarg, S., and Bale, R. (2016). "Accurate physics-based registration for the outcome validation of minimal invasive interventions and open liver surgeries." In: *IEEE Transactions on Biomedical Engineering* 64.2, pp. 362–371.

Lessmann, N., van Ginneken, B., de Jong, P. A., and Išgum, I. (2019). "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification." In: *Medical Image Analysis* 53, pp. 142–155.

Li, X., Chen, H., Qi, X., et al. (2017). "H-DenseUNet: hybrid densely connected UNet for liver and liver tumor segmentation from CT volumes." In: *arXiv:1709.07330*.

Lipton, A. (2004). "Pathophysiology of bone metastases: how this knowledge may lead to therapeutic intervention." In: *The Journal of Supportive Oncology* 2.3, pp. 205–13.

Liu, J., Napolitano, J. T., and Ebraheim, N. A. (2010). "Systematic review of cervical pedicle dimensions and projections." In: *Spine* 35.24, E1373–E1380.

Liu, J., Li, M., Wang, J., Wu, F., Liu, T., and Pan, Y. (2014). "A survey of MRI-based brain tumor segmentation methods." In: *Tsinghua Science and Technology* 19.6, pp. 578–595.

Lyksborg, M., Puonti, O., Agn, M., and Larsen, R. (2015). "An ensemble of 2D convolutional neural networks for tumor segmentation." In: *Scandinavian Conference on Image Analysis (SCIA)*, pp. 201–211.

Maccauro, G., Spinelli, M. S., Mauro, S., Perisano, C., Graci, C., and Rosa, M. A. (2011). "Physiopathology of spine metastasis." In: *International Journal of Surgical Oncology* 2011.

Maintz, J. A. and Viergever, M. A. (1998). "A survey of medical image registration." In: *Medical Image Analysis* 2.1, pp. 1–36.

Matschek, J., Bullinger, E., von Haeseler, F., Skalej, M., and Findeisen, R. (2017). "Mathematical 3D modelling and sensitivity analysis of multipolar radiofrequency ablation in the spine." In: *Mathematical Biosciences* 284, pp. 51–60.

Mauri, G., Cova, L., De Beni, S., Ierace, T., Tondolo, T., Cerri, A., Goldberg, S. N., and Solbiati, L. (2015). "Real-time US-CT/MRI image fusion for guidance of thermal ablation of liver tumors undetectable with

US: results in 295 cases." In: *Cardiovascular and Interventional Radiology* 38.1, pp. 143–151.

McCreedy, E. S., Cheng, R., Hemler, P. F., Viswanathan, A., Wood, B. J., and McAuliffe, M. J. (2006). "Radio frequency ablation registration, segmentation, and fusion tool." In: *IEEE Transactions on Information Technology in Biomedicine* 10, pp. 490–496.

Merten, N., Adler, S., Hille, G., Hanses, M., Becker, M., Saalfeld, S., and Preim, B. (2019). "A two-step risk assessment method for radiofrequency ablations of spine metastases." In: *Computers in biology and medicine* 108, pp. 174–181.

Miles, B., Ayed, I. B., Law, M. W., Garvin, G., Fenster, A., and Li, S. (2013). "Spine image fusion via graph cuts." In: *IEEE Transactions on Biomedical Engineering* 60.7, pp. 1841–1850.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-net: Fully convolutional neural networks for volumetric medical image segmentation." In: *International Conference on 3D Vision (3DV)*, pp. 565–571.

Nemec, S. F., Donat, M. A., Mehrain, S., Friedrich, K., Krestan, C., Matula, C., Imhof, H., and Czerny, C. (2007). "CT–MR image data fusion for computer assisted navigated neurosurgery of temporal bone tumors." In: *European Journal of Radiology* 62.2, pp. 192–198.

Neubert, A., Fripp, J., Engstrom, C., Schwarz, R., Lauer, L., Salvado, O., and Crozier, S. (2012). "Automated detection, 3D segmentation and analysis of high resolution spine MR images using statistical shape models." In: *Physics in Medicine & Biology* 57.24, p. 8357.

Neubert, A., Fripp, J., Shen, K., Salvado, O., Schwarz, R., Lauer, L., Engstrom, C., and Crozier, S. (2011). "Automated 3D segmentation of vertebral bodies and intervertebral discs from MRI." In: *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 19–24.

Ni, Y., Mulier, S., Miao, Y., Michel, L., and Marchal, G. (2005). "A review of the general aspects of radiofrequency ablation." In: *Abdominal Imaging* 30.4, pp. 381–400.

Palussiere, J., Pellerin-Guignard, A., Descat, E., Cornélis, F., and Dixmérias, F. (2012). "Radiofrequency ablation of bone tumours." In: *Diagnostic and Interventional Imaging* 93.9, pp. 680–684.

Panjabi, M. M. (2003). "Clinical spinal instability and low back pain." In: *Journal of Electromyography and Kinesiology* 13.4, pp. 371–379.

Parizel, P., Van der Zijden, T., Gaudino, S., Spaepen, M., Voormolen, M., et al. (2010). "Trauma of the spine and spinal cord: imaging strategies." In: *European Spine Journal* 19.1, pp. 8–17.

Passera, K., Selvaggi, S., Scaramuzza, D., Garbagnati, F., Vergnaghi, D., and Mainardi, L. (2013). "Radiofrequency ablation of liver tumors: quantitative assessment of tumor coverage through CT image processing." In: *BMC Medical Imaging* 13, p. 3.

Posteraro, A., Dupuy, D., and Mayo-Smith, W. (2004). "Radiofrequency ablation of bony metastatic disease." In: *Clinical Radiology* 59.9, pp. 803–811.

Prados, F., Cardoso, M. J., Yiannakas, M. C., Hoy, L. R., Tebaldi, E., et al. (2016). "Fully automated grey and white matter spinal cord segmentation." In: *Scientific Reports* 6, p. 36151.

Rak, M., Steffen, J., Meyer, A., Hansen, C., and Tönnies, K.-.-D. (2019). "Combining convolutional neural networks and star convex cuts for fast whole spine vertebra segmentation in MRI." In: *Computer Methods and Programs in Biomedicine* 177, pp. 47–56.

Rak, M. and Tönnies, K. D. (2017). "Star convex cuts with encoding swaps for fast whole-spine vertebra segmentation in MRI." In: *Interantional Conference on Vision, Modeling and Visualization (VMV)*, pp. 145–152.

Rak, M. and Tönnies, K.-D. (2016). "A learning-free approach to whole spine vertebra localization in MRI." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 283–290.

Rashad, A., Heiland, M., Hiepe, P., Nasirpour, A., Rendenbach, C., Keuchel, J., Regier, M., and Al-Dam, A. (2019). "Evaluation of a novel elastic registration algorithm for spinal imaging data: A pilot clinical study." In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 15.3, e1991.

Research, T. M. (2019). *Radiofrequency ablation devices market for pain management (product - RF generators, reusable products (probes and electrodes), and disposable products (cannulas and needles); end-user - specialty clinics, ambulatory surgical centers, and hospitals) - global industry analysis, size, share, growth, trends, and forecast 2019 - 2027*. https://www.transparencymarketresearch.com/radiofrequency-ablation-devices.html. Visited on 2020-04-05.

Rieder, C., Geisler, B., Bruners, P., Isfort, P., Na, H.-S., Mahnken, A. H., and Hahn, H. K. (2014). "Software-assisted post-interventional assessment of radiofrequency ablation." In: *Interantional Conference on Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 9036, p. 903604.

Rieder, C., Weihusen, A., Schumann, C., Zidowitz, S., and Peitgen, H.-O. (2010). "Visual support for interactive post-interventional assessment of radiofrequency ablation therapy." In: *Interantional Conference on Computer Graphics Forum (CGF)*. Vol. 29. 3, pp. 1093–1102.

Rieder, C., Wirtz, S., Strehlow, J., Zidowitz, S., Bruners, P., Isfort, P., Mahnken, A. H., and Peitgen, H.-O. (2012). "Automatic alignment of pre-and post-interventional liver CT images for assessment of radiofrequency ablation." In: *International Conference on SPIE Medical Imaging*. Vol. 8316. International Society for Optics and Photonics, 83163E.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, pp. 234–241.

Rosenthal, D. and Callstrom, M. R. (2012). "Critical review and state of the art in interventional oncology: benign and metastatic disease involving bone." In: *Radiology* 262, pp. 765–780.

Rosenthal, D. I., Hornicek, F. J., Wolfe, M. W., Jennings, L. C., Gebhardt, M. C., and Mankin, H. J. (1998). "Percutaneous radiofrequency coagulation of osteoid osteoma compared with operative treatment." In: *The Journal of Bone & Joint Surgery* 80, pp. 815–21.

Roth, H. R., Yao, J., Lu, L., Stieger, J., Burns, J. E., and Summers, R. M. (2015). "Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications." In: *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, pp. 3–12.

Rothrock, R. J., Barzilai, O., Bilsky, M. H., and Laufer, I. (2020). "Minimally invasive surgery for spinal metastases." In: *Central Nervous System Metastases*, pp. 575–582.

Salehi, S. S. M., Erdogmus, D., and Gholipour, A. (2017). "Tversky loss function for image segmentation using 3D fully convolutional deep networks." In: *Internetional Conference on Machine Learning in Medical Imaging and Analysis (ICMLIMIA)*, pp. 379–387.

Schwarzenberg, R., Freisleben, B., Nimsky, C., and Egger, J. (2014). "Cube-cut: vertebral body segmentation in MRI-data through cubic-shaped divergences." In: *PloS one* 9, e93389.

Scoles, P. V., Linton, A. E., Latimer, B., Levy, M. E., and Digiovanni, B. F. (1988). "Vertebral body and posterior element morphology: the normal spine in middle life." In: *Spine* 13.10, pp. 1082–1086.

Seitel, A., Engel, M., Sommer, C. M., Radeleff, B. A., Essert-Villard, C., et al. (2011). "Computer-assisted trajectory planning for percutaneous needle insertions." In: *Medical Physics* 38.6Part1, pp. 3246–3259.

Shah, L. M. and Salzman, K. L. (2011). "Imaging of spinal metastatic disease." In: *International Journal of Surgical Oncology* 2011.

Shaikh, M., Anand, G., Acharya, G., Amrutkar, A., Alex, V., and Krishnamurthi, G. (2017). "Brain tumor segmentation using dense fully convolutional neural network." In: *International Workshop on Brainlesion (BL-MICCAI)*, pp. 309–319.

Shorten, C. and Khoshgoftaar, T. M. (2019). "A survey on image data augmentation for deep learning." In: *Journal of Big Data* 6.1, p. 60.

Skanes, A. C., Klein, G., Krahn, A., and Yee, R. (2004). "Cryoablation: potentials and pitfalls." In: *Journal of Cardiovascular Electrophysiology* 15, S28–S34.

Smistad, E., Falch, T. L., Bozorgi, M., Elster, A. C., and Lindseth, F. (2015). "Medical image segmentation on GPUs–A comprehensive review." In: *Medical Image Analysis* 20.1, pp. 1–18.

Society, A. C. (2019). *Cancer facts & figures 2019. American Cancer Society (ACS) Atlanta, GA: American Cancer Society*.

Sohn, M.-J., Lee, D.-J., Yoon, S. W., Lee, H. R., and Hwang, Y. J. (2009). "The effective application of segmental image fusion in spinal radiosurgery for improved targeting of spinal tumours." In: *Acta Neurochirurgica* 151.3, pp. 231–238.

Starr, J. B., Gold, L., McCormick, Z., Suri, P., and Friedly, J. (2019). "Trends in lumbar radiofrequency ablation utilization from 2007 to 2016." In: *The Spine Journal* 19.6, pp. 1019–1028.

Steffen*, J., Hille*, G., Becker, M., Saalfeld, S., and Tönnies, K. (2020). "Automatic segmentation of necrosis zones after radiofrequency ablation of spinal metastases." In: *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. *both authors contributed equally to this work, pp. 96–102.

Štern, D., Likar, B., Pernuš, F., and Vrtovec, T. (2011). "Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images." In: *Phys Med Biol* 56.23, p. 7505.

Taha, A. A. and Hanbury, A. (2015). "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool." In: *BMC Medical Imaging* 15.1, p. 29.

Togawa, D. and Lewandrowski, K.-U. (2006). "The pathophysiology of spinal metastases." In: *Cancer in the Spine*, pp. 17–23.

Tomonari, A., Tsuji, K., Yamazaki, H., Aoki, H., Kang, J.-H., Kodama, Y., Sakurai, Y., and Maguchi, H. (2013). "Feasibility of the virtual needle tracking system for percutaneous radiofrequency ablation of hepatocellular carcinoma." In: *Hepatology Research* 43.12, pp. 1352–1355.

Tversky, A. (1977). "Features of similarity." In: *Psychological Review* 84.4, p. 327.

Van Dort, M. E., Rehemtulla, A., and Ross, B. D. (2008). "PET and SPECT imaging of tumor biology: new approaches towards oncology drug discovery and development." In: *Current Computer-Aided Drug Design* 4.1, pp. 46–53.

Viergever, M. A., Maintz, J. A., Klein, S., Murphy, K., Staring, M., and Pluim, J. P. (2016). *A survey of medical image registration–under review*.

Völzke, H., Alte, D., Schmidt, C. O., Radke, D., Lorbeer, R., et al. (2011). "Cohort profile: the Study of Health in Pomerania." In: *International Journal of Epidemiology* 40, pp. 294–307.

Von Dresky, C., Ballhausen, H., Becker, M., Serowy, S., Skalej, M., and Rieder, C. (2018). "Optimization of generator modulation and electrode positions for the radiofrequency ablation of spine metastases." In: *International Conference on Computer- und Roboterassistierte Chirurgie (CURAC)*, pp. 175–180.

Vujaskovic, Z., Gillette, S. M., Powers, B., LaRue, S., Gillette, E., et al. (1994). "Effects of intraoperative hyperthermia on peripheral nerves: neurological and electrophysiological studies." In: *International Journal of Hyperthermia* 10, pp. 41–49.

Wallace, A., Tomasian, A., Vaswani, D., Vyhmeister, R., Chang, R., and Jennings, J. (2016). "Radiographic local control of spinal metastases with percutaneous radiofrequency ablation and vertebral augmentation." In: *American Journal of Neuroradiology* 37.4, pp. 759–765.

Wallach, D., Toporek, G., Weber, S., Bale, R., and Widmann, G. (2014). "Comparison of freehand-navigated and aiming device-navigated targeting of liver lesions." In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 10.1, pp. 35–43.

Wang, J., Fang, Z., Lang, N., Yuan, H., Su, M.-Y., and Baldi, P. (2017). "A multi-resolution approach for spinal metastasis detection using deep

Siamese neural networks." In: *Computers in Biology and Medicine* 84, pp. 137–146.

Weihusen, A., Hinrichsen, L., Carus, T., Dammer, R., Rascher-Friesenhausen, R., Kröger, T., Peitgen, H.-O., and Preusser, T. (2010). "Towards a verified simulation model for radiofrequency ablations." In: *International Conference on Information Processing in Computer-Assisted Interventions (IPCAI)*. Springer, pp. 179–189.

Whyne, C. M., Hu, S. S., and Lotz, J. C. (2003). "Burst fracture in the metastatically involved spine: development, validation, and parametric analysis of a three-dimensional poroelastic finite-element model." In: *Spine* 28.7, pp. 652–660.

Wiese, T., Yao, J., Burns, J. E., and Summers, R. M. (2012). "Detection of sclerotic bone metastases in the spine using watershed algorithm and graph cut." In: *International Conference on SPIE Medical Imaging*. International Society for Optics and Photonics, pp. 831512–831512.

Wong, D. A., Fornasier, V. L., and MacNab, I. (1990). "Spinal metastases: the obvious, the occult, and the impostors." In: *Spine* 15, pp. 1–4.

Yamane, T., Tateishi, A., Cho, S., Manabe, S., Yamanashi, M., Dezawa, A., Yasukouchi, H., et al. (1992). "The effects of hyperthermia on the spinal cord." In: *Spine* 17, pp. 1386–1391.

Yao, J., O'Connor, S. D., and Summers, R. (2006). "Computer aided lytic bone metastasis detection using regular CT images." In: *International Conference on SPIE Medical Imaging*. International Society for Optics and Photonics, pp. 614459–614459.

Yin, J. J., Pollock, C. B., and Kelly, K. (2005). "Mechanisms of cancer metastasis to the bone." In: *Cell Research* 15.1, p. 57.

Zhang, P., Liu, H., Sun, Z., Wang, J., and Wang, G. (2020). "The application of O-arm and navigation system in precise localization of spinal cord lesions: a case series study." In: *Clinical Neurology and Neurosurgery*, p. 105922.

Zhang, Y., Matuszewski, B. J., Shark, L. K., and Moore, C. J. (2008). "Medical image segmentation using new hybrid level-set method." In: *International Conference on BioMedical Visualization (MEDIVIS)*, pp. 71–76.

Zhang, Y. (2016). "Radiofrequency ablation systems and operating mechanisms." In: *Radiofrequency Ablation for Small Hepatocellular Carcinoma*, pp. 25–33.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., et al. (2004). "Statistical validation of image segmentation quality based on a spatial overlap index 1: Scientific reports." In: *Academic Radiology* 11, pp. 178–189.

Zuckerman, S. L., Laufer, I., and Bilsky, M. (2018). "NOMS framework." In: *Metastatic Spine Disease: A Guide to Diagnosis and Management*, p. 41.

Zukić, D., Vlasák, A., Egger, J., Hořínek, D., Nimsky, C., and Kolb, A. (2014). "Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images." In: *Computer Graphics Forum* 33.6, pp. 190–204.

EHRENERKLÄRUNG

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich habe insbesondere nicht wissentlich

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht,
  um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, 16. Juni 2020

Georg Hille